Automated Labeling of Movement-Related Cortical Potentials Using Segmented Regression

Usman Rashid, Imran Khan Niazi, Mads Jochumsen, Laurens R. Krol, Nada Signal, and Denise Taylor

***Abstract*— The movement-related cortical potential (MRCP) is a brain signal related to planning and execution of motor tasks. From an MRCP, three notable features can be identified: the early Bereitschaftspotential (BP1), the late Bereitschaftspotential (BP2), and the negative peak (PN). These features have been used in past studies to quantify neurophysiological changes in response to motor training. Currently, either manual labeling or *a priori* specification of time points is used to extract these features. The limitation of these methods is the inability to fully model the features. This paper proposes the segmented regression along with a local peak method for automated labeling of the features. The proposed method derives the onsets, amplitudes at onsets, and slopes of BP1 and BP2 along with time and amplitude of the PN in a typical average MRCP. To choose the most suitable regression technique a bounded segmented regression method, a change point method and multivariate adaptive regression splines were evaluated using the root-mean-square error on a dataset of 6000 simulated MRCPs. The best-performing regression technique combined with the local peak method was then applied to a smaller set of 123 simulated MRCPs. Error in onsets of BP1 and BP2 and time of PN were compared with the errors in manual labeling by an expert. The performance of the proposed method was also evaluated on an experimental dataset of MRCPs derived from electroencephalography (EEG) recorded across two sessions from 22 healthy participants during a lower limb task. The Bland–Altman plots were used to evaluate the absolute reliability of the proposed method. On experimental data, the proposed method was also compared with manual labeling by an expert. Bounded segmented regression produced the smallest error on the simulation data. For the experimental data, our proposed method did not exhibit statistically significant bias in any of the modeled features. Furthermore, its performance was comparable to manual labeling by experts. We conclude that the proposed method can be used to automatically obtain robust estimates for the MRCP features with known measurement error.**

***Index Terms*— Electroencephalography (EEG), bereitschaftspotential (BP), early BP (BP1), late BP (BP2), movement-related cortical potential (MRCP), automatic detection, segmented regression.**

I. Introduction

MOVEMENT-RELATED cortical potentials (MRCPs) represent the cortical activity related to motor preparation and execution [1]–[3]. From an MRCP, three notable features can be identified: the early Bereitschaftspotential (BP1), the late Bereitschaftspotential (BP2), and the negative peak (PN). The Bereitschaftspotential (BP) starts around 2.0 seconds before the movement onset. Around 0.4 seconds before the onset, its slope abruptly steepens. These two slopes are characterized with reference to the baseline electroencephalography (EEG) activity as BP1 and BP2 [4]. These components of the MRCP have been studied across many different populations [5]–[7], notably in the context of motor training [8]–[12]. One motivation to use EEG data is to provide a low cost and simple method to quantify motor training compared to other approaches such as functional magnetic resonance imaging, magnetoencephalography, and transcranial magnetic stimulation [13]. For EEG data based quantification, measures of BP1, BP2, and PN play an important role.

A. Problem Statement

A range of methods have been employed to obtain BP1, BP2, and PN. In some of the past studies, onsets of BP1 and BP2 were specified *a priori* from previous research and their amplitudes were obtained for evaluation [8], [12], [14]. Whereas in other studies, BP1, BP2 and PN were identified from the MRCP with visual inspection by single or multiple experts [15], [16] who applied their knowledge of MRCPs to label points of interest on the signals. The limitation of the first method is that it limits the analysis to amplitudes as onsets are pre-specified. The limitation of the second method is that it cannot be applied automatically.

Manuscript received November 16, 2018; revised February 4, 2019 and March 31, 2019; accepted April 24, 2019. Date of publication May 7, 2019; date of current version June 6, 2019. *(Corresponding author: Usman Rashid.)*

U. Rashid, N. Signal, and D. Taylor are with the Health and Rehabilitation Research Institute, Auckland University of Technology, Auckland 0627, New Zealand (e-mail: urashid@aut.ac.nz).

I. K. Niazi is with the Health and Rehabilitation Research Institute, Auckland University of Technology, Auckland 0627, New Zealand, also with the Centre for Sensory-Motor Interaction, Department of Health Science and Technology, Aalborg University, 9220 Aalborg, Denmark, and also with the Centre for Chiropractic Research, New Zealand College of Chiropractic, Auckland 1060, New Zealand.

M. Jochumsen is with the Centre for Sensory-Motor Interaction, Department of Health Science and Technology, Aalborg University, 9220 Aalborg, Denmark.

L. R. Krol is with the Department of Biological Psychology and Neuroergonomics, Technische Universität Berlin, 10623 Berlin, Germany

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2019.2913880

Another limitation of these methods is that after specifying or labeling the onsets of BP1 and BP2, their amplitudes were obtained by taking an average over a range of the signal [11], [16]. Averaging over a range reduces the variability due to the local noise present in the signal. However, it discounts the negative slopes of BP1 and BP2 [4], as the mean statistic can only partially model a linear trend which has both a non-zero slope and a non-zero intercept. In yet another approach, the difference in amplitudes at two time points was used [15], which partially captures the change over time aspect of the slopes but the amplitudes taken at single time points from the signal are susceptible to local noise. Thus, there is a need for an automated method of identifying the MRCP features, which fully captures the underlying signal and is not susceptible to local noise.

B. Related Work

In the context of Lateralized Readiness Potential (LRP) which is another movement-related potential, there has been an effort to propose an automatic detector for its onset. Schwarzenau and Falkenstein proposed a novel method based on segmented regression for estimating the onset of LRPs [17]. Their method fitted two straight line segments to the LRP and labeled the LRP onset at the intersection of the two fitted lines. Alternatively, Miller, Patterson and Ulrich proposed the use of a set voltage threshold for the onset detection of the LRP from the group grand average followed by a jack-knifing procedure to obtain a standard error for the estimated onset [18]. Later, their jack-knifing method was further improved by Smulders [19]. A detailed investigation of these methods for detection of the LRP onset was conducted by Mordkoff and Gianaros [20]. They suggested that regression based methods should be used and the best performing method was the one which only allowed the slope of the second line to vary [20]. The obvious disadvantage of using the jack-knifing approach is that the onset and its standard error is computed for the entire group and an arbitrary threshold voltage has to be specified *a priori*. As these methods were proposed and applied to label a single onset of LRP, these methods can not be directly applied to label the two distinct onsets of BP1 and BP2 in a MRCP.

C. Novel Contributions

The aim of this research was to propose an automated method for the identification of BP1, BP2, and PN from a MRCP. To achieve this aim, we propose a local peak method for labeling of PN, and segmented regression for labeling of BP1, BP2. To choose the most suitable regression technique, we present a novel bounded segmented regression method which uses particle swarm optimization and evaluate it against a change point method and multivariate regression splines. To evaluate these techniques, we also propose a method to simulate MRCPs.

The simulation provides datasets of MRCPs in which the onsets and amplitudes of BP1, BP2 and PN are known. The simulation results are used to select the most suitable regression method for labeling of BP1 and BP2. The selected regression method combined with the local peak method are proposed for the identification of BP1, BP2, and PN from a MRCP. The performance of the proposed method was then evaluated both on simulated and experimental MRCP data and compared to manual labeling by experts. We also validated the simulated MRCPs against experimental MRCPs.

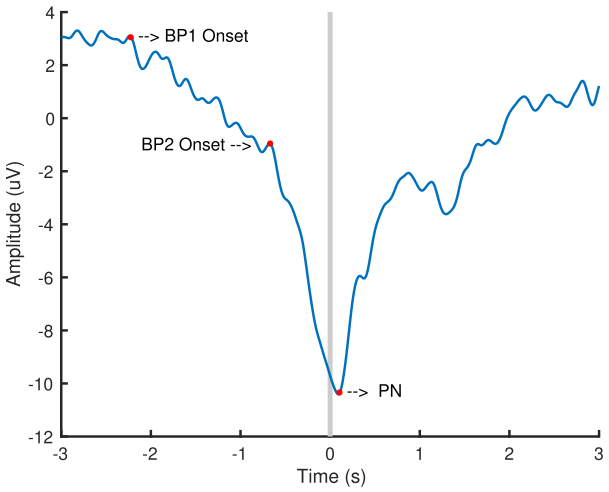

Fig. 1. An example MRCP obtained from averaging of EEG activity over fifty right foot ballistic dorsiflexions performed by a healthy person. '0' seconds represents time of the movement onset detected from two sEMG electrodes placed on the right Tibialis Anterior (TA) muscle.

II. Methods

Based on the comprehensive discussion of MRCPs found in "What is the Bereitschaftspotential?" by Shibasaki and Hallet [4], we make the following assumptions about the MRCP. Refer to Figure 1 for an example signal.

1) BP1 and BP2 can be modeled by two straight lines with non-zero slopes and intercepts.
2) PN can be modeled as a negative peak and is in close vicinity to the movement onset.
3) EEG activity before the onset of BP1 can be modeled as a straight line with a zero slope but a non-zero intercept.

A. Methods for Labeling MRCPs

In order to label the PN in an MRCP, we used the local peak method (LPM), which in addition to finding the minimum value, also ensures that the adjacent samples are at higher values [21]. LPM finds the PN in the vicinity of the movement onset. We defined this vicinity as a 1 second window before and after the movement onset to cater to the large variation across MRCPs. *findpeaks* function from MATLAB 2017b (MathWorks, Inc., Natick, MA, USA) was used to apply this method.

1) Labeling of BP1 and BP2: In order to label the onsets of BP1 and BP2, we used the MRCP signal from 3.0 seconds before the movement onset up to PN, represented by $y(n)$. To this signal, we fitted three line segments. The first segment corresponds to the baseline activity, the second to BP1 and the third to BP2, as given below.

$$\hat{y}(n) = \begin{cases} b_1 & n \leq n_1 \\ m_2 n + b_2 & n_1 < n \leq n_2 \\ m_3 n + b_3 & otherwise \end{cases} \quad (1)$$

where n is the sample number, $\hat{y}(n)$ is the fitted value of the MRCP signal at sample number n, $[m_2, m_3]$ and $[b_1, b_2, b_3]$ are the slopes and the intercepts of the three segments respectively; and n_1 and n_2 are the onsets of BP1 and BP2. The first segment has only an intercept term as it represents the baseline EEG which is assumed as such. For known n_1 and n_2, the coefficients of this model were obtained with least squares linear regression. To find n_1 and n_2, we evaluated the following three methods.

a) Bounded Segmented Regression (BSR): To fit the model defined by Equation 1, we set up the following optimization problem.

$$\begin{aligned} \min_{n_1,n_2} \; & \sum_{n=1}^{n_1} |b_1 - y(n)| \\ & + \sum_{n=n_1+1}^{n_2} |m_2 n + b_2 - y(n)| \\ & + \sum_{n=n_2+1}^{L} |m_3 n + b_3 - y(n)| \\ & n_1^l \le n_1 \le n_1^u \\ & n_2^l \le n_2 \le n_2^u \end{aligned} \tag{2}$$

This problem finds n_1 and n_2 corresponding to BP1 and BP2 within the provided bounds for the three straight line segments fitted to the recorded signal $y(n)$ which has L number of samples. Given a set of n_1, n_2, it also computes the coefficients (b_1, m_2, b_2, m_3, b_3) of the three line segments. Thus, this problem encompasses two optimization problems. (i) Finding n_1 and n_2. (ii) Nested inside (i), the second problem is finding the best fit line segments on intervals defined by n_1 and n_2.

The first problem is defined as minimizing the sum of L^1 norms of errors over the three segments. L^1 norm was chosen over the L^2 norm as the first segment which represents the baseline activity is modeled only with an intercept (b_1) and zero slope. As L^2 norm squares the errors, it would give more weight to the first segment which has only one degree of freedom. This formulation for finding n_1, n_2 does not allow guaranteed global optimum. Therefore, to increase the chances of finding the global or a near-global optimal solution, we used particle swarm optimization algorithm (PSO) which is a global approach to optimization [22]. PSO has been found to produce better results compared to traditional approaches in solving similar non-linear regression problems [23]–[25]. Furthermore, the optimization was run twice starting at randomly chosen n_1, n_2, and the solution with the smaller cost was selected to increase the chances of finding the global optimal or a near-global optimal solution. *particleswarm* function from MATLAB 2017b (MathWorks, Inc., Natick, MA, USA) was used for this purpose. Swarm size was set to 6. The lower and upper bounds for n_1 were set at sample numbers corresponding to -2.5 seconds and -1.0 seconds with respect to the movement onset respectively. The bounds for n_2 were set at -1.0 seconds to the time of PN with respect to the movement onset. These are reasonably large bounds keeping in mind the variations across MRCPs.

The second problem is finding three best fit line segments on intervals defined by n_1 and n_2. These three problems were solved by using least squares regression. This minimizes the L^2 norm of the errors and it is a convex optimization problem with a guaranteed global optimum. *mldivide* function from MATLAB 2017b (MathWorks, Inc., Natick, MA, USA) was used for this purpose.

b) Change Point Method (CPM): This method was selected from the change point literature. It uses an exhaustive algorithm based on dynamic programming to find n_1 and n_2 while minimizing a linear function for the underlying segments [26], [27]. The advantage of this method is that it is exhaustive while providing fast convergence. It has two disadvantages. First, it is unbounded and thus can find n_1 and n_2 at any point in $y(n)$. Second, it does not fully satisfy the third assumption which is that the baseline has a zero slope. Rather, it fits a line which has a non-zero slope to the baseline. To run this method, *findchangepts* function from MATLAB 2017b (MathWorks, Inc., Natick, MA, USA) was used with linear statistic and maximum number of change points set to 2.

c) Multivariate Adaptive Regression Splines (MRS): Multivariate adaptive regression splines with two knots corresponding to the onsets of BP1 and BP2 was also evaluated. This method fits a piece-wise linear function with three segments to $y(n)$. Traditionally the fitted model is stated in the following form.

$$\hat{y}(n) = a_0 + a_1 max(0, n - n_1) + a_2 max(0, n - n_2) \tag{3}$$

where $[a_0, a_1, a_2]$ are the coefficients of the fitted model. Although it satisfies the third assumption, like the change point method it is also unbounded. To determine n_1 and n_2, ARESLab: Adaptive Regression Splines toolbox version 1.13.0 was used [28], [29].

B. Proposed Method for Simulation of MRCPs

For appraisal of the above methods, we set up the simulation of MRCPs in SEREEGA [30]. It is a MATLAB (MathWorks, Inc., Natick, MA, USA) based open-source toolbox dedicated to the generation of simulated epochs of event-related EEG data.

We simulated a typical average MRCP by summing together two separate event-related potentials (ERPs) modeled using a Gaussian function. The peak latency, peak amplitude, and total width of the first ERP was set at -0.5 seconds, -2.5 uV and 3.0 seconds respectively. The second ERP was added at 0 seconds, with a peak amplitude of -10 uV and total width of 1.0 seconds. Summed together, these resulted in an MRCP with BP1 onset at -1.5 seconds, BP2 onset at -0.5 seconds, and PN at 0 seconds, as shown in Figure 2. The parameters of the simulation were based on the variability reported in the past research. Shibasaki and Hallet reported that BP1 starts around 2.0 seconds before the movement onset, followed by BP2 which starts around 0.4 seconds before the movement onset [4]. Wright et. al. in a study on motor planning reported that BP1 started around -1.79 $\pm$ SD 0.28 seconds and BP2 started around -0.72 $\pm$ SD 0.21 seconds in a group of non-musicians [16]. In the past, means and standard deviations for different features were not always stated explicitly. Thus, the values for these parameters were

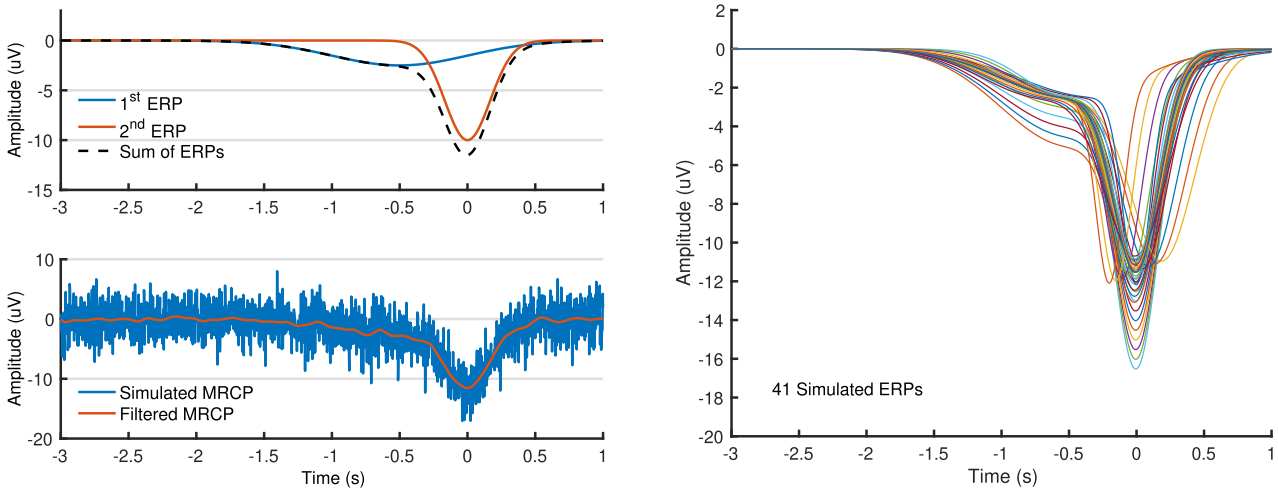

Fig. 2. On the left, a simulated MRCP obtained with SEREEGA/MATLAB (MathWorks, Inc., Natick, MA, USA) by adding two ERPs and white noise (6 dB). On the right, the simulated ERPs derived by applying the 41 variations.

derived from the figures in these articles [4], [7], [11], [13], [15], [16]. Following 41 random variations were added to the simulation. These variations are also shown in Figure 2. Only one variation was applied to each simulated MRCP.

1) BP1 onset. The onset of BP1 was varied from -1.5 to -2.0 seconds with a step of 0.1 seconds. This variation was applied by changing the width of the first ERP in the simulation.
2) BP2 onset. The onset of BP2 was varied from -0.3 to -0.7 seconds with a step of 0.05 seconds. It was applied by changing the latency of the first ERP. An equal change was also applied to the width of both ERPs to keep the remaining potentials at the same latencies.
3) PN time. The time of PN was varied from -0.2 to 0.2 seconds with a step of 0.05 seconds. It was applied by changing the latency of the second ERP. An equal change was also applied to the width of the second ERP to keep the remaining potentials at the same latencies.
4) BP2 amplitude. To produce variation in the amplitude of BP2, the peak amplitude of the first ERP was varied from -2.5 to -5 uV with a step of 0.5 uV.
5) PN amplitude. To produce variation in the amplitude of PN, the peak amplitude of the second ERP was varied from -10 to -15 uV with a step of 0.5 uV.

The latencies of BP1, BP2 and PN, and the amplitudes of BP1, BP2, PN obtained from the sum of the two ERPs at the latencies of BP1, BP2, time of PN provided the ground-truth data. A white noise was added to the sum of ERPs to produce simulated MRCPs under 3 signal to noise ratios (SNR) (6 dB, 3 dB, 0 dB) [31]. The signal to noise ratio was set by changing the root mean square value of the white noise with respect to the peak amplitude of the second ERP which corresponds to amplitude of PN. The simulated MRCPs were filtered with a low pass, 2nd order, zero-phase Butterworth filter with cut-off at 5 Hz. These filtered MRCPs were used for further analysis. Thus, in the subsequent text, a simulated MRCP refers to the simulated MRCP obtained at the end of the filtering process.

C. Simulated MRCPs

Three sets of MRCPs were simulated. Set I contained 2000 MRCPs for each SNR condition with random variations. Set II contained 41 MRCPs for each SNR condition with one MRCP corresponding to each variation. Set III was simulated in order to evaluate the validity of the simulated MRCPs. This set consisted of 42 simulated MRCPs, whose BP1, BP2 and PN latencies and amplitudes were copied from expert labels of experimental MRCPs (discussed next). To account for potential MRCPs with non-zero baselines, without having an explicit measure of baseline amplitude, BP1 amplitude was first subtracted from BP2 and PN amplitudes before simulation, and BP1 amplitude was subsequently added to the entire simulated MRCP after lowpass filtering. The simulation procedure was otherwise identical. This set allowed us to investigate the extent to which simulated MRCPs match experimental ones, as described in section II-F.3.

D. Experimental MRCPs

The experimental MRCPs used in this research were recorded from 22 healthy participants (Average age: 36 ± 6 years, 10 female) who were recruited through professional networks and local advertising [31]. Participants were excluded if they had a history of any neurological disorders or epilepsy. All the participants signed a written consent form before data collection. Ethical approval for the study (17/CEN/133) was obtained from Central Health and Disability Ethics Committee (HDEC), New Zealand in accordance with the Declaration of Helsinki.

The participants performed 50 self-paced ballistic dorsiflexions in two data collection sessions over two days. The laboratory setup is illustrated in Figure 3. The data was bifurcated into session I and II across the two recording days. Thus, session I corresponds to recording on the first day and session II corresponds to recording on the second day.

Movement onsets were obtained from two surface electromyography (sEMG) electrodes placed on the right Tibialis Anterior (TA) muscle. EEG data was cleaned with

Fig. 3. An illustration of the laboratory setup for recording of EEG and sEMG data from a participant performing self-paced ballistic dorsiflexions.

filtering, channel interpolation, manual and automatic epoch rejection, and independent component analysis (ICA) in EEGLAB/MATLAB (MathWorks, Inc., Natick, MA, USA). 47.1 ± 10.6 and 48.8 ± 1.5 epochs were retained after rejection at 125 uVpp threshold for session I and II respectively. Whilst the standard deviation of the rejection rate appears higher in session I, this can be explained by the rejection of all the epochs for one of the participants whose sEMG data was lost. During the ICA analysis, components corresponding to eye blinks, or limited to only one electrode and a few epochs, were removed. *runica* algorithm was used and 3 components were removed on average. MRCPs were obtained from FC3, FCz, FC4, C3, Cz, C4, CP3, CPz, CP4 channels by applying a small laplacian filter with center at Cz, followed by a 2nd order, zero phase, Butterworth filter with a low pass cut-off at 5 Hz, and averaging across epochs. SNR for the MRCPs in session I and II was 6.00 ± 2.39 dB and 5.70 ± 2.40 dB, respectively.

E. Manual Labeling of MRCPs

For comparison of the proposed method against experts, simulated MRCPs from set II (n = 123) and the experimental MRCPs were manually labeled. Experts with 3-8 years experience working with MRCPs, manually labeled both the simulated and the experimental MRCPs using a custom MATLAB 2017b (MathWorks, Inc., Natick, MA, USA) graphical user interface tool. This tool presented a MRCP as a 6 second epoch centered at the movement onset, with 3 seconds of data before and after the movement onset as shown in Figure 1. The experts labeled the onsets of BP1, BP2 and the PN using the mouse pointer. BP1 onset was labeled at the beginning of the negative slope, BP2 onset was labeled at the time when the slope increased abruptly, and PN was labeled at the time of the most negative amplitude. Amplitudes at the onsets of BP1, BP2 and the time of PN were obtained from the labeled time points. The slopes of BP1 and BP2 were mathematically computed from these time points and amplitudes.

F. Statistical Analysis

The analysis was performed in MATLAB 2017b (MathWorks, Inc., Natick, MA, USA). We used root-mean-square error (RMSE) to compare errors incurred by different methods on both the simulated and experimental datasets. RMSE was interpreted as the accuracy of the method with the ideal value being zero [20], [32]. RMSE was used instead of sample mean as RMSE incorporates both the central tendency and the variability of the data. Pair-wise permutation test was used to evaluate statistical differences in RMSE across different methods [33]. Beside the validity of the measure under consideration, a permutation test requires that the assumption of exchangeability is satisfied [34]. As in our case errors were obtained by different methods from the same data, the assumption of exchangeability required that the errors from the same pair be randomly permuted [35]. Significance level was set at 0.05.

1) Evaluation on Simulated MRCPs: For evaluation of LPM, CPM, MRS and BSR techniques set I of the simulated MRCPs was used (n = 6000). Failure rate, RMSE for the onsets of BP1, BP2 and time of PN were obtained for comparison. RMSEs were also obtained for the amplitudes of MRCPs at the onsets of BP1, BP2 and time of PN both from the simulated signals and the fitted models. From the fitted models, the amplitudes at the onsets of BP1, BP2 and time of PN were obtained from Equation 1 as b_1, $m_2 n_2 + b_2$ and $m_3 L + b_3$, respectively. This comparison was performed based on the hypothesis that the amplitudes obtained at single points from the simulated MRCPs would have larger error compared to the amplitudes obtained at single points from the fitted models. Essentially, this allowed us to see how much closer the model came to the ground-truth (pre-noise MRCP), compared to the simulated signal (noise added filtered MRCP). Using the results from these analyses, better performing techniques were combined to formulate our proposed method for automated labeling of the MRCPs.

From set II of the simulated MRCPs (n = 123), onsets of BP1, BP2 and time of PN were obtained using the proposed method and compared against those labeled by an expert. The error for the time points was obtained by subtracting the ground-truth from the labeled time. RMSE was used for comparison.

2) Evaluation on Experimental MRCPs: Onsets, amplitudes at onset and slopes of BP1, BP2, and time and amplitude of PN were obtained for the experimental MRCPs (n = 22 × 2) using the proposed method. Absolute reliability of the proposed method to identify these features across the two sessions was evaluated using Bland-Altman plots [36]. Bias, t-test for bias equal to zero, and coefficient of repeatability (CR) were reported. Bias was calculated as the mean of differences across the two sessions and interpreted as the systematic difference in the feature across the two sessions. Significance level for the t-test was set at 0.05. CR was calculated as 1.96 times the standard deviation of differences and interpreted as the measurement error below which the absolute differences between two sessions would lie with 0.95 probability [36].

The proposed method was compared to manual labeling by an expert in terms of the error across sessions in the onsets, amplitudes at onsets and slopes of BP1, BP2 along with time and amplitude of PN. This comparison was performed based on the hypothesis that the automated method would result in smaller error as compared to the manual labeling method. RMSEs was obtained for this comparison.

TABLE I
RMSES IN SECONDS FOR ONSETS OF BP1, BP2 AND TIME OF PN WITH RESPECT TO THE MOVEMENT ONSET UNDER DIFFERENT SNR CONDITIONS IN SIMULATED MRCPS. SIM. VAR. AND SIM. STEP STAND FOR THE SIMULATED VARIABILITY AND THE SIMULATED STEP. H_0 STANDS FOR NULL HYPOTHESIS

SNR	Potential	Method				Sim. Var.	Sim. Step	p-value H_0: BSR = CPM
		MRS	CPM	BSR	LPM			
6 dB	BP1	1.133	0.548	0.442	–	0.5	0.1	<0.001
	BP2	0.289	0.163	0.164	–	0.4	0.05	0.090
	PN	–	–	–	0.021	0.4	0.05	–
3 dB	BP1	0.919	0.627	0.518	–	0.5	0.1	<0.001
	BP2	0.247	0.169	0.170	–	0.4	0.05	0.168
	PN	–	–	–	0.034	0.4	0.05	–
0 dB	BP1	0.843	0.668	0.551	–	0.5	0.1	<0.001
	BP2	0.228	0.206	0.195	–	0.4	0.05	0.019
	PN	–	–	–	0.053	0.4	0.05	–

3) Validation of Simulated MRCPs: To evaluate the validity of the simulated MRCPs, the experimental MRCPs were compared to their simulated copies in Set III. Cosine similarity between the experimental MRCPs and their simulated versions up to the negative peak was obtained as follows [37].

$$r = \frac{\mathbf{u}.\mathbf{v}}{\|\mathbf{u}\| \times \|\mathbf{v}\|} \tag{4}$$

where $\mathbf{u}$, $\mathbf{v}$ represent the two MRCPs as vectors. '.' represents the dot product between the two vectors, and $\|.\|$ represents the L^2 norm of a vector.

III. RESULTS

A. Evaluation on Simulated MRCPs

1) Failure Rate: The evaluated techniques did not always succeed in fitting a model. LPM was successful in finding PN in all cases under the three SNR conditions. Similarly, BSR was also successful in all cases under all the SNR conditions. The failure rate for CPM was 0%, 2.2% and 11.3% for 6 dB, 3 dB and 0 dB, respectively. For MRS, the failure rate was 0.35%, 1.25% and 1.95% for SNR at 6 dB, 3 dB and 0 dB, respectively. Thus, the bounded segmented regression method for finding onsets of BP1 and BP2 is clearly superior with respect to failure rate.

2) RMSEs for BP1, BP2 Onsets and Time of PN: The root mean square errors for the onsets of BP1 and BP2 and the time of PN are given in Table I. The error corresponding to each technique was calculated from successful cases only. The root mean square error for PN time was smallest, followed by BP2 and BP1 respectively. MRS performed poorly under all SNR conditions for both BP1 and BP2 onsets, thus, it was excluded from further analysis.

For BP1 onset, BSR produced smaller error under all three SNR conditions compared to CPM and these differences were also statistically significant ($p < 0.05$). Compared to simulated variation, BSR produced smaller error for BP1 onset at 6 dB, and larger error at 3 dB and 0 dB.

For BP2 onset, CPM produced smaller error at 6 dB and 3 dB compared to BSR. However, these differences were smaller than 0.001 seconds and also did not achieve statistical significance ($p > 0.05$). Whereas, at 0 dB, BSR had smaller ($p < 0.05$) error compared to CPM. Compared to simulated variation, both BSR and CPM produced smaller error under all SNR conditions. These results suggest that BSR is superior in case of BP1 onset and both BSR and CPM detect very similar BP2 onsets. Thus, CPM was dropped from further analysis at this stage.

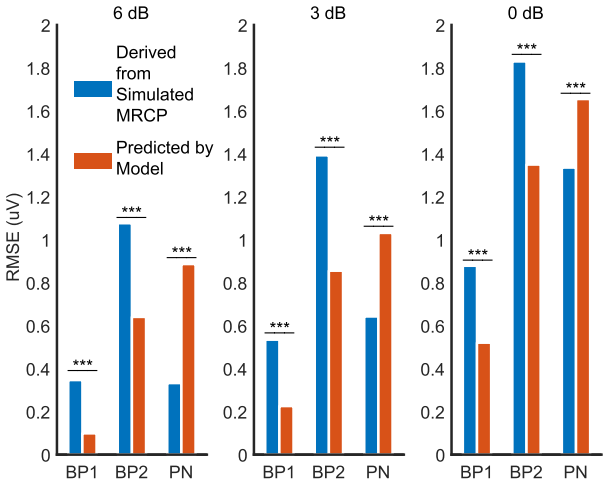

Fig. 4. RMSEs for amplitudes at onsets of BP1, BP2 and time of PN taken from the simulated MRCP and the fitted models using BSR and LPM under three signal to noise ratios (6 dB, 3dB, 0 dB). '***' denote $p < 0.001$ for the null hypothesis that the two RMSEs are equal.

3) Model Prediction: Using the BP1, BP2 onsets and time of PN, the amplitudes at these time points were obtained from both the fitted model as given by Equation 1 and the simulated MRCP signals. The errors are shown in Figure 4 for BSR and LPM. The error for the fitted model was smaller for BP1 and BP2 across all three SNR conditions compared to amplitudes taken from the simulated MRCP signals. Whereas, the error for PN was smaller for the signal amplitudes across all three SNR conditions. All these differences were statistically significant ($p < 0.05$).

4) Comparison With an Expert on Simulated MRCPs: Based on results from the previous sections, BSR was selected for identification of BP1 and BP2 onsets while LPM was selected for identification of the time of PN. These techniques were applied to the simulated data to fit models as given in Equation 1. The amplitudes for BP1, BP2 onsets were computed from the fitted model. The amplitude for PN was calculated from the MRCP signal. The slopes were taken

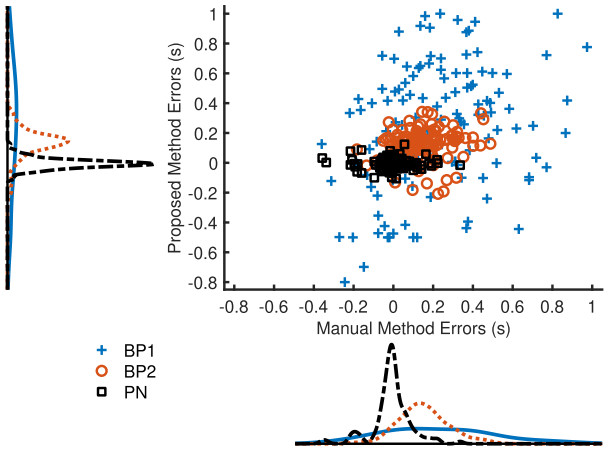

Fig. 5. Distribution of errors for onsets of BP1, BP2 and time of PN for manual and automatic labeling of the simulated MRCPs.

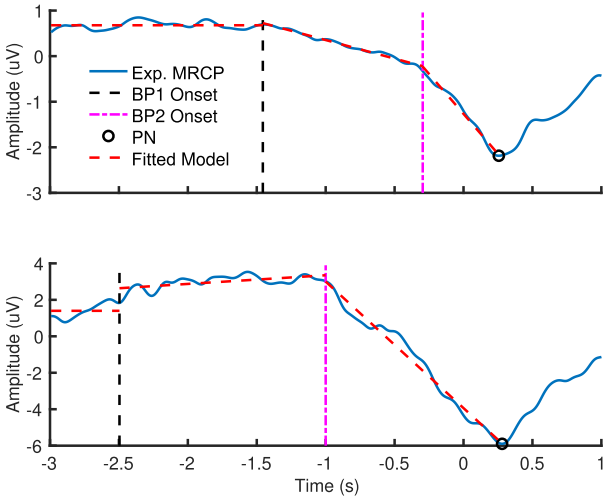

Fig. 6. Experimental MRCPs from two different participants labeled by the proposed method. The top MRCP had the lowest RMSE from the fitted model. Whereas, the bottom MRCP had the highest RMSE from the fitted model.

from the fitted model. The combination of these techniques is referred to as the proposed method in the text.

From simulated dataset II, the errors in the onsets of BP1, BP2, and time of PN identified by the proposed method are plotted against the error in labeling of these time points by an expert in Figure 5. This plot indicates that the errors from both the manual labeling and automated labeling were normally distributed.

The RMSE for onset of BP1 was 0.356 seconds for the expert and 0.478 seconds for the proposed method, respectively. Similarly, the RMSE for BP2 onset was 0.195 seconds and 0.163 seconds for the expert and the proposed method, respectively. In case of PN, the RMSE for the expert and the proposed method was 0.096 seconds and 0.036 seconds, respectively. The p-values for null hypotheses corresponding to RMSE in onsets of BP1, BP2 and time of PN were 0.001, 0.010 and <0.001 respectively. These results suggest that the proposed method incurred smaller error in identifying the onset of BP2 and the time of PN compared to the expert. Whereas, the proposed method had larger error in identifying the onset of BP1 compared to the expert.

B. Evaluation on Experimental MRCPs

The proposed method was applied to the experimental data. The results for two of the MRCPs are shown in Figure 6. The proposed method successfully found all the features in all the cases.

1) Absolute Reliability With Bland-Altman Plots: The Bland-Altman plots for the features obtained for the MRCPs from two sessions are given in Figure 7. The Bland-Altman plots do not exhibit any systematic trends, except in case of onset of BP1 and time of PN where the error variability appears to be non-uniform.

Furthermore, there was insufficient evidence to reject the null hypothesis that there was zero bias in any of the features. The coefficient of repeatability was smallest for the time of PN (0.279 seconds), followed by BP2 (0.571 seconds) and BP1 (1.279) respectively. On the other hand, for amplitudes an opposite trend in the size of CR was observed. CR was

TABLE II
RMSEs FOR ONSETS OF BP1, BP2, TIME OF PN, AMPLITUDES AT THESE TIME POINTS, AND SLOPES FOR BP1, BP2. THESE ERRORS WERE OBTAINED FROM BOTH THE MANUAL AND AUTOMATED LABELING OF EXPERIMENTAL MRCPs FROM TWO SESSIONS

RMSE for	Potential	Method		p-value
		Manual	Proposed	
Time (s)	BP1	0.408	0.640	0.126
	BP2	0.225	0.292	0.338
	PN	0.174	0.148	0.585
Amplitude (uV)	BP1	1.655	0.991	0.064
	BP2	1.189	1.505	0.060
	PN	2.277	2.547	0.488
Slope (uV/s)	BP1	2.100	1.531	0.496
	BP2	6.325	4.756	0.648

smallest for amplitude at onset of BP1 (1.988 uV), followed by amplitude at onset of BP2 (2.867 uV) and amplitude of PN (5.032 uV) respectively. Similarly, CR for slope of BP1 (3.016 uV/s) was smaller than that for BP2 (9.305 uV/s).

2) Comparison with an Expert on Experimental MRCPs: The comparison of RMSE from the proposed method and manual labeling for onsets of BP1, BP2, time of PN, amplitudes at these time points, and slopes for BP1, BP2 is given in Table II. There was not enough evidence ($p > 0.05$) to reject the null hypothesis that the two methods incurred same errors. Nonetheless, there were some notable differences in magnitude of the error across the two methods. The proposed method incurred larger RMSE in onset of BP1. Also, the proposed method achieved smaller error for the slopes of both BP1 and BP2.

C. Validation of Simulated MRCPs

Two experimental MRCPs along with their simulated counterparts in Set III, and the cosine similarity between the experimental MRCPs and their simulated versions is shown in Figure 8. The mean and standard deviation of the similarity index was 0.918 ± 0.078. These results suggest an excellent agreement between the experimental MRCPs and their simulated versions in majority of the cases.

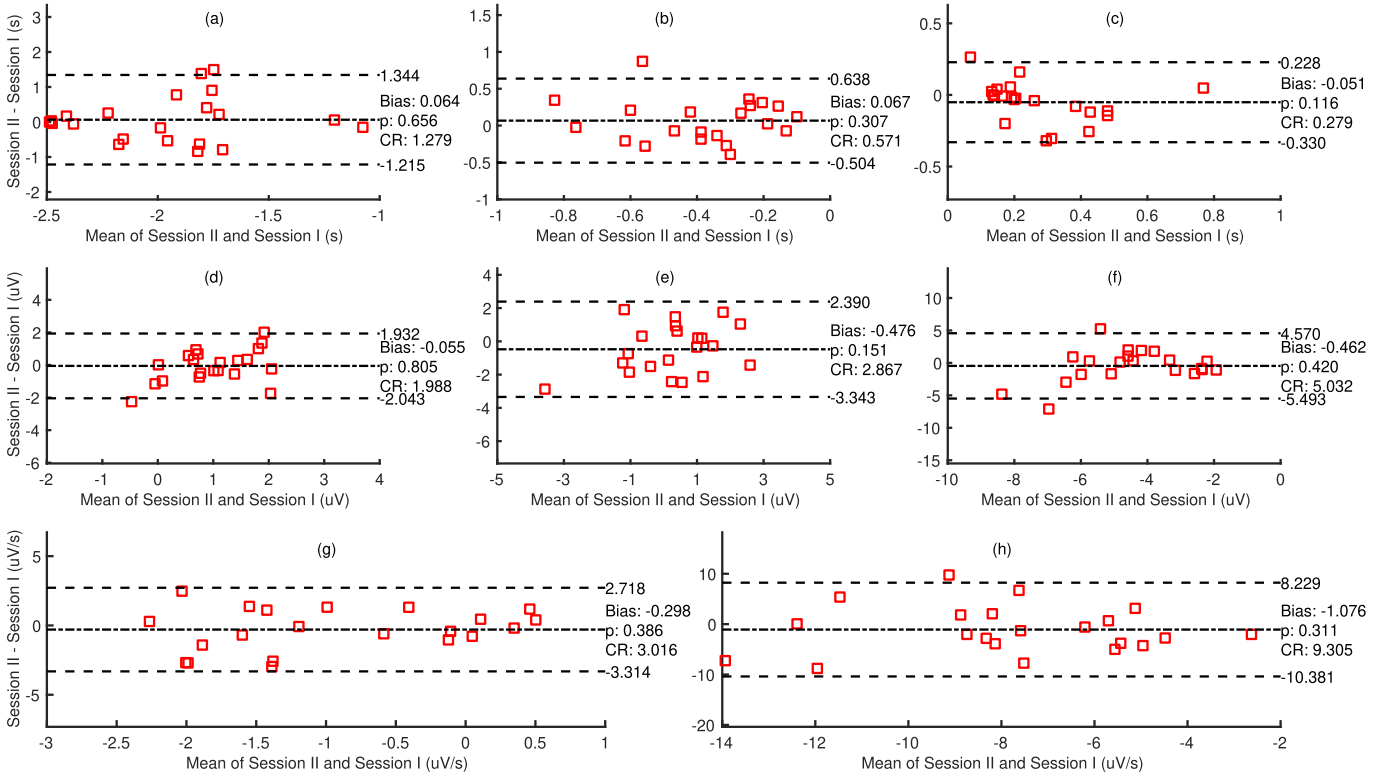

Fig. 7. Bland-Altman plot for onset of BP1 (a), onset of BP2 (b), time of PN (c), amplitude at onset of BP1 (d), amplitude at onset of BP2 (e), amplitude of PN (f), slope of BP1 (g), and slope of BP2 (h) for experimental MRCPs identified by the proposed method in session I and II.

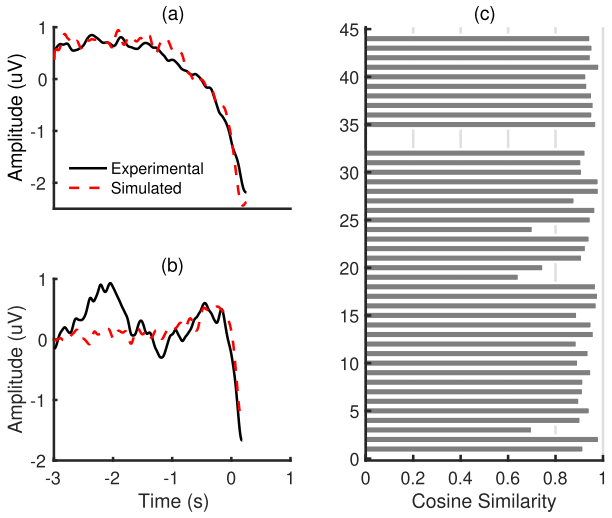

Fig. 8. (a), (b) Examples of experimental MRCPs along with their simulated counterparts. (a) corresponds to the case with highest cosine similarity of 0.985. (b) corresponds to the case with lowest cosine similarity of 0.646. (c) The cosine similarity between the experimental MRCPs and their simulated versions. MRCPs at no. 33 and 34 correspond to the participant whose sEMG data was lost during recording.

IV. Discussion

We have proposed and validated a method for automated labeling of movement-related cortical potentials. This method provides robust estimates of the MRCP features. On simulated MRCPs, BSR outperformed MSR and CPM in terms of failure rate, RMSEs in onsets of BP1, BP2, and time of PN. Also, simulation results showed that amplitudes obtained from fitted models result in smaller error compared to amplitudes directly taken from MRCP signals, except in the case of PN. Based on these results, BSR was selected for identification of BP1 and BP2 while LPM was selected for identification of PN. When compared against labeling of onsets of BP1, BP2, and time of PN on simulated MRCPs by an expert; the proposed method had larger ($p < 0.05$) error in the onset of BP1 while it had a smaller ($p < 0.05$) error in the onset of BP2 and time of PN. On the experimental MRCPs, the proposed method did not show any systematic trends in Bland-Altman plots of the features with minor exceptions discussed in the following paragraphs. There was insufficient evidence ($p > 0.05$) to suggest the presence of bias in features across the two recording sessions. Compared to manual labeling of MRCPs, the proposed method had larger RMSE in onset of BP1 and smaller RMSEs in slopes of both BP1 and BP2. However, these differences did not achieve statistical significance.

A. Evaluation on Simulated MRCPs

In terms of RMSE for onset of BP1 the BSR performed better than the other two methods, whereas its performance for BP2 onset was comparable to that of CPM. We also evaluated the failure rate. For CPM it was generally higher compared to BSR and considerably higher (approx. 11%) under 0 dB SNR. The MSR method consistently performed poorly in terms of both the failure rate and RMSE. Thus, MSR and CPM were excluded from further analysis.

The RMSEs of the amplitudes predicted from the fitted models at BP1 and BP2 onsets were smaller compared to amplitudes derived directly from signals. This difference can be explained by the fact that the fitted model smooths local

variability in the data. These results suggest that the MRCP amplitudes should not be measured directly from the signal at single time points. Rather, if necessary, the amplitudes predicted by the model should be used to reduce error due to local variability which can be a result of signal noise. However, for PN, opposite results were found. The error for the amplitude obtained from signal was smaller than that of the predicted amplitude. This can be explained by the fact that near PN, the MRCP signal violates the assumption that it can be modeled as a straight line. PN is a peak which deviates away from the fitted straight line. Thus, in case of PN, the amplitude should be calculated from the signal itself. These results have implications for researchers interested in amplitudes at single time points or differences in amplitudes measured at multiple single time points from MRCPs [15]. Furthermore, as highlighted in the introduction, amplitudes at single time points or differences across multiple time points do not fully capture the linear trends of BP1 and BP2, thus, we suggest the use of fitted slopes.

In comparison to manual labeling by an expert, the proposed method had smaller RMSEs in identifying the onset of BP2 and time of PN. However, it had larger errors in identifying the onset of BP1. One plausible explanation for this finding is that it is due to the break down of assumption 1 and 3 (refer to Section II) at the onset of BP1. Under current assumptions, the onset is assumed to be an intersection point between two lines. Perhaps, the actual transition from baseline to BP1 is radically smoother than an edge and can only be correctly modeled by a spline. We suggest it is investigated in future research.

B. Evaluation on Experimental MRCPs

The Bland-Altman plots did not show any systematic trends except in case of the onset of BP1 and time of PN where the error variability was not uniform. The implication of this trend is that the CR obtained under the assumption of normally distributed errors are not true representations of the agreement limits [36]. The coefficient of repeatability was also reported for each feature which can be used to interpret pre and post intervention scores in a future study involving MRCPs [36]. For example, for an individual participant any change greater than 9.305 uV/s in the slope of BP2 from pre to post can be interpreted as a real difference with 95% confidence.

In comparison to manual labeling by an expert, the proposed method had larger error for onset of BP1. This can again be explained by the spline hypothesis suggested by the results of the comparison between the expert and the proposed method in simulated MRCPs. The proposed method consistently achieved smaller measurement error in slopes of BP1 and BP2. These differences, however, did not achieve statistical significance and need further investigation in a larger sample.

C. Validation of Simulated MRCPs

The similarity between the experimental MRCPs and their simulated versions suggested an excellent agreement between the two in majority of the cases. Thus, the simulation method provided large test sets of valid MRCPs for which the ground truth was known. Additionally, simulated MRCPs gave us the ability to evaluate our proposed methods under large variations in time, amplitude and signal to noise ratio. Evaluation on experimental MRCPs under these variations would have not been possible without recording EEG data from a large number of participants from different populations.

The proposed simulation was based on findings of the past research. However, further work is required to adapt this simulation method to generate MRCPs at a single epoch level and demonstrate its validity.

D. Limitations

The findings of this research should be considered in light of a number of limitations. First, the MRCPs were labeled by a single expert. In some of the past studies, two experts labeled the MRCPs and the third expert selected the labeling of one of the experts. The reliability of these methods (single and multiple experts) has not yet been demonstrated. Second, the experimental MRCPs were recorded from 22 healthy participants. Thus, the estimates of measurement error and bias can not be applied across other populations without question. In future research, using large samples gathered from multiple populations, a full comparative analysis of the reliability of the proposed method and the different manual methods is needed.

E. Recommendations

1) The proposed method fully captures the two linear trends of BP1 and BP2 in the MRCP. Thus, we suggest that in future research, the slopes fitted by the proposed method be considered for studying the differences across conditions. Following our proposed method, one may automatically model MRCP parameters to quantify changes induced with an intervention or motor training paradigms [13], [38].
2) In studies where determining the onset of BP1 is of particular interest, we suggest the use of manual labeling. However, for the onset of BP2 and determining the time of PN; the use of the proposed method is suggested.
3) As the proposed method is fully automatic, a real-time implementation of particle swarm algorithm [39] can enable online neuroadaptive paradigms to support motor training [40].

F. Software Availability

The MATLAB (MathWorks, Inc., Natick, MA, USA) based implementation of the proposed method, the graphical user interface tool for visual labeling of MRCPs and tools for Bland-Altman analysis have been made available online[1]. These tools can be used to automatically obtain features from MRCPs using the proposed method or manually label the MRCPs for calculating the features discussed in this research.

V. Conclusion

We have proposed a method for automated labeling of features in movement-related cortical potentials. The proposed

[1] Available on Github at: https://github.com/GallVp/visualEEG.

method was formulated and validated using a large set of simulated MRCPs. Its absolute reliability was also evaluated on a set of experimental MRCPs. We conclude that the proposed method be used to automatically obtain robust estimates for the MRCP features with known measurement error in future studies involving MRCPs.

ACKNOWLEDGMENT

The authors would like to acknowledge the advice given by Associate Professor Alain C. Vandal (Auckland University of Technology, New Zealand) on optimization and analysis methods. The authors would also like to thank Rasmus Nedergaard, Samran Navid (Aalborg University, Denmark), and Sylvain Cremoux (University of Valenciennes and Hainaut-Cambresis, France) for software testing. We also appreciate the time taken by the anonymous reviewers to suggest important changes to the methods and the organization of this manuscript.

REFERENCES

[1] H. H. Kornhuber and L. Deecke, "Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale," *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere*, vol. 284, no. 1, pp. 1–17, 1965.

[2] M. Hallett, "Movement-related cortical potentials," *Electromyography Clin. Neurophysiol.*, vol. 34, no. 1, pp. 5–13, 1994.

[3] K. Sakamoto, H. Nakata, Y. Honda, and R. Kakigi, "The effect of mastication on human motor preparation processing: A study with CNV and MRCP," *Neurosci. Res.*, vol. 64, no. 3, pp. 259–266, 2009.

[4] H. Shibasaki and M. Hallett, "What is the bereitschaftspotential?" *Clin. Neurophysiol.*, vol. 117, no. 11, pp. 2341–2356, 2006.

[5] K. Kim, J. S. Kim, and C. K. Chung, "Increased gamma connectivity in the human prefrontal cortex during the bereitschaftspotential," *Frontiers Hum. Neurosci.*, vol. 11, p. 180, May 2017.

[6] A. L. Patil, S. K. Sood, V. Goyal, and K. P. Kochhar, "Cortical potentials prior to movement in Parkinson's disease," *J. Clin. Diagnostic Res.*, vol. 11, no. 3, pp. CC13–CC16, 2017.

[7] D. J. Wright, P. S. Holmes, and D. Smith, "Using the movement-related cortical potential to study motor skill learning," *J. Motor Behav.*, vol. 43, no. 3, pp. 193–201, 2011.

[8] M. J. Taylor, "Bereitschaftspotential during the acquisition of a skilled motor task," *Electroencephalogr. Clinical Neurophysiol.*, vol. 45, no. 5, pp. 568–576, 1978.

[9] W. Lang, M. Lang, A. Kornhuber, L. Deecke, and H. H. Kornhuber, "Human cerebral potentials and visuomotor learning," *Pflügers Archiv*, vol. 399, no. 4, pp. 342–344, 1983.

[10] J. Niemann, T. Winker, J. Gerling, B. Landwehrmeyer, and R. Jung, "Changes of slow cortical negative dc-potentials during the acquisition of a complex finger motor task," *Experim. Brain Res.*, vol. 85, no. 2, pp. 417–422, 1991.

[11] D. J. Wright, P. Holmes, F. Di Russo, M. Loporto, and D. Smith, "Reduced motor cortex activity during movement preparation following a period of motor skill practice," *PLoS ONE*, vol. 7, no. 12, 2012, Art. no. e51886.

[12] M.-K. Lu, N. Arai, C.-H. Tsai, and U. Ziemann, "Movement related cortical potentials of cued versus self-initiated movements: Double dissociated modulation by dorsal premotor cortex versus supplementary motor area rTMS," *Hum. Brain Mapping*, vol. 33, no. 4, pp. 824–839, 2012.

[13] M. Jochumsen *et al.*, "Quantification of movement-related eeg correlates associated with motor training: A study on movement-related cortical potentials and sensorimotor rhythms," *Frontiers Hum. Neurosci.*, vol. 11, p. 604, Dec. 2017.

[14] F. Fattapposta *et al.*, "Long-term practice effects on a new skilled motor learning: An electrophysiological study," *Electroencephalogr. Clin. Neurophysiol.*, vol. 99, no. 6, pp. 495–507, 1996.

[15] Y. Kita, A. Mori, and M. Nara, "Two types of movement-related cortical potentials preceding wrist extension in humans," *Neuroreport*, vol. 12, no. 10, pp. 2221–2225, 2001.

[16] D. J. Wright, P. S. Holmes, F. Di Russo, M. Loporto, and D. Smith, "Differences in cortical activity related to motor planning between experienced guitarists and non-musicians during guitar playing," *Hum. Movement Sci.*, vol. 31, no. 3, pp. 567–577, 2012.

[17] P. Schwarzenau, M. Falkenstein, J. Hoormann, and J. Hohnsbein, "A new method for the estimation of the onset of the lateralized readiness potential (LRP)," *Behav. Res. Methods, Instrum., Comput.*, vol. 30, no. 1, pp. 110–117, 1998.

[18] J. Miller, T. Patterson, and R. Ulrich, "Jackknife-based method for measuring LRP onset latency differences," *Psychophysiology*, vol. 35, no. 1, pp. 99–115, 1998.

[19] F. T. Y. Smulders, "Simplifying jackknifing of ERPs and getting more out of it: Retrieving estimates of participants' latencies," *Psychophysiology*, vol. 47, no. 2, pp. 387–392, 2010.

[20] J. T. Mordkoff and P. J. Gianaros, "Detecting the onset of the lateralized readiness potential: A comparison of available methods and procedures," *Psychophysiology*, vol. 37, no. 3, pp. 347–360, 2000.

[21] S. J. Luck, *An Introduction to the Event-Related Potential Technique*, vol. 1. Cambridge, MA, USA: MIT Press, 2014.

[22] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Proc. 6th Int. Symp. Micro Mach. Hum. Sci. (MHS)*, Oct. 1995, pp. 39–43.

[23] M. Adibifard, G. Bashiri, E. Roayaei, and M. A. Emad, "Using particle swarm optimization (PSO) algorithm in nonlinear regression well test analysis and its comparison with Levenberg–Marquardt algorithm," *Int. J. Appl. Metaheuristic Comput.*, vol. 7, no. 3, pp. 1–23, 2016.

[24] P. Erdoğmuş and S. Ekiz, "Nonlinear regression using particle swarm optimization and genetic algorithm," *Int. J. Comput. Appl.*, vol. 153, no. 6, pp. 28–36, Nov. 2016.

[25] N. Lu, J. Zhou, Y. He, and Y. Liu, "Particle swarm optimization for parameter optimization of support vector machine model," in *Proc. 2nd Int. Conf. Intell. Comput. Technol. Automat. (ICICTA)*, vol. 1, Oct. 2009, pp. 283–286.

[26] M. Lavielle, "Using penalized contrasts for the change-point problem," *Signal Process.*, vol. 85, no. 8, pp. 1501–1510, 2005.

[27] R. Killick, P. Fearnhead, and I. A. Eckley, "Optimal detection of changepoints with a linear computational cost," *J. Amer. Stat. Assoc.*, vol. 107, no. 500, pp. 1590–1598, 2012.

[28] J. H. Friedman, "Multivariate adaptive regression splines," *Ann. Statist.*, vol. 19, no. 1, pp. 1–67, 1991. doi: 10.1214/aos/1176347963.

[29] G. Jekabsons. *ARESLab: Adaptive Regression Splines Toolbox for MATLAB/Octave*. Accessed: 2016. [Online]. Available: http://www.cs.rtu.lv/jekabsons/

[30] L. R. Krol, J. Pawlitzki, F. Lotte, K. Gramann, and T. O. Zander, "SEREEGA: Simulating event-related EEG activity," *J. Neurosci. Methods*, vol. 309, pp. 13–24, Nov. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165027018302395

[31] U. Rashid, I. K. Niazi, N. Signal, and D. Taylor, "An EEG experimental study evaluating the performance of Texas instruments ADS1299," *Sensors*, vol. 18, no. 11, p. 3721, 2018.

[32] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature," *Geosci. Model Develop.*, vol. 7, no. 3, pp. 1247–1250, 2014.

[33] F. B. Butar and J.-W. Park, "Permutation tests for comparing two populations," *J. Math. Sci. Math. Edu.*, vol. 3, no. 2, pp. 19–30, 2008.

[34] V. W. Berger, "Pros and cons of permutation tests in clinical trials," *Statist. Med.*, vol. 19, no. 10, pp. 1319–1328, 2000.

[35] S. L. Simpson, R. G. Lyday, S. Hayasaka, A. P. Marsh, and P. J. Laurienti, "A permutation testing framework to compare groups of brain networks," *Frontiers Comput. Neurosci.*, vol. 7, p. 171, Nov. 2013.

[36] S. Vaz, T. Falkmer, A. E. Passmore, R. Parsons, and P. Andreou, "The case for using the repeatability coefficient when calculating test–retest reliability," *PLoS ONE*, vol. 8, no. 9, 2013, Art. no. e73990.

[37] N. Dehak, R. Dehak, J. Glass, D. Reynolds, and P. Kenny, "Cosine similarity scoring without score normalization techniques," in *Proc. Odyssey*, 2010, p. 15.

[38] N. Mrachacz-Kersting, S. R. Kristensen, I. K. Niazi, and D. Farina, "Precise temporal association between cortical potentials evoked by motor imagination and afference induces cortical plasticity," *J. Physiol.*, vol. 590, no. 7, pp. 1669–1682, 2012.

[39] L. Liu, W. Liu, D. A. Cartes, and N. Zhang, "Real time implementation of particle swarm optimiation based model parameter identification and an application example," in *Proc. IEEE Congr. Evol. Comput. (IEEE World Congr. Comput. Intell.)*, Jun. 2008, pp. 3480–3485.

[40] T. O. Zander, L. R. Krol, N. P. Birbaumer, and K. Gramann, "Neuroadaptive technology enables implicit cursor control based on medial prefrontal cortex activity," *Proc. Nat. Acad. Sci. USA*, vol. 113, no. 52, pp. 14898–14903, 2016.