



Markov Switching Model for Quick Detection of Event Related Desynchronization in EEG

Giuseppe Lisi*, Diletta Rivela, Asuka Takai and Jun Morimoto

ATR Computational Neuroscience Laboratories, Department of Brain Robot Interface, Kyoto, Japan

OPEN ACCESS

Edited by:

Mikhail Lebedev,
Duke University, United States

Reviewed by:

Kazutaka Takahashi,
University of Chicago, United States
Dan Zhang,
Tsinghua University, China
Thomas C. Bulea,
National Institutes of Health (NIH),
United States

*Correspondence:

Giuseppe Lisi
glisi@atr.jp

Specialty section:

This article was submitted to
Neuroprosthetics,
a section of the journal
Frontiers in Neuroscience

Received: 09 November 2017

Accepted: 12 January 2018

Published: 01 February 2018

Citation:

Lisi G, Rivela D, Takai A and
Morimoto J (2018) Markov Switching
Model for Quick Detection of Event
Related Desynchronization in EEG.
Front. Neurosci. 12:24.
doi: 10.3389/fnins.2018.00024

Quick detection of motor intentions is critical in order to minimize the time required to activate a neuroprosthesis. We propose a Markov Switching Model (MSM) to achieve quick detection of an event related desynchronization (ERD) elicited by motor imagery (MI) and recorded by electroencephalography (EEG). Conventional brain computer interfaces (BCI) rely on sliding window classifiers in order to perform online continuous classification of the rest vs. MI classes. Based on this approach, the detection of abrupt changes in the sensorimotor power suffers from an intrinsic delay caused by the necessity of computing an estimate of variance across several tenths of a second. Here we propose to avoid explicitly computing the EEG signal variance, and estimate the ERD state directly from the voltage information, in order to reduce the detection latency. This is achieved by using a model suitable in situations characterized by abrupt changes of state, the MSM. In our implementation, the model takes the form of a Gaussian observation model whose variance is governed by two latent discrete states with Markovian dynamics. Its objective is to estimate the brain state (i.e., rest vs. ERD) given the EEG voltage, spatially filtered by common spatial pattern (CSP), as observation. The two variances associated with the two latent states are calibrated using the variance of the CSP projection during rest and MI, respectively. The transition matrix of the latent states is optimized by the “quickest detection” strategy that minimizes a cost function of detection latency and false positive rate. Data collected by a dry EEG system from 50 healthy subjects, was used to assess performance and compare the MSM with several logistic regression classifiers of different sliding window lengths. As a result, the MSM achieves a significantly better tradeoff between latency, false positive and true positive rates. The proposed model could be used to achieve a more reactive and stable control of a neuroprosthesis. This is a desirable property in BCI-based neurorehabilitation, where proprioceptive feedback is provided based on the patient’s brain signal. Indeed, it is hypothesized that simultaneous contingent association between brain signals and proprioceptive feedback induces superior associative learning.

Keywords: Markov switching model, Bayesian estimation, quickest detection, event related desynchronization, sensorimotor rhythms, electroencephalogram, neuroprosthesis, brain computer interface

1. INTRODUCTION

Sensorimotor rhythm-based brain robot interfaces (BRI) have recently gathered attention in the field of neurorehabilitation (Daly and Wolpaw, 2008). In this context, rehabilitation is conducted by activating a device that assists movement using the brain signal. The assisted movement produces sensory input that is hypothesized to induce central nervous system (CNS) plasticity, leading to the restoration of normal motor control. Endogenous brain computer interfaces (BCI), such as the one based on sensorimotor rhythms (SMR), have been used in motor neurorehabilitation as an effective tool for promoting neuroplasticity of neuromuscular pathways (Silvani et al., 2011).

SMR are modulated by movement or motor imagery, and they are often referred to as event-related (de)synchronization (ERD/ERS). They can be detected by non-invasive methods such as the electroencephalogram (EEG). An ERD is a power decrease of mu (7–13 Hz) and beta (13–30 Hz) rhythms that occur in the sensorimotor areas during a motor-related task, while an ERS is a power increase following the offset (i.e., end) of the task (Pfurtscheller and Lopes da Silva, 1999). The ERD/ERS elicited by motor imagery have been used to control neuroprosthetic devices such as functional electrical stimulation (FES) to achieve the hand motion in spinal cord injury (Müller-Putz et al., 2005) or stroke patients (Daly et al., 2009), and to activate upper (Gomez-Rodriguez et al., 2011; Ramos-Murguialday et al., 2012; Sarac et al., 2013) or lower (Do et al., 2013; Lisi et al., 2014) limb exoskeleton robots (Lisi and Morimoto, 2017).

Most notably, clinical trials have been carried out to verify the effectiveness of ERD-based brain robot interfaces (BRI) for motor function recovery (Ang et al., 2009, 2014; Broetz et al., 2010; Caria et al., 2011; Shindo et al., 2011; Ramos-Murguialday et al., 2013; Naros and Gharabaghi, 2015). In these works, the motion of a neuroprosthesis is either triggered (Shindo et al., 2011) or continuously controlled (Ramos-Murguialday et al., 2012) based on the output of a BCI. In this context, minimizing the latency between the motor intention onset and the device activation is critical (Muralidharan et al., 2011), since it has been hypothesized that simultaneous contingent association between brain oscillations and proprioceptive feedback leads to superior associative learning and elicits motor learning (Ramos-Murguialday et al., 2012, 2013). For the same reason, an asynchronous BCI strategy, where a decoder continuously estimates the mental state (i.e., rest vs. motor imagery) of a subject, is more suitable. Here asynchronous refers to the fact that a decoder continuously analyses the EEG data, irrespectively of the cue given to the subject, trying to maximize true positives during the motor imagery and to minimize the false positives during the rest or idling state (Townsend et al., 2004).

Asynchronous BCI has received less attention compared to synchronous BCI (Lotte et al., 2007). The most conventional approach to real-time SMR-based asynchronous decoding is sliding window classification (Townsend et al., 2004; Muralidharan et al., 2011; Shindo et al., 2011; Ang et al., 2012; Lisi et al., 2016). In such systems, a sliding window is required in order to compute statistics, e.g., variance, associated with the spectral features of the signal. Usually, longer sliding windows

(i.e., 1 s) are chosen since they provide smooth variance estimates over time and a more stable output, which comes at the cost of a larger latency in ERD detection. Reducing the threshold of the classifier may reduce the latency, at the cost of a larger false positive rate (Muralidharan et al., 2011). On the other hand, shorter windows would minimize the latency, with the drawback of a higher feature variability and unstable output. Therefore, it becomes clear (**Figure 1**) that such systems are characterized by a

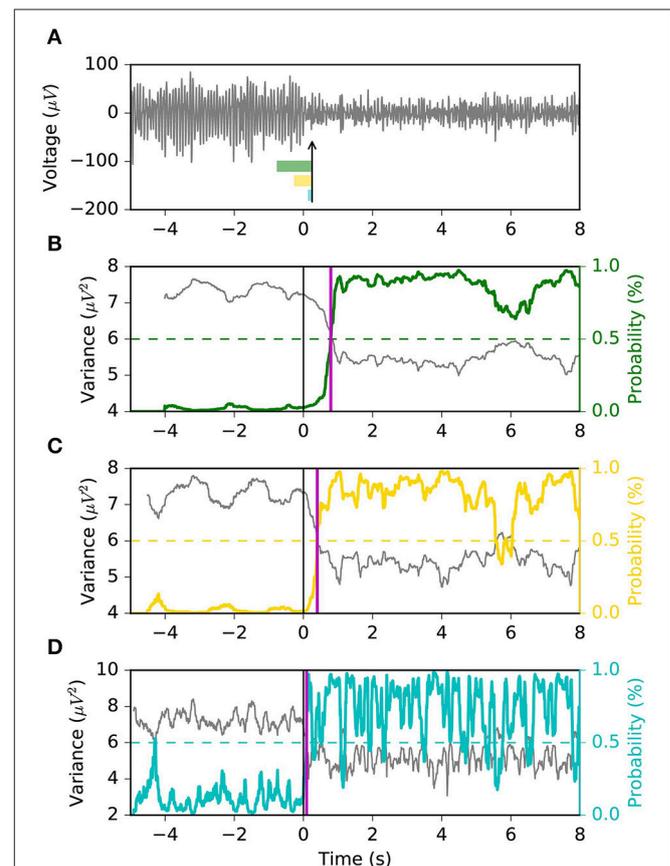


FIGURE 1 | Problem statement. The sliding window approach introduces a delay in the ERD detection. A long sliding window achieves a stable output (i.e., less false positives and more true positives) at the cost of a larger detection latency. A shorter sliding window is faster at detecting the ERD, but its output is less stable and has larger variability. Indeed, it is important to note that the left vertical axis of (**D**) covers a broader range. Panel (**A**) shows the EEG voltage obtained after optimal bandpass filtering and CSP spatial filtering (using the most discriminative CSP component) as described in section 2.4.3. The three rectangles represent the sliding window lengths (i.e., 1.0, 0.5, and 0.1 s) used for the models shown in (**B–D**). The vertical arrow indicates the point in time corresponding to the output of the depicted sliding windows. Panels (**B–D**) represent the decoding associated with sliding window lengths 1.0, 0.5, and 0.1 s, respectively. Each shows the variance, computed on the signal in (**A**) by the respective sliding window, with a solid gray line. The probability output of the logistic regression classifier is drawn with a solid colored line. The vertical solid black line shows the time of the motor imagery cue onset. The horizontal dashed colored line shows the classifier's threshold with respect to the classifier's probability output (solid colored line). The vertical solid magenta line shows the estimated latency of detection. This example is taken from a subject performing left motor imagery.

trade-off between detection latency, false positive rate (FPR) and true positive rate (TPR).

Here, we propose to avoid the use of a sliding window for the explicit calculation of the signal variance, and instead estimate the likelihood of the ERD state directly from the instantaneous EEG voltage. Based on the observation that often times the ERD occurs abruptly, we use a Markov switching model (MSM), a method that is suitable in applications where the latent state of a system changes suddenly, such as in the economics field (Hamilton, 2010). In our implementation, the model takes the form of a Gaussian observation model whose variance is governed by two latent discrete states with ergodic (i.e., fully connected) Markovian dynamics. The objective is to estimate the brain state (i.e., rest vs. ERD) given the EEG voltage, spatially filtered by common spatial pattern (CSP), as observation (Figure 2). The two variances associated with the two latent states are calibrated offline using the variance of the CSP projection during rest and motor imagery, respectively. Intuitively, the first state (S_0) has larger probability during the baseline resting period, while the second state (S_1) has larger probability during motor imagery. The transition matrix of the latent states is optimized by using a variant of the “quickest detection” strategy (Poor and Hadjiladis, 2009) that minimizes a cost function of detection latency and false positive rate. A similar approach has been taken in the development of an online seizure detection method based on intracranial EEG (Santaniello et al., 2011).

Markov and state space models in general, are not widespread within the non-invasive BCI community, but they are promising classifiers (or decoders) for BCI systems (Lotte et al., 2007). In literature, Hidden Markov models (HMM) have been proposed

with the objective of maximizing classification performance of left vs. right motor imagery, without explicitly targeting the rest class nor the trade-off between detection latency and FPR reduction. The HMM proposed in previous works (Obermaier et al., 2001; Chiappa and Bengio, 2003; Cincotti et al., 2003; Rezaei et al., 2006) are left-to-right finite automata meant to model the temporal changes of the EEG during a motor imagery task. The rationale is that sensorimotor rhythms have specific temporal characteristics, e.g., a short-latency ERD is often followed by a ERS. Each class of interest is assigned a separate HMM trained with trials from that specific class. An unknown trial is classified according to the HMM model with the highest probability, as calculated by Viterbi algorithm. Therefore, each HMM represents one class, and each state of a HMM represents a specific temporal state of that class. In Obermaier et al. (2001) and Rezaei et al. (2006) the classification is done synchronously, meaning that the rest class is not modeled. In Cincotti et al. (2003) and Chiappa and Bengio (2003) the classification is asynchronous, however the rest class is not modeled and not evaluated (i.e., FPR not computed). The observations of an HMM are spectral (Obermaier et al., 2001; Cincotti et al., 2003) or statistical features (Rezaei et al., 2006) of the EEG signal computed on a sliding window. Such models are not suitable to classify rest vs. motor imagery, since temporal dynamics during rest are not well determined. Moreover, the need of a sliding window for feature extraction makes them comparable to a sliding window classifier from a detection latency point of view. On the other hand, the proposed MSM is an ergodic model where no specific temporal sequence is modeled, and any state can be visited at any time. In the MSM the rest class is modeled as a state with high variance, and is included in the evaluation criteria (i.e., FPR). Moreover, a sliding window is not necessary, as previously explained, making it possible to track closely abrupt changes associated with the ERD.

2. METHODS

In this section we describe the modules composing the neural decoding pipeline (Figure 3). We begin by explaining data acquisition, followed by online decoding pipeline, the MSM and offline parameter estimation. The last subsection describes the methodology used to assess the performance of the proposed method.

2.1. Data Acquisition

To explore the MSM model and its application to non-invasive neural decoding, we have carried out motor imagery experiments with 50 healthy subjects: age 24 ± 3 , 9 females and 41 males. Only one subject was familiar with motor imagery BCI, while all the others had never performed a motor imagery BCI task before. During a single run, a subject performed 10 trials of cued motor imagery, interleaved by a period of rest. For each subject, 4 runs were collected, interleaved by a few minutes of rest. It should be noted that even though the experiment is cued, the subsequent decoding is done irrespectively of the cue (i.e., asynchronous decoding), and that a cued experiment is needed in order to know the ground truth onset of the motor imagery. For

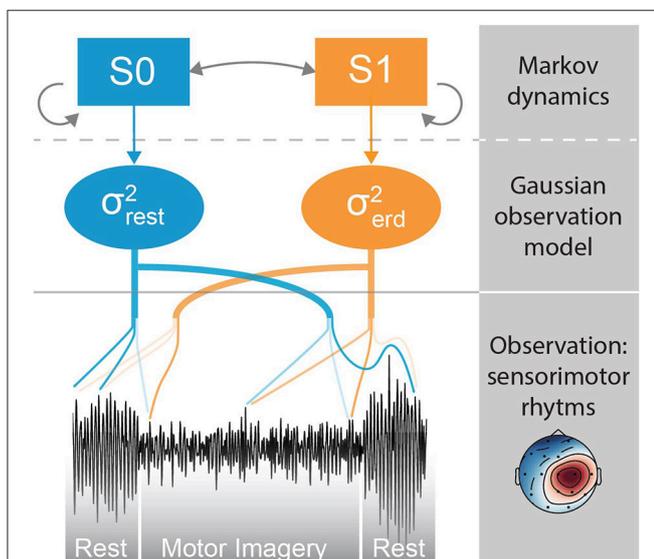


FIGURE 2 | Conceptual representation of the Markov switching model (MSM). The sensorimotor rhythms are obtained by CSP spatial filtering. The MSM is composed of a Gaussian observation model whose variance is governed by Markovian dynamics. If the variance of the sensorimotor rhythm is high, the probability of state S_0 is high; if the variance is low, the probability of S_1 is high. The probability of S_1 represents the probability of an ERD.

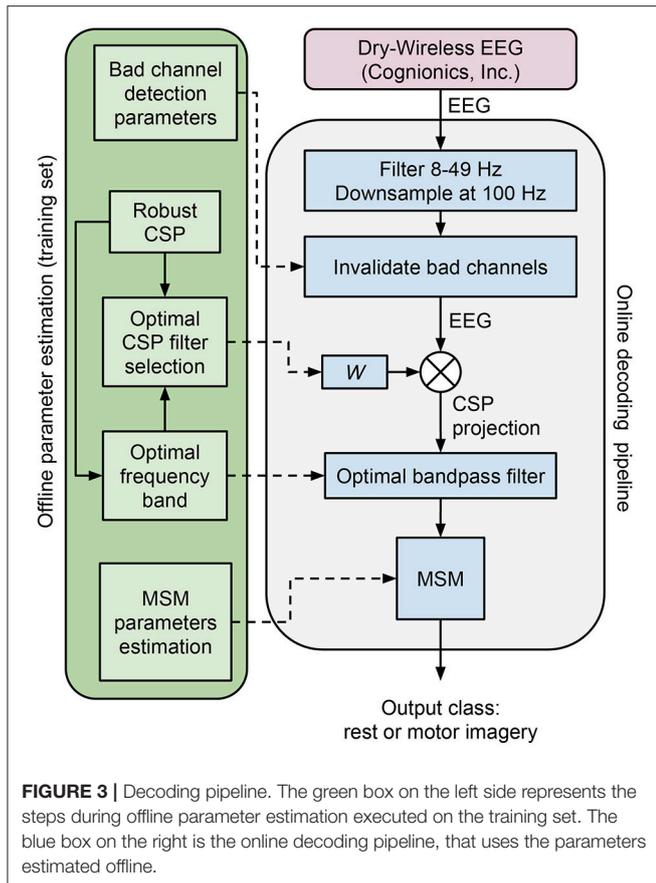


FIGURE 3 | Decoding pipeline. The green box on the left side represents the steps during offline parameter estimation executed on the training set. The blue box on the right is the online decoding pipeline, that uses the parameters estimated offline.

a given subject, the motor imagery type was fixed. The majority of the participants ($N = 47$) was randomly assigned either to a left ($N = 25$) or right ($N = 22$) hand motor imagery task, with 7 s rest duration and 5 s motor imagery duration. The additional 3 subjects performed a foot motor imagery task with 8 s rest duration and 4 s motor imagery duration. The EEG signal was collected at a sampling rate of 500 Hz, by the Quick-20 dry-wireless headset (Cognionics, Inc.), which is a full 10–20 array, with 19 channels (F7, Fp1, Fp2, F8, F3, Fz, F4, C3, Cz, P8, P7, Pz, P4, T3, P3, O1, O2, C4, T4) plus reference on A1 and ground on A2. Participants gave written informed consent for the experimental procedures, which were approved by the ATR Human Subject Review Committee.

2.2. Online Decoding Pipeline

Here we describe the online decoding pipeline (Figure 3, blue box on the right), that uses the optimal parameters computed offline by the method detailed in section 2.4. A subset of channels that is typically associated with sensorimotor activity is selected (i.e., F3, Fz, F4, C3, Cz, Pz, P4, P3, C4). The signal is bandpass filtered in the 8–49 Hz range and downsampled at 100 Hz. Outlying channels are identified using the deviation criterion (see section 2.4.1), and temporarily invalidated by assigning them the average of the valid channels. Subsequently, the projection of the most discriminative CSP filter is computed and further bandpass

filtered within the optimal frequency band. The bandpass filtered CSP projection is then given as an observation to the MSM in order to estimate the current class (i.e., rest or motor imagery).

2.3. Markov Switching Model (MSM)

The MSM was first introduced in the field of econometrics (Hamilton, 1989; Turner et al., 1989), but to the authors' knowledge it has never been applied on EEG motor imagery data. The model is sketched in Figure 2, and it consists of a Gaussian observation model whose variance is governed by two latent states with Markovian state transitions. The observation model is a simple zero-mean Gaussian with variance σ^2 :

$$y_t \sim \mathcal{N}(0, \sigma_t^2) \quad (1)$$

where y_t is the observation at time t and σ_t^2 is dependent on a discrete state S_t :

$$\sigma_t^2 = \begin{cases} \sigma_{rest}^2 & \text{if } S_t = 0 \\ \sigma_{erd}^2 & \text{if } S_t = 1 \end{cases} \quad (2)$$

where $S_t = 0$ is associated with the resting state, and $S_t = 1$ is associated with a motor imagery ERD. Since S_t is hidden (i.e., we do not know when to apply each sub-model), we use a weighted combination of each sub-model (i.e., *soft switching*), where the weights are given by $P(S_t = i | y_{1:t})$. Therefore, the resulting system can be thought of as a mixture of Gaussian models (Murphy, 1998).

The discrete state $S_t = 0$ or 1 denotes the latent space of the system and it is generated by a realization of a first-order Markov process with transition probabilities defined by the matrix:

$$\mathbf{Z} = \begin{bmatrix} p & 1-p \\ 1-q & q \end{bmatrix}, \quad (3)$$

where $p = P(S_t = 0 | S_t = 0)$ and $q = P(S_t = 1 | S_t = 1)$ represent the probability of remaining in $S = 0$ or remaining in $S = 1$, respectively. The offline calibration of the variances σ_{rest}^2 , σ_{erd}^2 , and probabilities p and q is described in section 2.4.4.

At each iteration, the new probability of each discrete state is computed according to Bayes' theorem:

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{normalization factor}} \quad (4)$$

The *likelihood* for each discrete state i is:

$$\mathcal{L}_i = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-y^2/2\sigma_i^2} \quad (5)$$

where σ_i^2 is either σ_{rest}^2 or σ_{erd}^2 and y is the EEG voltage spatially filtered by a CSP filter (see section 2.4.2). The *prior* in Equation 4, which we denote by \bar{c} , is computed by propagating the previous probabilities of the discrete states according to the transition probability matrix \mathbf{Z} :

$$\bar{c} = \mathbf{Z} \cdot \mu \quad (6)$$

where μ is the vector containing the previous probabilities of the discrete states. The numerator of Equation 4 can simply be computed as:

$$\tilde{\mu}_i = \bar{c}_i \times \mathcal{L}_i \quad (7)$$

and then normalized (i.e., *normalization factor*) so that it sums to one:

$$\mu_i = \frac{\tilde{\mu}_i}{\sum_{j=0}^1 \tilde{\mu}_j} \quad (8)$$

The new probability of each discrete state is contained in μ_i .

Given a CSP projection observation y_t at time t , the probability of state $S_t = 0$ associated with variance at rest (σ_{rest}^2) is larger when the subject is in resting state, while the probability of state $S_t = 1$ associated with variance during motor imagery (σ_{erd}^2) is larger when the subject does motor imagery.

2.4. Offline Parameters Estimation

Offline parameter estimation (Figure 3, green box on the left) is performed on the training data set, before running online decoding. Signal preprocessing is equivalent to the one used for online decoding.

2.4.1. Bad Channel Detection Parameters

We implemented two types of bad channel detection methods, based on the *correlation* and *deviation* criteria, respectively (Bigdely-Shamlo et al., 2015). The former criterion labels a channel as bad if its maximum correlation coefficient with the other channels is below a threshold (0.4 by default). Here, the correlation coefficients are computed on the same epochs used by the CSP algorithm (section 2.4.2). This procedure is carried out offline and, once a channel is rejected by the correlation criterion, it is removed completely from both offline parameters estimation and online decoding. On the other hand, the *deviation* criterion is applied at each time sample of both offline and online processing. In the deviation criterion, outlying channels are identified using the modified Z-score method (Iglewicz and Hoaglin, 1993): a channel i at time t whose modified Z-score $M_i(t) = 0.6745(x_i(t) - \bar{x}_i)/MAD_i$ is larger than 5 is considered as an outlier. The parameters \bar{x}_i and MAD_i are computed using the entire EEG of the training set and they represent the median and the median absolute value (MAD) of the EEG channel, respectively. Channels marked as bad at time t are temporarily invalidated by assigning them the average of the valid channels. In the case that most channels are bad, the minority would still be used online to compute final decoding and evaluate the performance measures.

2.4.2. Robust CSP

A robust version of the CSP algorithm (Yong et al., 2008) is executed on the training set to estimate the spatial filters, robust to outliers, that maximize the difference in variance between the two classes (i.e., rest vs. motor imagery). For this purpose, from the bandpass filtered EEG signal (8–49 Hz) of the training set, we cut epochs of 2 s length from 1 s to 3 s with respect to the motor imagery onset, for the motor imagery class. For the longer rest class, the epochs start 1 s after rest onset and end 1 s before rest offset. The length discrepancy between motor imagery and

rest epochs is due to the fact that we want the covariance matrix of the rest class to capture as much as possible the variability of the rest EEG, to make the CSP filters more robust. Finally, we select the first 3 and last 3 columns of the CSP unmixing matrix, representing the most discriminative spatial filters (Figure 4, top row). The CSP filters selected at this point are redundant, so that the CSP filter selection at the next step has enough options to choose from.

2.4.3. Optimal Frequency Band and CSP Filter Selection

For each CSP filter, the most discriminative frequency band (Figure 4) is computed using the heuristic proposed in Blankertz et al. (2008). Frequency bands with a high correlation coefficient between the logpower and classes labels are iteratively added to the optimal frequency band: the frequency with the largest correlation is selected and then the adjacent frequencies (i.e. above and below) are added if their correlation is at least 90% of the best correlation. Correlation coefficients are computed on the spectrogram of the same epochs used for CSP (section 2.4.2). Likewise, in order to select the most discriminative CSP filter (Figure 4), we compute the average logpower within the optimal frequency band of each filter, and keep the one having the largest correlation coefficient with classes labels.

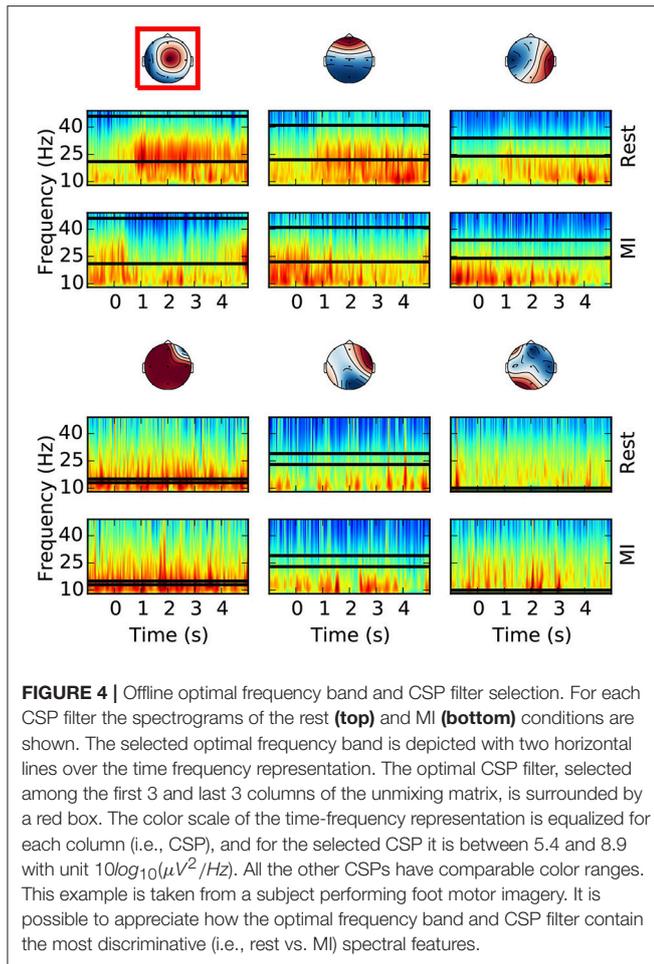
2.4.4. MSM Parameters Estimation

The parameters σ_{rest}^2 and σ_{erd}^2 are assigned the variances of the CSP projection during rest and motor imagery, respectively. Contrary to the CSP epochs, the durations of the rest (i.e., $[-3s, -1s]$ with respect to onset) and motor imagery (i.e., $[1s, 3s]$ with respect to onset) time epochs are chosen to be the same, so that the chance of them containing outliers is minimized. Indeed, it is important to remove epochs with extreme values, in order to obtain an unbiased estimate of variance. Therefore, variance is computed for all the epochs; then, epochs whose variance exceeds 3 standard deviations from the mean are removed. Once data is clean, the average variance of the rest and motor imagery epochs are computed.

For convenience sake, the transition probabilities (p, q) of the latent states are defined using the exponential duration model $\pi_d = 1 - \frac{1}{d}$, where d is the duration in number of frames and π_d is the probability of staying in a given state Rabiner (1989). Then, according to a modified “quickest detection” strategy (Poor and Hadjiliadis, 2009; Santaniello et al., 2011), the tuple (p, q) is optimized by minimizing a cost function of detection latency (δt) and false positive rate (FPR):

$$C = \delta t + FPR \quad (9)$$

The optimization is carried out by Sequential Least Squares Programming (Kraft, 1988), and the starting point of the optimization is set to the probabilities (p, q) computed from the durations of rest and motor imagery, respectively. With respect to the cost function, FPR is the false positive rate computed using the samples during the *rest* condition. Computing the detection latency δt is more complex, since it is important to avoid irrelevant short-lasting false positives. For this purpose, we use a sliding window approach to find the most robust state



transition $S_0 \rightarrow S_1$ across a trial. Accordingly, the detection time is the t_d that maximizes the difference $\bar{S}_{post} - \bar{S}_{pre}$, where \bar{S}_{post} is the average estimated state S in the range $[t_d, t_d + 3s]$ and \bar{S}_{pre} is the average state S in the range $[t_d - 1s, t_d]$. The rationale for having a longer range for \bar{S}_{post} is that in case of several $S_0 \rightarrow S_1$ transitions, the one with the longer stay in S_1 should be considered as the real ERD. Once t_d is computed, $\delta t = t_d - t_o$, where t_o is the time of the motor imagery cue onset. Trials that cannot be improved (i.e., $FPR = 0$) or trials where ERD does not exist ($TPR = 0$) should be excluded from the optimization. Based on this rationale and in order to reduce computational cost, only the five most improvable trials within the training set are used for optimization. This is done by looking for the trials in the training set that jointly minimize the absolute difference between FPR and the average FPR (i.e., $|FPR - \overline{FPR}|$) and maximize TPR.

2.5. Performance Evaluation

Leave-one-run-out cross-validation is used, separately for each subject, to evaluate the online performance of the proposed decoder: $n - 1$ runs are used as training set and 1 run as test set, so that every run is used as test set once.

2.5.1. Baseline Models

The MSM model is compared against a sliding window logistic regression (LR) classifier. Specifically, three different window's lengths are used: 1.0 s, 0.5 s, 0.1 s (i.e., $LR\ 1.0\ s$, $LR\ 0.5\ s$, $LR\ 0.1\ s$). The sliding window includes time samples prior to the current time, as shown in **Figure 1**, in order to avoid using future information. A logistic regression classifier is trained and tested on the log-transformed sliding window variance of the optimal CSP projection (Townsend et al., 2004). The decoding pipeline up to classification is the same as the one used for MSM. During classifiers' training, the time ranges used to represent the *rest* and *motor imagery* classes are equal to the ones used to tune the MSM: the sliding window whose last sample is within $[-3\ s, -1\ s]$ and $[1\ s, 3\ s]$, with respect to motor imagery onset, is assigned to the *rest* and *motor imagery* class, respectively. During online decoding and performance evaluation, the positive class of the classifier (i.e., motor imagery) corresponds to S_1 of the MSM.

2.5.2. Online Performance Measures

Three measures are used within each trial of each subject for model evaluation: the latency of detection (δ_t), the false positive rate (FPR), and the geometric mean (Barandela et al., 2003) of true positive rate (TPR) and true negative rate (TNR).

The latency of detection δ_t has been described in section 2.4.4. Here, however, we need to find consistent δ_t values across the four models being compared (i.e., MSM, $LR\ 1.0\ s$, $LR\ 0.5\ s$, $LR\ 0.1\ s$), and make sure that state estimates instability would not lead to incompatible δ_t values across models. Indeed, noisy state estimates of $LR\ 0.5\ s$ and $LR\ 0.1\ s$ may have a maximum of the value $\bar{S}_{post} - \bar{S}_{pre}$ at a very different δ_t compared to the more stable MSM and $LR\ 1.0\ s$, making the comparison difficult and often unfair. Such noisy state estimates would already be accounted for by the TPR and FPR. Therefore, we decided not to judge them again from the latency point of view, and force their latency to be close to that of the more stable models. We do this by extending the procedure described in section 2.4.4: first we find a common point of reference Δ_t between the more stable models MSM and $LR\ 1.0\ s$, then we compute the δ_t of each model with respect to Δ_t . The common point of reference Δ_t is found by applying the sliding window procedure (described in section 2.4.4) over the average estimated class of MSM and $LR\ 1.0\ s$. Then, for each model we look for the time t_d closest to Δ_t , where a transition $S_0 \rightarrow S_1$ occurs. Latency is computed as $\delta t = t_d - t_o$, where t_o is the time of the motor imagery cue onset.

The FPR, TPR, TNR and G-mean are computed for each trial (i.e., consecutive rest and motor imagery) assuming that samples up to 4 s before motor imagery cue onset belong to the negative class, and samples during the motor imagery cue belong to the positive class. The geometric mean ($G\text{-mean} = \sqrt{TPR \cdot TNR}$) is used here as an estimator of accuracy separately for each trial. It should be noted that if in a given trial there is no ERD, even an optimal decoder would produce an uninformative output: correct classification only 50% of the time samples during rest (i.e., $TNR = 0.5$) and motor imagery (i.e., $TPR = 0.5$), yielding $G\text{-mean} = 0.5$. Conversely, for a trial containing an ERD, a sufficiently good decoder is expected to have at least $TNR = 0.6$ and $TPR = 0.6$, yielding $G\text{-mean} = 0.6$. One attractive

property of G-mean (Barandela et al., 2003) is that it penalizes unbalanced TPR and TNR, compared to the arithmetic mean, e.g., $\sqrt{0.8 \cdot 0.4} < \sqrt{0.6 \cdot 0.6}$, $(0.8 + 0.4)/2 = (0.6 + 0.6)/2$. G-mean and FPR measure the ability of a decoder of estimating the correct state at each point in time; therefore, they are indicative of the decoder output stability (Figure 1).

Models are evaluated on all the trials from all the subjects according to their TPR and FPR (see Table 1), in order to verify that the baseline classification performance is in agreement with previous studies of asynchronous BCI, and to verify that decoding performance is similar across motor imagery tasks (i.e., foot, left and right hand motor imagery).

Not every subject is able to produce an ERD at every trial, due to the so called BCI illiteracy (Vidaurre and Blankertz, 2010) and to the fact that sensorimotor rhythms are noisy signals. This means that several trials do not contain an ERD at all. In such cases, it is uninformative to compare different models, since all will fail due to the absence of an ERD. Therefore, to obtain a meaningful comparison across models, it is necessary to remove such trials. Here, a trial with no ERD is defined as a trial where neither the MSM nor the best sliding window classifier (i.e., LR 1.0 s) are able to produce a sufficiently accurate and stable output. Based on this assumption, a trial with a G-mean smaller than 0.6 is considered as a trial with no ERD. This threshold is a compromise between chance level (i.e., 0.5) and the threshold commonly accepted in BCI (i.e., 0.7) as the minimum needed for communication (Nijboer et al., 2008). Subjects who could not produce an ERD in at least 25% of the trials are removed from the analysis. This threshold was chosen based on the fact that one trial contains two classes. Therefore, the probability of a successful trial by chance is 25%: the combination of majority of true negatives (TN) and majority of true positives (TP) against all the other 3 combinations (e.g., majority of false negatives (FN) and majority of TN). Based on this approach, 2 out of 50 subjects were removed. The percentage of successful trials for the retained subjects was on average 58 ± 18 (SD) (chance level at 25%).

2.5.3. Statistical Testing

For each of the three performance measures (i.e., latency, FPR and G-mean) we compare MSM with the sliding window classifiers (i.e., LR 1.0 s, LR 0.5 s, LR 0.1 s) by Analysis of variance (ANOVA) with a linear mixed-effects model (Bates et al., 2015), specified by the formula, in Wilkinson notation:

$$\text{Measure} = \text{Decoder} + (1 + \text{Decoder} | \text{Subject}) + (1 | \text{Subject} : \text{Trial})$$

For each measure (i.e., latency, FPR or G-mean), one new model is inferred. Therefore, the dependent variable Measure is either

TABLE 1 | Mean and standard deviation, of true positive rate (TPR) and false positive rate (FPR) of each model, computed across across 50 subjects and including all the trials, i.e., before removing trials with no ERD.

	MSM	LR1.0	LR0.5	LR0.1
TPR	0.64 ± 0.14	0.62 ± 0.11	0.60 ± 0.10	0.55 ± 0.08
FPR	0.37 ± 0.11	0.37 ± 0.10	0.38 ± 0.08	0.41 ± 0.06

latency, FPR or G-mean. The fixed-effect predictor Decoder represents the type of decoder (i.e., MSM, LR 1.0 s, LR 0.5 s, LR 0.1 s). The term (1 + Decoder | Subject) represents a random intercept and random slope for Decoder, which allows for different random relationships among the decoders for each subject. The term (1 | Subject : Trial) allows for different random intercepts for each trial of each subject. Tukey's all-pair comparisons *post-hoc* tests are carried out by z-test, corrected for multiple comparisons by the Holm-Bonferroni procedure (Bretz et al., 2010).

3. RESULTS

The true positive rate (TPR) and false positive rate (FPR) of the models before removing trials with no ERD are shown in Table 1. This allows us to verify whether the baseline classification performance is in agreement with previous studies of asynchronous BCI. The performance measures of MSM and LR 1.0 s are approximately 10% lower than the results reported in Townsend et al. (2004) in which 3 healthy subjects, achieved

TABLE 2 | Result of multiple comparisons.

	Latency		
	Estimate	z-value	Pr(> z)
LR1.0 - MSM	0.155	4.48	2.22×10^{-5}
LR0.5 - MSM	-0.0908	-2.46	0.014
LR0.1 - MSM	-0.152	-5.29	4.96×10^{-7}
LR0.5 - LR1.0	-0.246	-15.1	0
LR0.1 - LR1.0	-0.307	-12.6	0
LR0.1 - LR0.5	-0.0616	-3.63	0.000561
	False positive rate (FPR)		
	Estimate	z-value	Pr(> z)
LR1.0 - MSM	0.0011	0.0726	0.942
LR0.5 - MSM	0.0419	2.78	0.0108
LR0.1 - MSM	0.0984	6.43	3.79×10^{-10}
LR0.5 - LR1.0	0.0408	11.1	0
LR0.1 - LR1.0	0.0973	16.3	0
LR0.1 - LR0.5	0.0565	11.5	0
	G-mean(TPR,TNR)		
	Estimate	z-value	Pr(> z)
LR1.0 - MSM	-0.0144	-1.77	0.0774
LR0.5 - MSM	-0.0508	-5.88	8.21×10^{-9}
LR0.1 - MSM	-0.11	-13.7	0
LR0.5 - LR1.0	-0.0365	-12.1	0
LR0.1 - LR1.0	-0.0958	-23.4	0
LR0.1 - LR0.5	-0.0594	-17.7	0

Each subtable contains the results of latency, FPR and G-mean, respectively. Each row contains one pairwise comparison between two models. The column Estimate contains the estimated average difference between the two models being compared. P-values below 2×10^{-16} are assigned $P = 0$.

on average an FPR of 0.27 and a TPR of 0.74. This may be due to the fact that while the 3 subjects of Townsend et al. (2004) were all familiar with the BCI, in our experiment 49 out of 50 subjects had never experienced a motor imagery BCI task. Thus, a consistent portion of these naive subjects may have been labeled as BCI illiterate (Vidaurre and Blankertz, 2010). Indeed, our averaged performance is in line with the least performing subject in Townsend et al. (2004). Moreover, before removing trials with no ERD, we verify that MSM decoding performance is similar across motor imagery tasks (i.e., foot, left and right hand motor imagery) as measured by G-mean: 0.66 ± 0.07 for foot, 0.61 ± 0.08 for left hand and 0.64 ± 0.07 for right hand. Such similarity across tasks is confirmed, for every decoder, by the large p -values ($p > 0.1$) of a one-way ANOVA with G-mean as response variable and motor imagery tasks as groups.

The linear mixed-effects ANOVA tests, computed on the trials with ERD, showed that the effect of the type of decoder (i.e., MSM, LR 1.0 s, LR 0.5 s, LR 0.1 s) was significant on latency [$F_{(3)} = 79.09$, $P < 2.2 \times 10^{-16}$], on FPR [$F_{(3)} = 92.99$, $P < 2.2 \times 10^{-16}$] and on G-mean [$F_{(3)} = 212.36$, $P < 2.2 \times 10^{-16}$]. The magnitude and significance of all the pairwise comparisons are shown in Table 2.

The main result is that MSM achieves the best tradeoff between latency, FPR and G-mean, as depicted in Figure 5. The latency of MSM is on average 155 ms shorter than that of LR 1.0 s, with significance at $P = 2.2 \times 10^{-5}$; while FPR and G-mean are not significantly different. Indeed, the estimated averages show that MSM improves G-mean of only 1.4% with significance at $P = 0.077$, and increases FPR of only 0.1% with significance at $P = 0.94$. Moreover, LR 0.5 s and LR 0.1 s achieve significantly better latency but significantly worse FPR and G-mean.

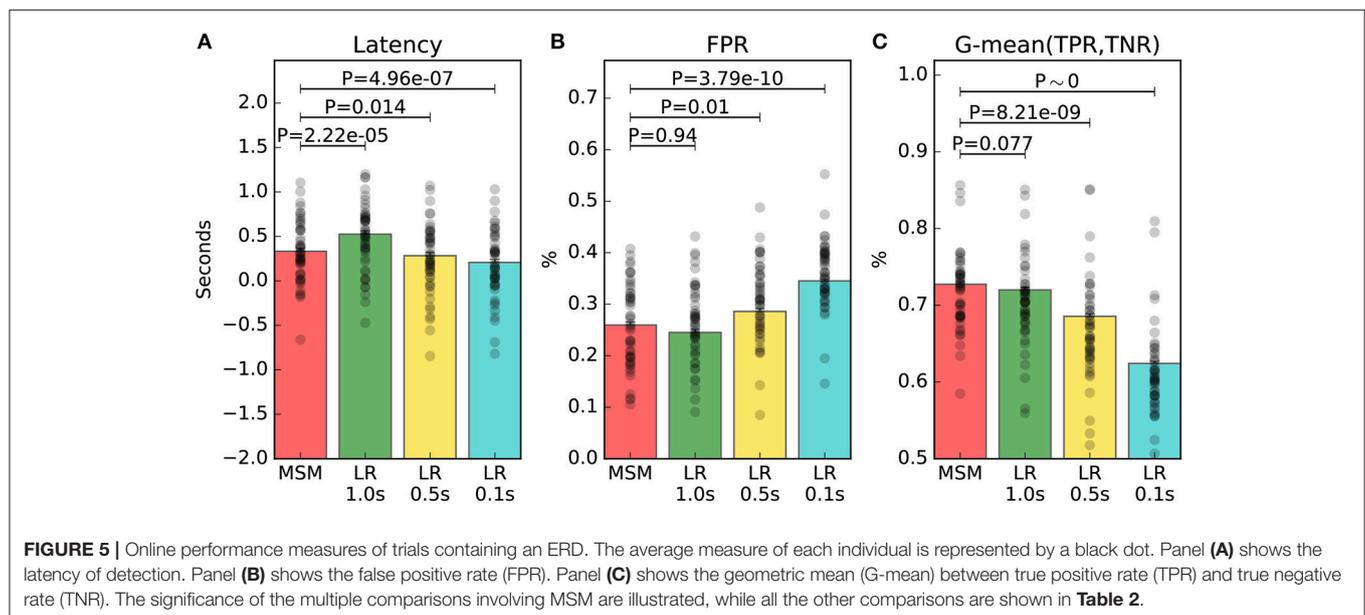
We appreciate the importance of achieving the best tradeoff between latency, FPR and G-mean, from Figures 6, 7. The former depicts the same left motor imagery trial as in Figure 1, while the latter shows another example from a different subject performing foot motor imagery. In Figure 6 we observe how MSM detects

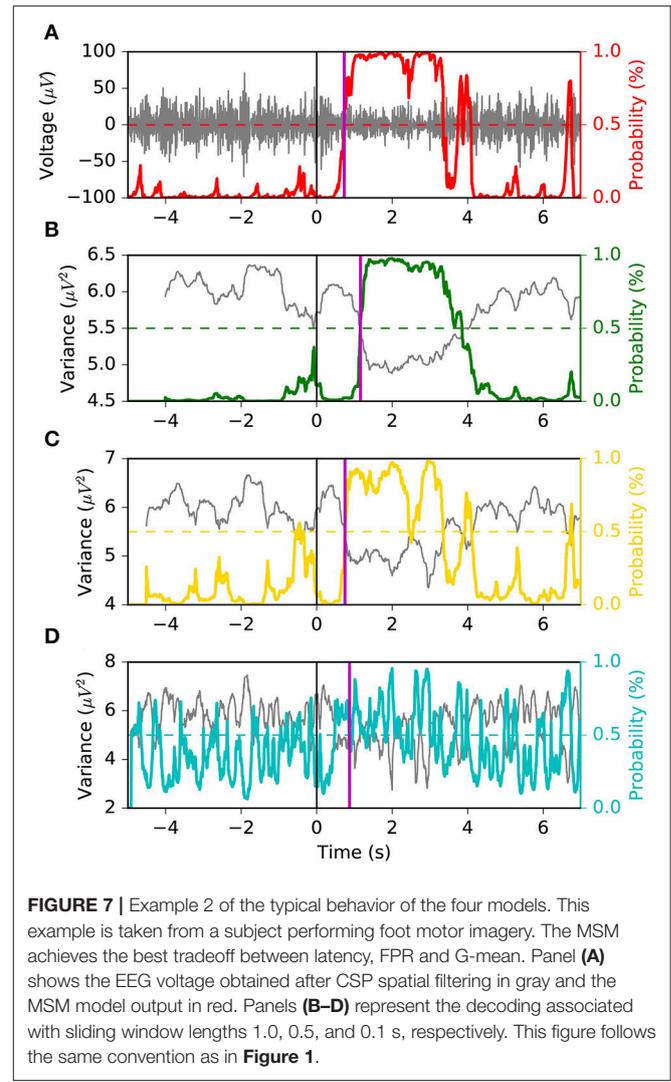
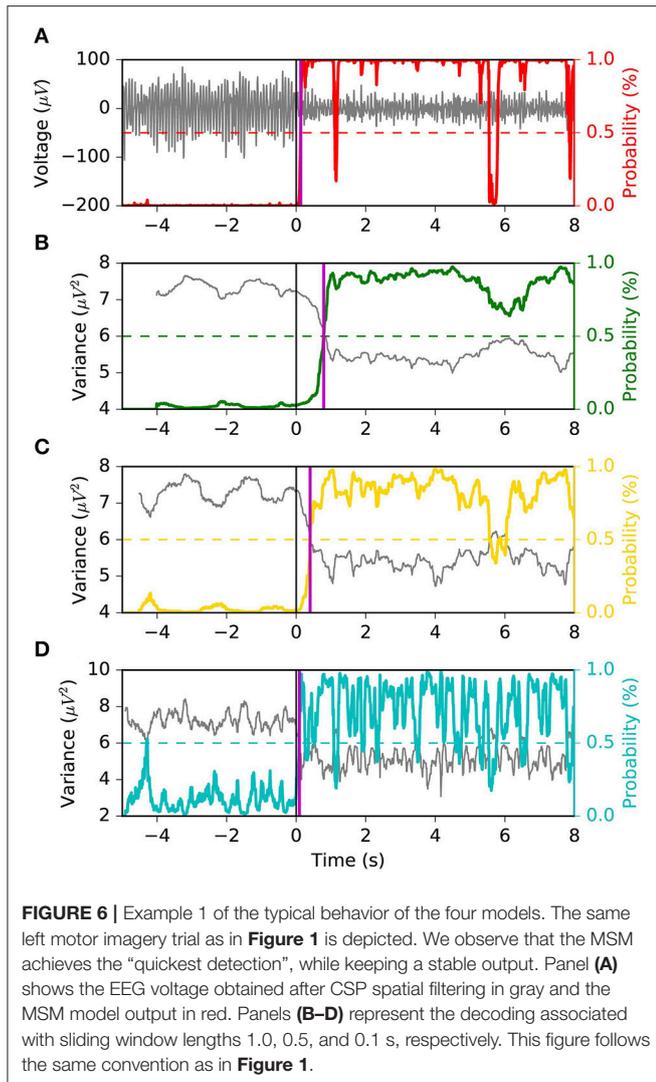
the ERD faster than LR 1.0 s, LR 0.5 s and as quick as LR 0.1 s, while keeping a stable output. In the example of Figure 7, when comparing MSM with LR 1.0 s we observe that latency is shorter, FPR is equivalent and G-mean is slightly larger due to more true positives. On the other hand, MSM detects the ERD as quick as LR 0.5 s, but with less false positives and more stability. Also, we observe how unstable the output of LR 0.1 s can be.

4. DISCUSSION

We showed that, in asynchronous motor imagery detection, the proposed MSM achieves the best tradeoff between latency of detection, false positive rate and G-mean of true positive rate and true negative rate. This translates into a quicker motor imagery detection (i.e., latency) at no cost of output stability (i.e., FPR and G-mean), compared with sliding window classifiers. On average, the proposed MSM was able to detect an ERD 155 ms faster than the most performant and commonly used sliding window approach (i.e., LR 1.0 s). The main reason for such improvement is due to the fact that the ERD is a phenomenon that occurs abruptly, and MSM have been historically used in applications where the latent state of a system changes suddenly (Hamilton, 2010).

Given the relatively low signal-to-noise ratio of the EEG signal, the most realistic application of the proposed model is BCI-based neurorehabilitation, where proprioceptive feedback is provided by a neuroprosthesis, based on the patient's brain signal. The control of only one degree of freedom is typically required, and false negatives or positives do not have the potential of becoming disastrous. In this context, the short latency of detection and stable output of the MSM allows for a more reactive proprioceptive feedback and for a reduced delay of contingent feedback. This is a desirable property, since it is hypothesized that simultaneous contingent association between volitionally evoked SMR and proprioceptive feedback may lead to





superior associative learning and elicit motor learning (Ramos-Murguialday et al., 2012, 2013).

It should be noted that the ERD does not always occur at the cue onset, but on average it happens a few tenths of a second later. However, the cue onset is the only ground truth available. As a result, the average latencies shown in **Figure 5A** include the intrinsic delay of ERD occurrence with respect to cue onset. Such delay of ERD occurrence cannot be improved, and should not be taken into account when evaluating the models. Indeed, the relative difference among models’ latencies is the most important aspect. In this context, the MSM is able to detect the ERD 155 ms earlier than *LR* 1.0 s, meaning that the time required between the ERD occurrence (not necessarily at time 0 s) and the detection is 155 ms shorter, on average. In addition, associative learning is hypothesized to occur with respect to the ERD onset, not the cue onset.

Whether a latency reduction of 155 ms is crucial for improving temporal association, inducing neural plasticity and restoring function is yet to be confirmed. It has been hypothesized that the maximum proprioceptive feedback delay still inducing

co-activation in Hebbian plasticity should be 200 or 300 ms, but this remains an open question that needs to be addressed by the community (Grosse-Wentrup et al., 2011; Xu et al., 2015). Muralidharan et al. (2011) discussed that increasing decoding accuracy at the expense of longer latencies (i.e., 200, 400, or 600 ms) would cause delayed neuroprosthetic activation and may limit therapeutic benefits. Another study (Xu et al., 2014) proposed a new method for the detection of movement-related cortical potentials, reporting a 145 ms latency reduction. They discussed that such an improvement is fundamental in order to induce neuroplastic changes in closed-loop BCIs, since the temporal association between movement-related brain signals and the afferent input is crucial for plasticity (Mrachacz-Kersting et al., 2012).

In the neurorehabilitation field of application, a question arises: whether a neurological injury would have impact on the CSP algorithm. We expect the CSP patterns to be different between healthy subjects and patients (Lei et al., 2017). However, previous reports of CSP-based BCI in clinical trials have shown that the algorithm successfully finds the most discriminant spatial filters in stroke patients (Ang et al., 2015). Therefore, retraining

the CSP filters individually for each subject should be sufficient to discover subject-specific patterns that are customized according to the brain injury.

As for the limitations of the MSM, the computation of the posterior probability has a time complexity of $O(nm^2)$, where m is the number of latent variables, and n is the length of the sequence of the observed variable. This procedure has to be repeated for each training trial and each iteration during the optimization of the MSM's transition matrix. On the other hand, the logistic regression model is not time dependent, therefore the samples of log-transformed variance are only computed once. This results in a large difference of calibration time between the MSM and the sliding window logistic regression model: the MSM may take up to tens of seconds to converge, while training the logistic regression classifier is a matter of several hundredths of a second. However, such difference in calibration time may not be prohibitive in practice, since subjects are usually asked to rest for several minutes between runs.

The MSM and sliding window classifiers achieved classification performance levels (i.e., TPR, FPR) comparable to previous studies (Townsend et al., 2004; Muralidharan et al., 2011). However, a similar comparison with previous literature is more challenging with respect to the latency of detection. This is due to the fact that, in most asynchronous ERD-based BCI studies, latency or delay are regarded as the time required to achieve the peak classification performance between two classes (i.e. left and right motor imagery) (Pfurtscheller et al., 2009).

The proposed model could be applied to phenomena other than motor imagery, such as gait related ERD (Wagner et al., 2012; Lisi and Morimoto, 2015) or processing of sensory and cognitive information (Pfurtscheller and Lopes da Silva, 1999), as long as an abrupt change in power occurs in the signal of

interest. The MSM could be extended to adaptively handle the non-stationarities of the EEG signal (Vidaurre et al., 2011), by means of Bayesian update techniques. It may also be possible to handle multi-class problems: for each additional class, a new hidden state should be included, and the Gaussian observation models should be parameterized with a covariance rather than a variance in order to handle multiple CSP projections. Given the Bayesian nature of the model, we could also include priors related to external environment, or to the current state of a neuroprosthesis in order to achieve context-dependent behavior.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: GL, AT, and JM. Performed the experiments: GL and AT. Analyzed the data: GL, DR, and AT. Contributed reagents, materials, analysis tools: GL, JM, and DR. Wrote the paper: GL. Drafted the article: GL. Revised the article: GL, DR, AT, and JM. Final approval: GL, DR, AT, and JM. Agreed to be accountable for all aspects of the work: GL, DR, AT, and JM.

FUNDING

This paper is based on results obtained from a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO) and from research commissioned by the National Institute of Information and Communications Technology (NICT), JAPAN. This work was also funded by ImPACT Program of Council for Science, Technology and Innovation (Cabinet Office, Government of Japan) and supported by the Strategic Research Program for Brain Sciences from Japan Agency for Medical Research and development, AMED.

REFERENCES

- Ang, K. K., Chin, Z. Y., Wang, C., Guan, C., and Zhang, H. (2012). Filter bank common spatial pattern algorithm on bci competition iv datasets 2a and 2b. *Front. Neurosci.* 6:39. doi: 10.3389/fnins.2012.00039
- Ang, K. K., Chua, K. S. G., Phua, K. S., Wang, C., Chin, Z. Y., Kuah, C. W. K., et al. (2015). A randomized controlled trial of eeg-based motor imagery brain-computer interface robotic rehabilitation for stroke. *Clin. EEG Neurosci.* 46, 310–320. doi: 10.1177/1550059414522229
- Ang, K. K., Guan, C., Chua, K. S. G., Ang, B. T., Kuah, C., Wang, C., et al. (2009). "A clinical study of motor imagery-based brain-computer interface for upper limb robotic rehabilitation," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009 (Minneapolis, MN: IEEE), 5981–5984. doi: 10.1109/IEMBS.2009.5335381
- Ang, K. K., Guan, C., Phua, K. S., Wang, C., Zhou, L., Tang, K. Y., et al. (2014). Brain-computer interface-based robotic end effector system for wrist and hand rehabilitation: results of a three-armed randomized controlled trial for chronic stroke. *Front. Neuroeng.* 7:30. doi: 10.3389/fneng.2014.00030
- Barandela, R., Sánchez, J. S., García, V., and Rangel, E. (2003). Strategies for learning in class imbalance problems. *Pattern Recognit.* 36, 849–851. doi: 10.1016/S0031-3203(02)00257-1
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., and Robbins, K. A. (2015). The prep pipeline: standardized preprocessing for large-scale eeg analysis. *Front. Neuroinform.* 9:16. doi: 10.3389/fninf.2015.00016
- Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., and Müller, K.-R. (2008). Optimizing spatial filters for robust eeg single-trial analysis. *Signal Process. Magazine IEEE*, 25, 41–56. doi: 10.1109/MSP.2008.4408441
- Bretz, F., Hothorn, T., and Westfall, P. (2010). *Multiple Comparisons Using R*. CRC Press. Available online at: <https://www.crcpress.com/Multiple-Comparisons-Using-R/Bretz-Hothorn-Westfall/p/book/9781584885740>
- Broetz, D., Braun, C., Weber, C., Soekadar, S. R., Caria, A., and Birbaumer, N. (2010). Combination of brain-computer interface training and goal-directed physical therapy in chronic stroke: a case report. *Neurorehabil. Neural Repair* 24, 674–679. doi: 10.1177/1545968310368683
- Caria, A., Weber, C., Brötz, D., Ramos, A., Ticini, L. F., Gharabaghi, A., et al. (2011). Chronic stroke recovery after combined bci training and physiotherapy: a case report. *Psychophysiology* 48, 578–582. doi: 10.1111/j.1469-8986.2010.01117.x
- Chiappa, S., and Bengio, S. (2003). *Hmm and Iohmm Modeling of Eeg Rhythms for Asynchronous Bci Systems*. Technical report, IDIAP.
- Cincotti, F., Scipione, A., Timperi, A., Mattia, D., Marciani, A. G., Millan, J., et al. (2003). "Comparison of different feature classifiers for brain computer interfaces," in *First International IEEE EMBS Conference on Neural Engineering*, 2003. *Conference Proceedings* (Capri Island), 645–647.
- Daly, J. J., Cheng, R., Rogers, J., Litinas, K., Hrovat, K., and Dohring, M. (2009). Feasibility of a new application of noninvasive brain computer interface (bci): a case study of training for recovery of volitional motor control after stroke. *J. Neurol. Phys. Ther.* 33, 203–211. doi: 10.1097/NPT.0b013e3181c1fc0b
- Daly, J. J., and Wolpaw, J. R. (2008). Brain-computer interfaces in neurological rehabilitation. *Lancet Neurol.* 7, 1032–1043. doi: 10.1016/S1474-4422(08)70223-0
- Do, A. H., Wang, P. T., King, C. E., Chun, S. N., and Nenadic, Z. (2013). Brain-computer interface controlled robotic gait orthosis. *J. Neuroeng. Rehabil.* 10:111. doi: 10.1186/1743-0003-10-111

- Gomez-Rodriguez, M., Peters, J., Hill, J., Schölkopf, B., Gharabaghi, A., and Grosse-Wentrup, M. (2011). Closing the sensorimotor loop: haptic feedback facilitates decoding of motor imagery. *J. Neural Eng.* 8:036005. doi: 10.1088/1741-2560/8/3/036005
- Grosse-Wentrup, M., Mattia, D., and Oweiss, K. (2011). Using brain-computer interfaces to induce neural plasticity and restore function. *J. Neural Eng.* 8:025004. doi: 10.1088/1741-2560/8/2/025004
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357–384. doi: 10.2307/1912559
- Hamilton, J. D. (2010). “Regime switching models,” in *Macroeconometrics and Time Series Analysis*, eds S. N. Durlauf and L. E. Blume (London: Palgrave Macmillan UK), 202–209. doi: 10.1057/9780230280830_23
- Iglewicz, B., and Hoaglin, D. C. (1993). *How to Detect and Handle Outliers*. Milwaukee, WI: ASQC Quality Press. Available online at: <https://trove.nla.gov.au/work/10589055?q&versionId=12350384>
- Kraft, D. (1988). *A Software Package for Sequential Quadratic Programming*. Forschungsbericht- Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt.
- Lei, X., Wang, L., Kong, W., Peng, Y., Hu, S., Zeng, H., et al. (2017). “Identification of eeg features in stroke patients based on common spatial pattern and sparse representation classification,” in *2017 8th International IEEE/EMBS Conference on Neural Engineering (NER)* (Shanghai), 114–117.
- Lisi, G., Hamaya, M., Noda, T., and Morimoto, J. (2016). “Dry-wireless eeg and asynchronous adaptive feature extraction towards a plug-and-play co-adaptive brain robot interface,” in *2016 IEEE International Conference on Robotics and Automation (ICRA)* (Stockholm), 959–966.
- Lisi, G., and Morimoto, J. (2015). Eeg single-trial detection of gait speed changes during treadmill walk. *PLoS ONE* 10:e0125479. doi: 10.1371/journal.pone.0125479
- Lisi, G., and Morimoto, J. (2017). “Noninvasive brain machine interfaces for assistive and rehabilitation robotics: a review” in *Human Modelling for Bio-Inspired Robotics*, eds J. Ueda and Y. Kurita (Academic Press), 187–216. doi: 10.1016/B978-0-12-803137-7.00006-9
- Lisi, G., Noda, T., and Morimoto, J. (2014). Decoding the ERD/ERS: influence of afferent input induced by a leg assistive robot. *Front. Syst. Neurosci.* 8:85. doi: 10.3389/fnsys.2014.00085
- Lotte, F., Congedo, A., Lécuyer, A., Lamarche, F., and Arnaldi, B. (2007). A review of classification algorithms for eeg-based brain-computer interfaces. *J. Neural Eng.* 4:R1. doi: 10.1088/1741-2560/4/2/R01
- Mrachacz-Kersting, N., Kristensen, S. R., Niazi, I. K., and Farina, D. (2012). Precise temporal association between cortical potentials evoked by motor imagination and afference induces cortical plasticity. *J. Physiol.* 590, 1669–1682. doi: 10.1113/jphysiol.2011.222851
- Müller-Putz, G. R., Scherer, R., Pfurtscheller, G., and Rupp, R. (2005). EEG-based neuroprosthesis control: a step towards clinical practice. *Neurosci. Lett.* 382, 169–174. doi: 10.1016/j.neulet.2005.03.021
- Muralidharan, A., Chae, J., and Taylor, D. M. (2011). Extracting attempted hand movements from eegs in people with complete hand paralysis following stroke. *Front. Neurosci.* 5:39. doi: 10.3389/fnins.2011.00039
- Murphy, K. P. (1998). *Switching Kalman Filters*. Technical report, Citeseer.
- Naros, G., and Gharabaghi, A. (2015). Reinforcement learning of self-regulated β -oscillations for motor restoration in chronic stroke. *Front. Hum. Neurosci.* 9:391. doi: 10.3389/fnhum.2015.00391
- Nijboer, F., Sellers, E., Mellinger, J., Jordan, M. A., Matuz, T., Furdea, A., et al. (2008). A p300-based brain-computer interface for people with amyotrophic lateral sclerosis. *Clin. Neurophysiol.* 119, 1909–1916. doi: 10.1016/j.clinph.2008.03.034
- Obermaier, B., Guger, C., Neuper, C., and Pfurtscheller, G. (2001). Hidden markov models for online classification of single trial eeg data. *Pattern Recogn. Lett.* 22, 1299–1309. doi: 10.1016/S0167-8655(01)00075-7
- Pfurtscheller, G., Linortner, P., Winkler, R., Korisek, G., and Müller-Putz, G. (2009). Discrimination of motor imagery-induced eeg patterns in patients with complete spinal cord injury. *Comput. Intel. Neurosci.* 2009:104180. doi: 10.1155/2009/104180
- Pfurtscheller, G., and Lopes da Silva, F. H. (1999). Event-related eeg/meg synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* 110, 1842–1857. doi: 10.1016/S1388-2457(99)00141-8
- Poor, H. V., and Hadjiladis, O. (2009). *Quickest Detection*. Cambridge: Cambridge University Press.
- Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proc. IEEE* 77, 257–286. doi: 10.1109/5.18626
- Ramos-Murguialday, A., Broetz, D., Rea, M., Lärer, L., Yilmaz, O., Brasil, F. L., et al. (2013). Brain-machine interface in chronic stroke rehabilitation: a controlled study. *Ann. Neurol.* 74, 100–108. doi: 10.1002/ana.23879
- Ramos-Murguialday, A., Schürholz, M., Caggiano, V., Wildgruber, M., Caria, A., Hammer, E. M., et al. (2012). Proprioceptive feedback and brain computer interface (bci) based neuroprostheses. *PLoS ONE* 7:e47048. doi: 10.1371/journal.pone.0047048
- Rezaei, S., Tavakolian, K., Nasrabadi, A. M., and Setarehdan, S. K. (2006). Different classification techniques considering brain computer interface applications. *J. Neural Eng.* 3, 139–144. doi: 10.1088/1741-2560/3/2/008
- Santaniello, S., Burns, S. P., Golby, A. J., Singer, J. M., Anderson, W. S., and Sarma, S. V. (2011). Quickest detection of drug-resistant seizures: an optimal control approach. *Epilep. Behav.* 22(Suppl. 1), S49–S60. doi: 10.1016/j.yebeh.2011.08.041
- Sarac, M., Koyas, E., Erdogan, A., Cetin, M., and Patoglu, V. (2013). “Brain computer interface based robotic rehabilitation with online modification of task speed,” in *2013 IEEE International Conference on Rehabilitation Robotics (ICORR)* (Seattle, WA), 1–7.
- Shindo, K., Kawashima, K., Ushiba, J., Ota, N., Ito, M., Ota, T., et al. (2011). Effects of neurofeedback training with an electroencephalogram-based brain-computer interface for hand paralysis in patients with chronic stroke: a preliminary case series study. *J. Rehabil. Med.* 43, 951–957. doi: 10.2340/16501977-0859
- Silvoni, S., Ramos-Murguialday, A., Cavinato, M., Volpato, C., Cisotto, G., Turolla, A., et al. (2011). Brain-computer interface in stroke: a review of progress. *Clin. EEG Neurosci.* 42, 245–252. doi: 10.1177/155005941104200410
- Townsend, G., Graimann, B., and Pfurtscheller, G. (2004). Continuous eeg classification during motor imagery-simulation of an asynchronous bci. *IEEE Trans. Neural Syst. Rehabil. Eng.* 12, 258–265. doi: 10.1109/TNSRE.2004.827220
- Turner, C. M., Startz, R., and Nelson, C. R. (1989). A markov model of heteroskedasticity, risk, and learning in the stock market. *J. Financ. Econ.* 25, 3–22. doi: 10.1016/0304-405X(89)90094-9
- Vidaurre, C., and Blankertz, B. (2010). Towards a cure for bci illiteracy. *Brain Topogr.* 23, 194–198. doi: 10.1007/s10548-009-0121-6
- Vidaurre, C., Kawanabe, M., von Büna, P., Blankertz, B., and Müller, K. (2011). Toward unsupervised adaptation of lda for brain-computer interfaces. *Biomed. Eng. IEEE Trans.* 58, 587–597. doi: 10.1109/TBME.2010.2093133
- Wagner, J., Solis-Escalante, T., Grieshofer, P., Neuper, C., Müller-Putz, G., and Scherer, R. (2012). Level of participation in robotic-assisted treadmill walking modulates midline sensorimotor EEG rhythms in able-bodied subjects. *Neuroimage* 63, 1203–1211. doi: 10.1016/j.neuroimage.2012.08.019
- Xu, R., Jiang, N., Lin, C., Mrachacz-Kersting, N., Dremstrup, K., and Farina, D. (2014). Enhanced low-latency detection of motor intention from eeg for closed-loop brain-computer interface applications. *IEEE Trans. Biomed. Eng.* 61, 288–296. doi: 10.1109/TBME.2013.2294203
- Xu, R., Jiang, N., Mrachacz-Kersting, N., Dremstrup, K., and Farina, D. (2015). Factors of influence on the performance of a short-latency non-invasive brain switch: evidence in healthy individuals and implication for motor function rehabilitation. *Front. Neurosci.* 9:527. doi: 10.3389/fnins.2015.00527
- Yong, X., Ward, R. K., and Birch, G. E. (2008). “Robust common spatial patterns for EEG signal preprocessing,” in *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008* (Vancouver, BC: IEEE), 2087–2090. doi: 10.1109/IEMBS.2008.4649604

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

Copyright © 2018 Lisi, Rivela, Takai and Morimoto. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.