Spectrally Adaptive Common Spatial Patterns

Anonymous Author(s)
Affiliation
Address
email

Abstract

The method of Common Spatial Patterns (CSP) is widely used for feature extraction of electroencephalography (EEG) data such as in motor imagery Brain-computer Interface (BCI) systems. It is a data-driven method estimating a set of spatial filters so that the power of the filtered EEG data is maximally separated between imagery classes. This method, however, is prone to overfitting and is known to suffer from poor generalization especially with limited calibration data. On the other hand, due to the high heterogeneity in brain data and the non-stationarity of brain activity, CSP is usually trained for each user separately resulting in long calibration sessions or frequent re-calibrations that are tiring for the user. In this work, we propose a novel algorithm called *Spectrally Adaptive Common Spatial Patterns* (SACSP) that improves CSP by learning a temporal/spectral filter for each spatial filter so that the spatial filters are concentrated on the most relevant temporal frequencies. We show the efficacy of SACSP in motor imagery BCI in providing better generalizability and higher classification accuracy from calibration to online control compared to existing methods while providing neurophysiologically relevant information about the temporal frequencies of the filtered signals.

The method of Common Spatial Patterns (CSP) [1] is among the most commonly used methods for feature extraction in electroencephalography (EEG) data [2, 3]. One of its main applications is to learn relevant features in motor imagery brain-computer interface (MI-BCI) systems [4]. Imagined movement, similar to actual movement, results in a decrease in power in the mu (7-13 Hz) and often beta (13-30Hz) frequency bands over the contralateral sensorimotor cortex. CSP is a data-driven supervised approach that learns a set of spatial filters to maximally discriminate the two classes. That is, the power of the projected EEG data through the selected filters is maximally different for the two imagery classes [5]. CSP counters the spatial blurring caused by measuring electrical activity through bone and several layers of tissue [5]. However, CSP suffers from overfitting and poor generalizability from calibration to online control [6].

After learning the feature space (logarithm of the power through the selected filters), a classifier – commonly linear discriminant analysis (LDA) [5] – is trained on the selected features.

Due to the difficulty of the task and differences between people, there is currently no universal motor-imagery BCI classifier, and at least a short calibration session (usually supervised) for every user is required to achieve reliable control. During the calibration session, the user is instructed to imagine the movement of different body parts (e.g., the right and the left hand) and is provided with no feedback or sham feedback [2]. The recorded EEG data during this calibration session – calibration data – is then used to train the classifier (CSP filters + linear classifier) and deployed for real-time online control of the BCI.

In this work, we propose a novel approach to learn spatial and spectral filters in motor imagery data by solving an optimization problem that generalizes the cost function in CCACSP [7]. We use an alternating direction method [8] to optimize our proposed cost function. We apply our method to motor imagery data that we recorded from 12 participants and show improved classification accuracy

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.

as well as more interpretable spatial patterns and spectral filters compared to the closely related methods of CSP [1], CCACSP [7] and spec-CSP [9].

CSP estimates the spatial covariance for the two classes and uses a Raleigh quotient cost function to find the optimal directions for such projection [1, 5]. In SACSP [1] however, we define the spectrally-weighted spatial covariance instead, to weight the motor imagery related frequencies.

Let $X_i \in \mathbb{R}^{n \times t}$ represent the bandpassed EEG data epoch over n channels with t time samples that belongs to class $i \in \{1, 2\}$, e.g. the two motor imagery classes. Since EEG data is bandpassed, $\mathbb{E}[X_1] = \mathbb{E}[X_2] = 0$. Let C be a circulant matrix. We define the spectrally-weighted spatial covariance matrices for classes 1 and 2 as $\Gamma_{S1} = \mathbb{E}[X_1 C X_1^T]$ and $\Gamma_{S2} = \mathbb{E}[X_2 C X_2^T]$, where $(\cdot)^T$ indicates the transpose of a matrix/vector. As C is circulant, we can write it as $C = F\text{diag}(h)F^H$ where h is a vector, $diag(h)$ is a matrix whose diagonal elements are the elements of the vector h, F is the $t \times t$ Fourier matrix and $(\cdot)^H$ indicates the complex conjugate transpose of a matrix. Replacing this in the above and defining $\hat{X}_1 = X_1 F$ and $\hat{X}_2 = X_2 F$ as the Fourier transform of an epoch along its temporal dimension we can write the following set of optimization problems:

$$\underset{w,h}{\arg\max} \quad \frac{w^T \mathbb{E}[\hat{X}_1 \text{diag}(h) \hat{X}_1^H] w}{w^T (\Sigma_1 + \Sigma_2) w}, \qquad \underset{v,l}{\arg\max} \quad \frac{v^T \mathbb{E}[\hat{X}_2 \text{diag}(l) \hat{X}_2^H] v}{v^T (\Sigma_1 + \Sigma_2) v}, \tag{1}$$

where $\Sigma_i = \mathbb{E}[X_i X_i^T]$ is the spatial covariance of each class. We solve the two optimization problems in (1) via an alternating maximization procedure.

Classification: After finding the optimal $2 \times R$ spatial and spectral filter pairs (f_{wv}, f_{hl}), then $2 \times R$ features were extracted by filtering each epoch, x as follows:

$$\log(f_{wv}^T \, (\hat{x} \, \text{diag}(f_{hl}) \hat{x}^H) \, f_{wv}), \tag{2}$$

where $\hat{x} = xF$. Next, linear discriminant analysis (LDA) [10] with automatic shrinkage using the Ledoit-Wolf lemma [11] was trained as the classifier.

Data: Participants were instructed to use movement imagination of their right or left hand to move a cursor on the screen. During the first three blocks, the cursor moved according to a pre-determined pseudo-randomly generated sequence of movements but the participants were not aware of this and were led to believe that they were in control of the cursor movements. The recorded EEG data from the first three blocks were used for calibration. CSP was trained for feature extraction where the top three CSP filters for each class were selected. The logarithm of the power of the CSP-filtered EEG data were selected as features and shrinkage-LDA [11] was trained for classification. This classifier was used during the online control of the cursor in the latter 6 blocks which had a similar structure as the calibration data except that the cursor moved based on the participant's motor imagery data.

In offline analysis, we used the data from the calibration blocks to train CSP, CCACSP, spec-CSP and SACSP separately. Each method was then tested on data recorded during online control blocks.

Pre-processing: Data from calibration and online blocks were downsampled to 100 Hz and epoched 0-1 second with respect to each cursor movement excluding the last. After downsampling and epoching, data were re-referenced to the common average and bandpass filtered in 7–30 Hz to include the mu (about 7–13 Hz) and beta (about 13–30 Hz) frequency bands [12, 13].

We balanced the right and left classes (including movements towards and away from the target in each class) for both calibration and online blocks to alleviate any potential bias in learning features and for easier interpretation of the results. Balancing the classes was done 10 times to allow steps to be in the training (from calibration data) or test (from online data) sets.

Classification accuracy: The classification accuracy of the CSP, CCACSP, spec-CSP and SACSP when the feature extraction and classifiers were trained on the calibration data and tested on the online data, are presented in table 1. The first number in each entry for participants 1 to 12 (P1-P12), represents the mean classification accuracy over 10 instances of train-test data and the second number

[1] We provide the anonymized and epoched EEG data at: `https://github.com/anonymsquirrel/SACSP`. We release our developed code under the MIT license.

Table 1: Motor imagery (right/left hand) classification accuracy. For P1-P12, the first number in each entry reports the average accuracy across 10 instances of the data and the second number reports the standard deviation of the mean. In the last row, the first number in each entry reports the average classification accuracy across participants for the corresponding column and the second number reports the standard error of the mean.

PID	CSP	spec-CSP	CCACSP	SACSP
P1	0.55/0.035	0.61/0.039	0.74/0.033	0.81/0.022
P2	0.56/0.028	0.63/0.091	0.69/0.012	0.69/0.030
P3	0.63/0.038	0.61/0.037	0.61/0.057	0.72/0.096
P4	0.72/0.019	0.70/0.051	0.71/0.020	0.73/0.009
P5	0.59/0.118	0.59/0.105	0.56/0.071	0.62/0.117
P6	0.50/0.014	0.52/0.032	0.54/0.017	0.52/0.013
P7	0.62/0.027	0.65/0.035	0.68/0.019	0.70/0.040
P8	0.57/0.023	0.59/0.034	0.71/0.023	0.72/0.036
P9	0.68/0.048	0.67/0.033	0.73/0.019	0.76/0.023
P10	0.49/0.021	0.49/0.024	0.49/0.017	0.50/0.019
P11	0.60/0.019	0.64/0.028	0.64/0.034	0.67/0.027
P12	0.47/0.032	0.45/0.025	0.49/0.024	0.51/0.010
Average	0.58/0.022	0.60/0.022	0.63/0.027	**0.66/0.030**

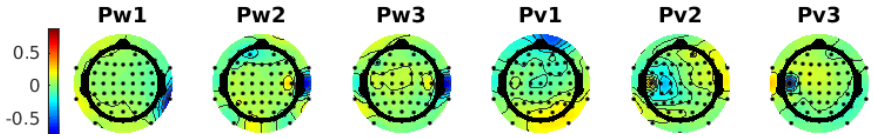
reports the standard deviation. In the last row, the average performance across participants is reported where in each entry, the first number indicates the average for the corresponding column and the second number indicates the standard error of the mean. Across participants, SACSP performs significantly better compared to CCACSP, spec-CSP and CSP (Wilcoxon signed rank test, $p < 0.01$).

We also looked at the classification accuracy when each method was trained on calibration data and tested on calibration data through cross-validation (we call it 'calibCV' case) as well as when trained on online data and tested on online data through cross-validation (we call it 'onlineCV' case). Note that both of these conditions are different from the practical scenario of interest in this work, where the methods are trained on the calibration data and tested on the online data (we call it 'test' case). Our results show that for CSP, calibCV and onlineCV are not different across participants (average 0.73 and 0.71, Wilcoxon signed-rank test, $p = 0.42$) However, as table 1 shows, there is a large drop in performance of CSP in the test condition (across participants significantly lower that calibCV and onlineCV, Wilcoxon signed rank test, $p < 0.001$). Our results show that spec-CSP, CCACSP, and SACSP were all able to close this gap from calibration to online performance with SACSP performing significantly better than all other methods (Wilcoxon signed rank test, $p < 0.01$).

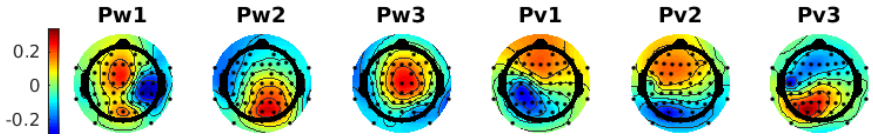
Visualization: We visualize the optimal spatial patterns and spectral filters to compare their neurophysiological interpretation. Figure 1 compares the spatial patterns and spectral filters of the SACSP with CSP, CCACSP, and spec-CSP. Note that these are the filters and patterns that were trained on the calibration data and used later for classification of the online data. The last two rows show the results for SACSP. The filters and patterns for each class (Pw_1, Pw_2 and Pw_3, h_1, h_2 and h_3) and (Pv_1, Pv_2 and Pv_3, l_1, l_2 and l_3) converge to similar values. This is because our algorithm, different from the others presented here, does not result in spatial filters/patterns that are necessarily orthogonal to each other and convergence to the same solution from different initializations in fact shows that SACSP converged to its optimal solution. Furthermore, we tested SACSP with random initialization and received the same filter and patterns.

Comparing the spectral filters for spec-CSP and SACSP, we can see that the spec-CSP spectral filters fail at selectively extracting the underlying features in an interpretable way. SACSP shows smoother and less noisy filters with a peak at/around the mu band for both the right and left hand motor imagery classes and a small beta peak for the left hand motor imagery class.

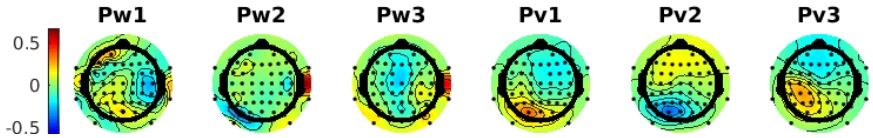
SACSP shows patterns typical for right and left hand motor imagery. Pv_3 in spec-CSP can be regarded as a similar pattern to Pv_3 in SACSP but the rest of the patterns in spec-CSP and those in CSP are neurophysiologically implausible for motor imagery. Pv_1 and Pv_3 in CCACSP are motor imagery patterns. However, CCACP fails at finding appropriate patterns for the other class.



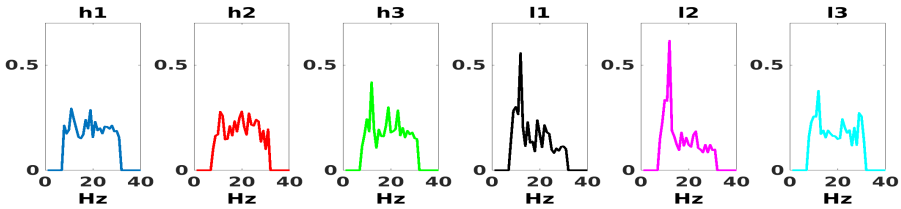
(a) P1 – CSP spatial patterns.



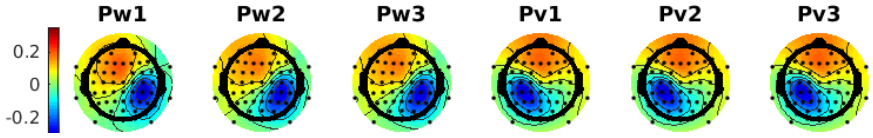
(b) P1 – CCACSP spatial patterns.



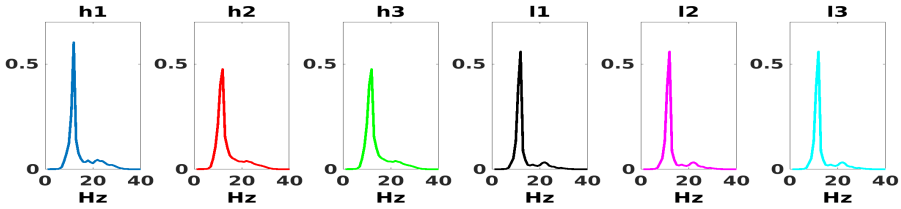
(c) P1 – spec-CSP spatial patterns.



(d) P1 – spec-CSP spectral filters.



(e) P1 – SACSP spatial patterns.



(f) P1 – SACSP spectral filters.

Figure 1: Spatial patterns and spectral filters trained on the calibration data for P1. Compared to other methods, SACSP results in spatial patterns that are typical for right and left hand motor imagery while providing neurophysiologically relevant frequency selectivity in spectral filters. Normalized spectral filters were plotted against frequency (Hz) as line plots using MATLAB. Spatial patterns were estimated for each method accordingly and `topoplot` from EEGLAB was used to plot them as heatmap on a 2-D scalp.

Supplemental Information - Full Paper

1 Introduction

The method of Common Spatial Patterns (CSP) [1] is among the most commonly used methods for feature extraction in electroencephalography (EEG) data [2, 3]. One of its main applications is to learn relevant features in motor imagery brain-computer interface (MI-BCI) systems [4]. To operate an MI-BCI, users imagine movement of different parts of their body – for instance, their right or left hand – without actually moving them and the goal of the BCI is to infer the user's intended imagery class from the EEG data [12]. Imagined movement, similar to actual movement, results in a decrease in power in the mu (7-13 Hz) and often beta (13-30Hz) frequency bands over the contralateral sensorimotor cortex. However, as movement imagination is self-generated with varying success and there are differences in cortical folding patterns across people, there is a large amount of heterogeneity in the motor imagery signal between BCI users, including differences in the precise temporal frequencies that are modulated. Furthermore, as EEG data is recorded non-invasively through electrodes that are located on the scalp several tissue layers away from the neural activity, the recorded EEG is spatially blurred and can be very similar between the different motor imagery classes while differing significantly among people [14].

CSP is a data-driven supervised approach that learns a set of spatial filters to maximally discriminate the two classes. That is, the power of the projected EEG data through the selected filters is maximally different for the two imagery classes [5]. After learning the feature space (logarithm of the power through the selected filters), a classifier – commonly a linear one such as a linear discriminant analysis (LDA) [5] – is trained on the selected features.

Due to the difficulty of the task and the above mentioned differences between people, there is currently no universal motor-imagery BCI classifier, and at least a short calibration session (usually supervised) for every user is required to achieve reliable control. In the case of a binary BCI, during the calibration session, the user is instructed to imagine the movement of two body parts (e.g., the right and the left hand) and is provided with no feedback or sham feedback [2]. The recorded EEG data during this calibration session – calibration data – is then used to train the classifier (CSP filters + linear classifier) and deployed for real-time online control of the BCI. Movement imagination of different body parts is usually translated to real-world application, for instance switching a light on/off, moving a wheelchair or a robotic arm, etc.

CSP extracts the class-correlated spatial features and acts as supervised dimensionality reduction that also reduces the inherent noise in the EEG data and counters the spatial blurring caused by measuring electrical activity through bone and several layers of tissue [5]. The corresponding EEG data is usually temporally filtered to the 7–30 Hz frequency band (which contains the relevant motor imagery information - decrease in power in the alpha and beta bands [12, 5]) before applying CSP. However, CSP suffers from overfitting and poor generalizability from calibration to online control [6].

Variants of CSP have been proposed to learn the spatial patterns together with the optimal temporal filters (e.g., [15, 16, 7, 17, 18, 9, 19, 20, 21, 22, 23, 24, 25, 26, 27]). Canonical correlation approach to common spatial patterns (CCACSP) [7] combines the canonical correlation analysis (CCA) and CSP approaches and its cost function considers a temporal filter with a forward and backward one sample temporal shift. Spectrally-weighted CSP (spec-CSP) attempts at finding a set of spatial and spectral (frequency transformed temporal) filters through alternating optimization of the cost function [9].

Instead of learning the optimal temporal/spectral filters, another line of CSP variants finds the relevant sub-bands through a filter bank approach [22, 23, 24, 25]. These methods require a priori assumptions on the length and location of the sub-bands and require significant amounts of calibration data to find a robust feature space to be transferred from calibration data to online BCI control [28].

In this work, we propose a novel approach to learn spatial and spectral filters in motor imagery data by solving an optimization problem that generalizes the cost function in CCACSP [7]. We use an alternating direction method [8] to optimize our proposed cost function. We apply our method to motor imagery data that we recorded from 12 participants and show improved classification accuracy as well as more interpretable spatial patterns and spectral filters compared to the closely related methods of CSP, CCACSP and spec-CSP.

2 Methods

CSP estimates the spatial covariance for the two classes and uses a Raleigh quotient cost function to find the optimal directions for such projection [1, 5]. In SACSP [2] however, we define the spectrally-weighted spatial covariance instead, to weight the motor imagery related frequencies in the estimation of the covariance.

Let $X_i \in \mathbb{R}^{n \times t}$ represent the bandpassed EEG data epoch over n channels with t time samples that belongs to class $i \in \{1, 2\}$, e.g. the two motor imagery classes. Since EEG data is bandpassed, $\mathbb{E}[X_1] = \mathbb{E}[X_2] = 0$. Let C be a circulant matrix. We define the spectrally-weighted spatial covariance matrices for classes 1 and 2 as $\Gamma_{S1} = \mathbb{E}[X_1 C X_1^T]$ and $\Gamma_{S2} = \mathbb{E}[X_2 C X_2^T]$, where $(\cdot)^T$ indicates the transpose of a matrix/vector. As C is circulant, we can write it as $C = F\text{diag}(h)F^H$ where h is a vector, $diag(h)$ is a matrix whose diagonal elements are the elements of the vector h, F is the $t \times t$ Fourier matrix and $(\cdot)^H$ indicates the complex conjugate transpose of a matrix. Replacing this in the above and defining $\hat{X}_1 = X_1 F$ and $\hat{X}_2 = X_2 F$ as the Fourier transform of an epoch along its temporal dimension we can write the following set of optimization problems:

$$\underset{w,h}{\arg\max} \quad \frac{w^T \mathbb{E}[\hat{X}_1 \text{diag}(h) \hat{X}_1^H] w}{w^T (\Sigma_1 + \Sigma_2) w}, \qquad \underset{v,l}{\arg\max} \quad \frac{v^T \mathbb{E}[\hat{X}_2 \text{diag}(l) \hat{X}_2^H] v}{v^T (\Sigma_1 + \Sigma_2) v}, \tag{3}$$

where $\Sigma_i = \mathbb{E}[X_i X_i^T]$ is the spatial covariance of each class. We will solve the two optimization problems in (3) via an alternating maximization procedure. If h and l are fixed, the optimal w must satisfy the first equation below and the optimal v the second.

$$\mathbb{E}[\hat{X}_1 diag(h) \hat{X}_1^H] w = \lambda_w (\Sigma_1 + \Sigma_2) w, \qquad \mathbb{E}[\hat{X}_2 diag(l) \hat{X}_2^H] v = \lambda_v (\Sigma_1 + \Sigma_2) v. \tag{4}$$

If v and w are fixed, solving for h subject to $||h||_2 \leq 1$ leads to the following solution:

$$h_k = (\mathbb{E}[\hat{X}_1^H w w^T \hat{X}_1])_{kk}, \;\; k \in \{1, 2, .., t\}, \tag{5}$$

where kk indicates the $k-$ th diagonal element of $(\mathbb{E}[\hat{X}_1^H w w^T \hat{X}_1])$. Similarly, the solution for l subject to the norm-2 constraint $||l||_2 \leq 1$ leads to the following solution:

$$l_k = (\mathbb{E}[\hat{X}_2^H v v^T \hat{X}_2])_{kk}, \;\; k \in \{1, 2, .., t\}, \tag{6}$$

where kk indicates the $k-$ th diagonal element of $(\mathbb{E}[\hat{X}_2^H v v^T \hat{X}_2])$. Putting the above together, we propose Alg. 1. Note that for each spatial filter, our algorithm also estimates the spatial patterns as Pw and Pv which we will visualize later for neurophysiological interpretation of the features.

Since EEG data is often rank deficient (due to for instance re-referencing to the common average), we recommend projecting the EEG data to a full-rank subspace by finding L and Ψ such that $\Sigma_1 + \Sigma_2 = L \Psi L^T$. Then we find the projection operator as $Q = \Psi^{-\frac{1}{2}} L^T$ and project the EEG data as $Y_1 = Q X_1$ and $Y_2 = Q X_2$. We apply algorithm 1 and eventually project the spatial filters back to the original space by multiplying Q^T to the left side of the selected spatial filters.

Classification: After finding the optimal $2 \times R$ spatial and spectral filter pairs (f_{wv}, f_{hl}), then $2 \times R$ features were extracted by filtering each epoch, x as follows:

$$\log(f_{wv}^T \; (\hat{x} \; \text{diag}(f_{hl}) \hat{x}^H) \; f_{wv}), \tag{7}$$

where $\hat{x} = xF$. Next, linear discriminant analysis (LDA) [10] with automatic shrinkage using the Ledoit-Wolf lemma [11] was trained as the classifier.

Choice of parameters in SACSP: We chose the top 3 filters for each class, i.e., $R = 3$ as this is usually the choice for CSP and other similar methods [29, 5, 28]. We set $\epsilon = 1e - 6$; however, our investigations show that SACSP is not highly dependent on this parameter. Of course if the results for a specific dataset vary significantly for different values of ϵ, one can always set it through cross-validation. Furthermore, the spectral filters h and l were initialized with 3 initializations each(i.e., $M = 3$): 1) $h_k = l_k = 1, \forall k \in \{0, 1, ..., t-1\}$, 2) $h_k = l_k = 1$ only in 7–15 Hz to include mu frequency, and 3) $h_k = l_k = 1$ only in 15–30 Hz to include beta frequency.

[2] We provide the anonymized and epoched EEG data at: `https://github.com/anonymsquirrel/SACSP`. We release our developed code under the MIT license.

Data: X_1 and X_2, R = number of spectral/spatial filters for each class
Result: w_r^o, v_r^o, h_r^o, l_r^o, Pw_r^o and Pv_r^o for $r \in \{1, 2, .., R\}$
initialization: $h_m^i/||h_m^i||_2$ and $l_m^i/||l_m^i||_2$ for $m \in \{1, 2, .., M\}$;
for m = 1 *to* M **do**
  select w_1 to w_R with largest λ_w from $\mathbb{E}[\hat{X}_1 \text{diag}(h_m^i)\hat{X}_1^H]w = \lambda_w(\Sigma_1 + \Sigma_2)w$;
  let A concatenate the sorted eigenvectors w; select Pw_1 to Pw_R as the first R columns of $(A^{-1})^T$;
  select v_1 to v_R with largest λ_v from $\mathbb{E}[\hat{X}_2 \text{diag}(l_m^i)\hat{X}_2^H]v = \lambda_v(\Sigma_1 + \Sigma_2)v$;
  let B concatenate the sorted eigenvectors v; select Pv_1 to Pv_R as the first R columns of $(B^{-1})^T$;
  for r = 1 *to* R **do**
    while *the cost function for class 1 (Eq. 3) increases more than* ϵ **do**
      update the spectral filter h_{m_r} as $\{h_{m_r}\}_k = (\mathbb{E}[\hat{X}_1^H w_r w_r^T \hat{X}_1])_{kk}$, $k \in \{1, 2, .., t\}$;
      $h_{m_r} \leftarrow h_{m_r}/||h_{m_r}||_2$;
      update the spatial filter, w_r, by selecting the eigenvector of the largest λ_w from $\mathbb{E}[\hat{X}_1 \text{diag}(h_{m_r})\hat{X}_1^H]w = \lambda_w(\Sigma_1 + \Sigma_2)w$;
      let A concatenate the sorted eigenvectors w; update the spatial pattern Pw_r as the first column of $(A^{-1})^T$;
    end
    while *the cost function for class 2 (Eq. 3) increases more than* ϵ **do**
      update the spectral filter l_{m_r} as $\{l_{m_r}\}_k = (\mathbb{E}[\hat{X}_2{}^T v_r v_r^T \hat{X}_2])_{kk}$, $k \in \{1, 2, .., t\}$;
      $l_{m_r} \leftarrow l_{m_r}/||l_{m_r}||_2$;
      update the spatial filter, v_r, by selecting the eigenvector of the largest λ_v from $\mathbb{E}[\hat{X}_2 \text{diag}(l_{m_r})\hat{X}_2^H]v = \lambda_v(\Sigma_1 + \Sigma_2)v$;
      let B concatenate the sorted eigenvectors v; update the spatial pattern Pv_r as the first column of $(B^{-1})^T$;
    end
  end
end
Among the $M \times R$ spatial and spectral filter pairs pairs for each class, select R that maximizes the cost function on the training data for the corresponding class (Eq. 3) to find w_r^o, v_r^o, h_r^o, l_r^o, Pw_r^o and Pv_r^o for $r \in \{1, 2, .., R\}$;

Algorithm 1: Spectrally adaptive common spatial patterns (SACSP).

Comparison with other methods: We compared our proposed method with CSP and two closely related methods in the literature, namely spec-CSP [9] and CCACSP [7]. For every algorithm, we chose the top 3 filters for each class. SACSP reduces to CSP if $h_k = l_k = 1, \forall k \in \{0, 1, ..., t-1\}$ and reduces to CCACSP if $h_k = l_k = \cos(\frac{2\pi k}{t}), \forall k \in \{0, 1, ..., t-1\}$. On the other hand, Spec-CSP finds spatial and spectral filters through an alternating fashion and has two hyperparameters that were set through cross-validation [9]. Per the authors' recommendation in the original paper [9], the number of iteration steps in spec-CSP was set to 10 (with one spatial and spectral filter update in each step), hyperparameters $p\prime \in \{0, 0.5, 1\}$ and $q' \in \{0, 0.5, 1, 1.5, 2\}$ were optimized using cross-validation on the calibration data and the 7–30 Hz frequency band was selected as apriori knowledge on the motor imagery signal similar to [9].

2.1 Experiment

After the study was approved by the institutional review board of (removed for anonymity), data were recorded from 12 participants (7 females, 1 left-handed, average age = 20.4 $\pm$ 1.0) who were naive to BCI systems. Each participant signed an informed consent form prior to participating in one session of motor imagery. They were compensated with experimental course credit or at the rate of (amount/currency omitted for anonymity) based on their preference. During the experiment which took roughly 2.5 hours, participants were comfortably sitting in an arm chair about one meter away from the screen (Dell 19" CRT monitor). They were asked to minimize their movements to minimize the occurrence of eye movements and muscular artifacts in the EEG recordings. Furthermore, participants could take as much break between the blocks as needed.

Each session had two main parts: in the first part, participants were instructed (over 30 trials) on how to perform movement imagination of their right or left hand (reference omitted for anonymity). Note that this part of the study was only designed to provide some exposure to the otherwise naive participants and the recorded EEG data in this part was not used in the analysis that is presented next.

For the second part of the experiment, participants were instructed to use movement imagination of their right or left hand to move a cursor on the screen. Each trial began with a cursor (blue circle about 2 cm in diameter) in the center and a target (white circle about 2 cm in diameter) at either side of the screen and three steps away from the cursor. The cursor moved one step per 1.2 seconds towards or away from the target until it reached the target location or the corresponding location on the other side of the screen. Each participant participated in 9 blocks comprising 20 trials each. During the first three blocks, the cursor moved according to a pre-determined pseudo-randomly generated sequence of movements but the participants were not aware of this and were led to believe that they were in control of the cursor movements. The recorded EEG data from the first three blocks were used for calibration. CSP was trained for feature extraction where the top three CSP filters for each class were selected. The logarithm of the power of the CSP-filtered EEG data were selected as features and shrinkage-LDA [11] was trained for classification. This classifier was used during the online control of the cursor in the latter 6 blocks which had a similar structure as the calibration data except that the cursor moved based on the participant's motor imagery data. An example of a trial is provided in the extra supplementary material section 1.

In an offline analysis, we used the recorded data from the calibration blocks to train CSP, CCACSP, spec-CSP and SACSP separately. Each method was then tested on the data that were recorded during the online control blocks. Note that we train and test the methods on every cursor movement (step) and the results are reported accordingly in the next section.

Data collection: EEG data were recorded with a 64-channel BrainAmp system (Brain Products GmbH) at 5000 Hz sampling rate and downsampled to 100 Hz for online control and later for the offline analysis. Electrodes were located according to the international 10-20 system [30], the impedance of the electrodes was set to be below 20 $K\Omega$. The online reference and ground electrodes were located at FCz and AFz, respectively.

We used the python-based Simulation and Neuroscience Application Platform (SNAP) [31] to design the stimuli and Lab Streaming Layer (LSL) [32] as an interface between the computer and the EEG system as well as to record the EEG data. For data processing and classification in online control and later for offline analysis of the data, we used Numpy [33], Scipy [34] and Scikit-learn [10]. MATLAB [35] and EEGLAB [36] were used for epoching the EEG data and for plotting the results.

Pre-processing: All 64 channels were used in this study. Data from calibration and online blocks were downsampled to 100 Hz and epoched 0-1 second with respect to each cursor movement excluding the last cursor movement. This is because participants stopped motor imagery when the trial ended with the last cursor movement. After downsampling and epoching, data were re-referenced to the common average and bandpass filtered in 7–30 Hz to include the mu (about 7–13 Hz) and beta (about 13–30 Hz) frequency bands [12, 13]. The bandpass filter was designed as a 6th order Butterworth filter and was applied to each epoch with the `filtfilt` function from Scipy. This function applies the filter once forward and once backward and results in a zero-phase filtering.

Since in calibration blocks, all participants were shown the same pre-determined sequence of cursor movements, the number of calibration steps in each class did not vary across participants. However, the number of steps in each class in online blocks varied across participants depending on their online performance. We balanced the right and left classes (including movements towards and away from the target in each class) for both calibration and online blocks to alleviate any potential bias in learning features and for easier interpretation of the results. After balancing the number of right and left motor imagery steps, there were a total of 136 steps in the calibration data and on average across participants, a total of 272.7 ± 38.4 steps in the online blocks. Balancing the classes was done 10 times to allow steps to be in the training (from calibration data) or test (from online data) sets.

Visualization: We visualize the optimal spatial patterns and spectral filters to compare the neurophysiological interpretation of of the methods under study. Normalized spectral filters were plotted against frequency (Hz). as line plots using MATLAB. Spatial patterns were estimated for each method accordingly and `topoplot` from EEGLAB was used to plot them as heatmap on a 2-D scalp.

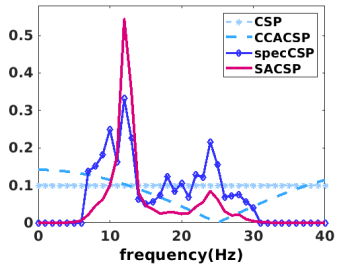


Figure 2: Comparison of the normalized spectral filters in P9. SACSP and spec-CSP lines show the top spectral filter for the right hand motor imagery class. Note that CSP and CCACSP have fixed spectral filters.

3 Results

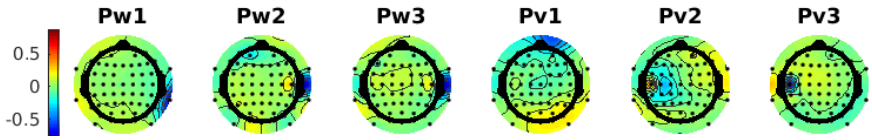
Classification accuracy: The classification accuracy of the CSP, CCACSP, spec-CSP and SACSP when the feature extraction and classifiers were trained on the calibration data and tested on the online data, are presented in table 2. The first number in each entry for participants 1 to 12 (P1-P12), represents the mean classification accuracy over 10 instances of train-test data and the second number reports the standard deviation. In the last row, the average performance across participants is reported where in each entry, the first number indicates the average for the corresponding column and the second number indicates the standard error of the mean. Across participants, SACSP performs significantly better compared to CCACSP, spec-CSP and CSP (Wilcoxon signed rank test, $p < 0.01$).

We also looked at the classification accuracy when each method was trained on calibration data and tested on calibration data through cross-validation (we call it 'calibCV' case) as well as when trained on online data and tested on online data through cross-validation (we call it 'onlineCV' case). Note that both of these conditions are different from the practical scenario of interest in this work, where the methods are trained on the calibration data and tested on the online data (we call it 'test' case). Our results show that for CSP, calibCV and onlineCV are not different across participants (average 0.73 and 0.71, Wilcoxon signed-rank test, $p = 0.42$)– results for each participant are presented in extra supplementary material section 2. However, as table 2 shows, there is a large drop in performance of CSP in the test condition (across participants significantly lower that calibCV and onlineCV, Wilcoxon signed rank test, $p < 0.001$). Our results show that spec-CSP, CCACSP, and SACSP were all able to close this gap from calibration to online performance with SACSP performing significantly better than all other methods (Wilcoxon signed rank test, $p < 0.01$).

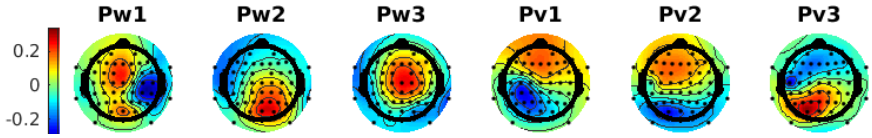
Visualization of spatial patterns and spectral filters: We visualize the optimal spatial patterns and spectral filters to compare the neurophysiological interpretation of of the methods under study. Normalized spectral filters were plotted against frequency (Hz). Figure 3 compares the spatial patterns and spectral filters of the SACSP with CSP, CCACSP, and spec-CSP. Note that these are the filters and patterns that were trained on the calibration data and used later for classification of the online data. We present the results for the rest of the participants in the extra supplementary material section 3. The last two rows show the results for SACSP. The filters and patterns for each class (Pw_1, Pw_2 and Pw_3, h_1, h_2 and h_3) and (Pv_1, Pv_2 and Pv_3, l_1, l_2 and l_3) converge to similar values. This is because our algorithm, different from the others presented here, does not result in spatial filters/patterns that are necessarily orthogonal to each other and convergence to the same solution from different initializations in fact shows that SACSP converged to its optimal solution. Furthermore, we tested SACSP with random initialization and received the same filter and patterns.

Comparing the spectral filters for spec-CSP and SACSP, we can see that the spec-CSP spectral filters fail at selectively extracting the underlying features in an interpretable way. SACSP shows smoother and less noisy filters with a peak at/around the mu band for both the right and left hand motor imagery classes and a small beta peak for the left hand motor imagery class. Another example comparing spectral filters for P9 is presented in figure 2.

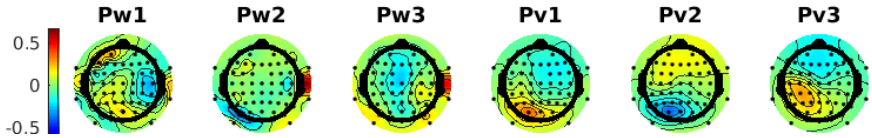
Comparing the spatial patterns, SACSP shows patterns that are typical for right and left hand motor imagery. Pv_3 in spec-CSP can be regarded as a similar pattern to Pv_3 in SACSP but the rest of the



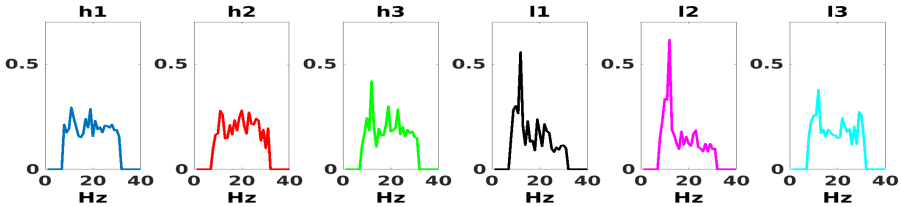
(a) P1 – CSP spatial patterns.



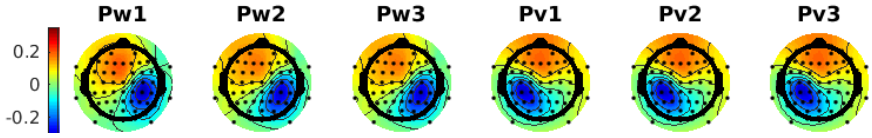
(b) P1 – CCACSP spatial patterns.



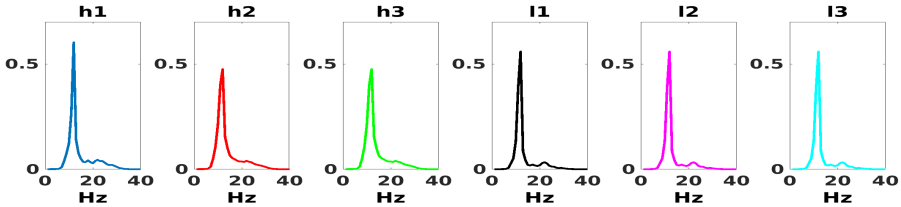
(c) P1 – spec-CSP spatial patterns.



(d) P1 – spec-CSP spectral filters.



(e) P1 – SACSP spatial patterns.



(f) P1 – SACSP spectral filters.

Figure 3: Spatial patterns and spectral filters trained on the calibration data for P1. Compared to other methods, SACSP results in spatial patterns that are typical for right and left hand motor imagery while providing neurophysiologically relevant frequency selectivity in spectral filters.

Table 2: Motor imagery (right/left hand) classification accuracy. For P1-P12, the first number in each entry reports the average accuracy across 10 instances of the data and the second number reports the standard deviation of the mean. In the last row, the first number in each entry reports the average classification accuracy across participants for the corresponding column and the second number reports the standard error of the mean.

PID	CSP	spec-CSP	CCACSP	SACSP
P1	0.55/0.035	0.61/0.039	0.74/0.033	0.81/0.022
P2	0.56/0.028	0.63/0.091	0.69/0.012	0.69/0.030
P3	0.63/0.038	0.61/0.037	0.61/0.057	0.72/0.096
P4	0.72/0.019	0.70/0.051	0.71/0.020	0.73/0.009
P5	0.59/0.118	0.59/0.105	0.56/0.071	0.62/0.117
P6	0.50/0.014	0.52/0.032	0.54/0.017	0.52/0.013
P7	0.62/0.027	0.65/0.035	0.68/0.019	0.70/0.040
P8	0.57/0.023	0.59/0.034	0.71/0.023	0.72/0.036
P9	0.68/0.048	0.67/0.033	0.73/0.019	0.76/0.023
P10	0.49/0.021	0.49/0.024	0.49/0.017	0.50/0.019
P11	0.60/0.019	0.64/0.028	0.64/0.034	0.67/0.027
P12	0.47/0.032	0.45/0.025	0.49/0.024	0.51/0.010
Average	0.58/0.022	0.60/0.022	0.63/0.027	**0.66/0.030**

patterns in spec-CSP and the ones in CSP are neurophysiologically irrelevant for motor imagery. Pv_1 in CCACSP is a motor imagery pattern and Pv_3 could also be interpreted as such. However, CCACP fails at finding appropriate patterns for the other class.

4 Discussion

In this work, we proposed SACSP, a novel feature extraction method for motor imagery data, and showed its superiority in classification accuracy when trained on calibration data and tested on online data as well as neurophysiologically plausible spatial patterns and spectral filters. On the classification accuracy, we should emphasize that this is performance for one second of motor imagery data and still more than half of the participants perform around 0.7 or higher with the proposed algorithm. The majority of existing work in the literature focus on cross-validation accuracy (not calibration to online transfer) and over longer lengths of motor imagery data. Furthermore, the SACSP spatial patterns are in fact very typical of the ones from motor imagery data [5] and the three methods we compare with SACSP in this paper, fully or partially fail at learning them. Furthermore, the spectral filters are also as expected, revealing existing information in mu and beta frequency bands [4, 12]. One important and valuable observation from an engineering and scientific perspective, is the difference among the spectral filters learned for the right and left hand motor imagery classes and the ones learned for different participants. Note that in SACSP, different from sub-band and filterbank variations of CSP, we learn the spectral filters directly instead of a priori assuming multiple frequency bands and searching for class-correlated information in those bands [22, 23, 24, 25].

SACSP has a few hyperparameters to choose: 1) the number of spectral/spatial filters, 2) initialization for the spectral filters, and 3) the threshold for the stopping of the iterations. We selected the number of spatial filters for each class to be 3 as is usually selected for CSP in the literature [5, 28]. In the case of CSP, three orthogonal filters are selected and the LDA classifier weighs them according to their relevance. However, our proposed algorithm initializes the spatial filters in each class to be orthogonal to each other but through iterations, they do not necessarily remain orthogonal. In fact, our results show that often the top three selected spatial filters that maximized the cost function for each class, converge to the same solution. This is an interesting observation and highlights the fact that SACSP converges with different initial conditions and paths to convergence.

We also investigated SACSP with different initializations of the spectral filters including random initialization; but neither the classification accuracy nor the spatial patterns/spectral filters were different upon convergence. Moreover, we varied ϵ from $1e-8$ to $1e-3$, but the number of steps to convergence remained the same varying from 1 to 15 for different initializations and for different

participants/classes. On a Linux machine, this took a few seconds for each participant to converge as opposed to spec-CSP that took a few minutes for each participant. Therefore, the main parameter to choose for in SACSP is the number of spatial/spectral filters and 3 seems to be a good choice given that it often converges to the same solution and increasing the number is not of benefit.

As mentioned earlier, one major downside of the CSP method is its poor generalizability from calibration to online control [6]. Various approaches can be taken to solve this issue. The first approach, as discussed in the Introduction section, is to design neurophysiologically plausible cost functions to extract robust spectral/temporal or spatial information from EEG data. A second approach is to make use of available EEG data from other users to learn a classifier and then fine-tune it to the new user either through a supervised or unsupervised approach. For a review, please see [28]. A third approach is to use other available classifiable information (such as the user response to the BCI erroneous output [37]) in combination with the motor imagery signal in a hybrid BCI fashion [38, 13, 39]. The proposed method of SACSP belongs to the first category of approaches above. As future work, the performance of SACSP when transferred from one participant to another should be further studied. Nevertheless, it can be easily combined with other classifiers in a hybrid fashion replacing CSP to further improve the BCI performance.

Even though one of the main goals for BCI applications is to provide means of communication and locomotion for patients who suffer from the locked-in syndrome [2], current BCI systems are for the most part limited to the laboratory environment. One important limitation is that the trained classifier does not transfer well to online control resulting in long calibration times or frequent recalibration. Together with our proposed SACSP method, we also make our dataset and code freely available (`https://github.com/anonymsquirrel/SACSP`) so that other researchers can test how their methods would perform from calibration to online control. We hope that through open science and open data/code, the BCI research community can make progress in developing better techniques for use outside of the laboratory.

References

[1] Herbert Ramoser, Johannes Muller-Gerking, and Gert Pfurtscheller. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Transactions on Rehabilitation Engineering*, 8(4):441–446, 2000.

[2] Luis Fernando Nicolas-Alonso and Jaime Gomez-Gil. Brain computer interfaces, a review. *Sensors*, 12(2):1211–1279, 2012.

[3] Piotr Wierzgała, Dariusz Zapała, Grzegorz M Wojcik, and Jolanta Masiak. Most popular signal processing methods in motor-imagery bci: a review and meta-analysis. *Frontiers in neuroinformatics*, 12:78, 2018.

[4] Gert Pfurtscheller and Christa Neuper. Motor imagery and direct brain-computer communication. *Proceedings of the IEEE*, 89(7):1123–1134, 2001.

[5] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, and K-R Muller. Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Processing Magazine*, 25(1):41–56, 2008.

[6] Pradeep Shenoy, Matthias Krauledat, Benjamin Blankertz, Rajesh PN Rao, and Klaus-Robert Müller. Towards adaptive classification for BCI. *Journal of Neural Engineering*, 3(1):R13, 2006.

[7] Eunho Noh and Virginia R De Sa. Canonical correlation approach to common spatial patterns. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 669–672. IEEE, 2013.

[8] Stephen P. Boyd, N. Parikh, E. Chu, B. Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3:1–122, 2011.

[9] Ryota Tomioka, Guido Dornhege, Guido Nolte, Benjamin Blankertz, Kazuyuki Aihara, and Klaus-Robert Müller. Spectrally weighted common spatial pattern algorithm for single trial eeg classification. *Dept. Math. Eng., Univ. Tokyo, Tokyo, Japan, Tech. Rep*, 40, 2006.

[10] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(Oct):2825–2830, 2011.

[11] Olivier Ledoit and Michael Wolf. Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119, 2004.

[12] Gert Pfurtscheller, Clemens Brunner, Alois Schlögl, and FH Lopes Da Silva. Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks. *NeuroImage*, 31(1):153–159, 2006.

[13] Mahta Mousavi, Adam S Koerner, Qiong Zhang, Eunho Noh, and Virginia R de Sa. Improving motor imagery BCI with user response to feedback. *Brain-Computer Interfaces*, 4(1-2):74–86, 2017.

[14] Brendan Z Allison and Christa Neuper. Could anyone use a bci? In *Brain-computer interfaces*, pages 35–54. Springer, 2010.

[15] Steven Lemm, Benjamin Blankertz, Gabriel Curio, and K-R Muller. Spatio-spectral filters for improving the classification of single trial EEG. *IEEE transactions on biomedical engineering*, 52(9):1541–1548, 2005.

[16] Qibin Zhao, Liqing Zhang, and Andrzej Cichocki. Multilinear generalization of common spatial pattern. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 525–528. IEEE, 2009.

[17] Guido Dornhege, Benjamin Blankertz, Matthias Krauledat, Florian Losch, Gabriel Curio, and K-R Muller. Combined optimization of spatial and temporal filters for improving brain-computer interfacing. *IEEE transactions on biomedical engineering*, 53(11):2274–2281, 2006.

[18] Roel M Willems, Ivan Toni, Peter Hagoort, and Daniel Casasanto. Body-specific motor imagery of hand actions: neural evidence from right-and left-handers. *Frontiers in Human Neuroscience*, 3:39, 2009.

[19] Wei Wu, Zhe Chen, Xiaorong Gao, Yuanqing Li, Emery N Brown, and Shangkai Gao. Probabilistic common spatial patterns for multichannel eeg analysis. *IEEE transactions on pattern analysis and machine intelligence*, 37(3):639–653, 2014.

[20] Wei Wu, Xiaorong Gao, Bo Hong, and Shangkai Gao. Classifying single-trial eeg during motor imagery by iterative spatio-spectral patterns learning (isspl). *IEEE Transactions on Biomedical Engineering*, 55(6):1733–1743, 2008.

[21] Hiroshi Higashi and Toshihisa Tanaka. Simultaneous design of fir filter banks and spatial patterns for eeg signal classification. *IEEE transactions on biomedical engineering*, 60(4):1100–1110, 2012.

[22] Quadrianto Novi, Cuntai Guan, Tran Huy Dat, and Ping Xue. Sub-band common spatial pattern (sbcsp) for brain-computer interface. In *2007 3rd International IEEE/EMBS Conference on Neural Engineering*, pages 204–207. IEEE, 2007.

[23] Kai Keng Ang, Zheng Yang Chin, Haihong Zhang, and Cuntai Guan. Filter bank common spatial pattern (FBCSP) in brain-computer interface. In *Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on*, pages 2390–2397. IEEE, 2008.

[24] Kai Keng Ang, Zheng Yang Chin, Chuanchu Wang, Cuntai Guan, and Haihong Zhang. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. *Frontiers in Neuroscience*, 6:39, 2012.

[25] Amirhossein S Aghaei, Mohammad Shahin Mahanta, and Konstantinos N Plataniotis. Separable common spatio-spectral patterns for motor imagery bci systems. *IEEE Transactions on Biomedical Engineering*, 63(1):15–29, 2015.

[26] Feifei Qi, Yuanqing Li, and Wei Wu. Rstfc: A novel algorithm for spatio-temporal filtering and classification of single-trial eeg. *IEEE transactions on neural networks and learning systems*, 26(12):3070–3082, 2015.

[27] Xinyang Li, Cuntai Guan, Haihong Zhang, and Kai Keng Ang. A unified fisher's ratio learning method for spatial filter optimization. *IEEE transactions on neural networks and learning systems*, 28(11):2727–2737, 2016.

[28] Fabien Lotte, Laurent Bougrain, Andrzej Cichocki, Maureen Clerc, Marco Congedo, Alain Rakotomamonjy, and Florian Yger. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *Journal of Neural Engineering*, 15(3):031005, 2018.

[29] Fabien Lotte, Marco Congedo, Anatole Lécuyer, Fabrice Lamarche, and Bruno Arnaldi. A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, 4(2):R1, 2007.

[30] George H Klem, Hans Otto Lüders, HH Jasper, C Elger, et al. The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, 52(3):3–6, 1999.

[31] UC San Diego Swartz Center for Computational Neuroscience. Simulation and Neuroscience Application Platform (SNAP). `https://github.com/sccn/SNAP`.

[32] UC San Diego Swartz Center for Computational Neuroscience. Lab Streaming Layer (LSL). `https://github.com/sccn/labstreaminglayer`.

[33] Travis E Oliphant. *A guide to NumPy*, volume 1. Trelgol Publishing USA, 2006.

[34] Eric Jones, Travis Oliphant, Pearu Peterson, et al. Scipy: Open source scientific tools for Python. 2001.

[35] MATLAB and Statistics Toolbox Release 2018b. The MathWorks Inc., Natick, Massachusetts, United States, 2018.

[36] Arnaud Delorme and Scott Makeig. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21, 2004.

[37] Ricardo Chavarriaga, Aleksander Sobolewski, and José del R Millán. Errare machinale est: the use of error-related potentials in brain-machine interfaces. *Frontiers in Neuroscience*, 8:208, 2014.

[38] Gert Pfurtscheller, Brendan Z Allison, Günther Bauernfeind, Clemens Brunner, Teodoro Solis Escalante, Reinhold Scherer, Thorsten O Zander, Gernot Mueller-Putz, Christa Neuper, and Niels Birbaumer. The hybrid BCI. *Frontiers in Neuroscience*, 4:3, 2010.

[39] Mahta Mousavi, Laurens Ruben Krol, and Virginia de Sa. Hybrid brain-computer interface with motor imagery and error-related brain activity. *Journal of Neural Engineering*, 2020.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] See Abstract.
   (b) Did you describe the limitations of your work? [Yes] See Methods and Discussion sections.
   (c) Did you discuss any potential negative societal impacts of your work? [N/A] Our work does not have negative societal impacts.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]

2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] The code and data are available in a public (currently anonymous during review process) repository `https://github.com/anonymsquirrel`
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See Methods section.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] All results have standard deviation or standard error of the mean reported with them.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] See Methods section.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] A URL is provided to an anonymous github account with the developed code.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [Yes] See Methods section.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [Yes] See Methods section.
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [Yes] Described in the supplementary material.
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [Yes] Described in the supplementary material. Our study was considered as posing no more than minimal risk to the participants in that the probability and magnitude of harm or discomfort anticipated from the study participation are no greater than the ones encountered in daily life or during physical or psychological examinations or tests. The IRB approval letters are not provided for anonymity but will be provided with the final submission.
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [Yes] See Methods section.

Spectrally Adaptive Common Spatial Patterns – Extra supplementary material

Anonymous Author(s)
Affiliation
Address
email

1 Methods – experiment

Our study was considered as posing no more than minimal risk to the participants in that the probability and magnitude of harm or discomfort anticipated from the study participation are no greater than the ones encountered in daily life or during physical or psychological examinations or tests. The IRB approval letter and the consent form are not provided for anonymity but will be provided with the final submission.

An example of one trial in our experiment is presented in figure 1. Target disappears after 1.2 seconds from the beginning of the trial to minimize eye artifacts. The experiment instructions for the participants are attached after page 15.

2 Results – cross-validation accuracy

Table 1 shows the classification accuracy of the CSP method when trained and tested on calibration data through cross-validation (called 'calibCV' and depicted in blue color), when trained and tested on online data through cross-validation (called 'onlineCV and depicted in red color) and when trained on calibration data and tested on the online data as reported in the main article (called 'test' and depicted in bold font). Similarly, table 2 shows the classification accuracy of the SACSP method when trained and tested on calibration data through cross-validation (called 'calibCV' and depicted in blue color), when trained and tested on online data through cross-validation (called 'onlineCV and depicted in red color) and when trained on calibration data and tested on the online data as reported in the main article (called 'test' and depicted in bold font).

Note that the 'test' condition (reported in bold font here) is the natural scenario that is relevant for practical applications, i.e., trained on calibration data and tested on online data. This is also reported in the main text in Table 1. As for the CV rates, note that CSP performs significantly better in calibCV and onlineCV conditions (red and blue fonts) compared to the test condition in bold font (Wilcoxon signed rank test $p < 0.001$). This suggests that CSP is overfitting to the data in calibration and online blocks. On the other hand, SACSP does not perform different in the onlineCV (red font) and the test condition in black bold font (Wilcoxon signed-rank test $p = 0.97$) suggesting that SACSP is not suffering from overfitting as CSP does.

3 Results – spatial patterns and spectral filters

Figures 2 to 13 present the spatial patterns and spectral filters for right and left motor imagery classes for all participants (P1-P12) with CSP, CCACSP, spec-CSP and SACSP methods.

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.

Table 1: Comparison of motor imagery (right/left hand) classification accuracy with CSP when trained and tested on calibration data (CSP-calibCV), trained and tested on online data (CSP-onlineCV) and trained on calibration and tested on online data (**CSP-test**). For P1-P12, the first number in each entry reports the average accuracy across 10 instances of the data and the second number reports the standard deviation of the mean. In the last row, the first number in each entry reports the average classification accuracy across participants for the corresponding column and the second number reports the standard error of the mean.

PID	CSP-calibCV	CSP-onlineCV	**CSP-test**
P1	0.73/0.019	0.84/0.023	**0.55/0.035**
P2	0.74/0.021	0.70/0.022	**0.56/0.028**
P3	0.75/0.024	0.78/0.050	**0.63/0.038**
P4	0.87/0.027	0.72/0.017	**0.72/0.019**
P5	0.79/0.051	0.81/0.026	**0.59/0.118**
P6	0.60/0.046	0.57/0.022	**0.50/0.014**
P7	0.63/0.062	0.77/0.036	**0.62/0.027**
P8	0.67/0.039	0.70/0.044	**0.57/0.023**
P9	0.81/0.040	0.73/0.017	**0.68/0.048**
P10	0.73/0.039	0.65/0.023	**0.49/0.021**
P11	0.81/0.032	0.64/0.050	**0.60/0.019**
P12	0.67/0.038	0.59/0.019	**0.47/0.032**
Average	0.73/0.023	0.71/0.024	**0.58/0.022**

Table 2: Comparison of motor imagery (right/left hand) classification accuracy with SACSP when trained and tested on calibration data (SACSP-calibCV), trained and tested on online data (SACSP-onlineCV) and trained on calibration and tested on online data (**SACSP-test**). For P1-P12, the first number in each entry reports the average accuracy across 10 instances of the data and the second number reports the standard deviation of the mean. In the last row, the first number in each entry reports the average classification accuracy across participants for the corresponding column and the second number reports the standard error of the mean.

PID	SACSP-calibCV	SACSP-onilneCV	**SACSP-test**
P1	0.79/0.020	0.82/0.023	**0.81/0.022**
P2	0.68/0.018	0.67/0.021	**0.69/0.030**
P3	0.77/0.046	0.79/0.027	**0.72/0.096**
P4	0.89/0.013	0.71/0.017	**0.73/0.009**
P5	0.65/0.103	0.52/0.040	**0.62/0.117**
P6	0.56/0.022	0.55/0.022	**0.52/0.013**
P7	0.68/0.036	0.79/0.015	**0.70/0.040**
P8	0.72/0.030	0.71/0.032	**0.72/0.036**
P9	0.85/0.022	0.69/0.018	**0.76/0.023**
P10	0.75/0.030	0.57/0.035	**0.50/0.019**
P11	0.76/0.023	0.63/0.032	**0.67/0.027**
P12	0.67/0.061	0.51/0.018	**0.51/0.010**
Average	0.73/0.026	0.66/0.031	**0.66/0.030**

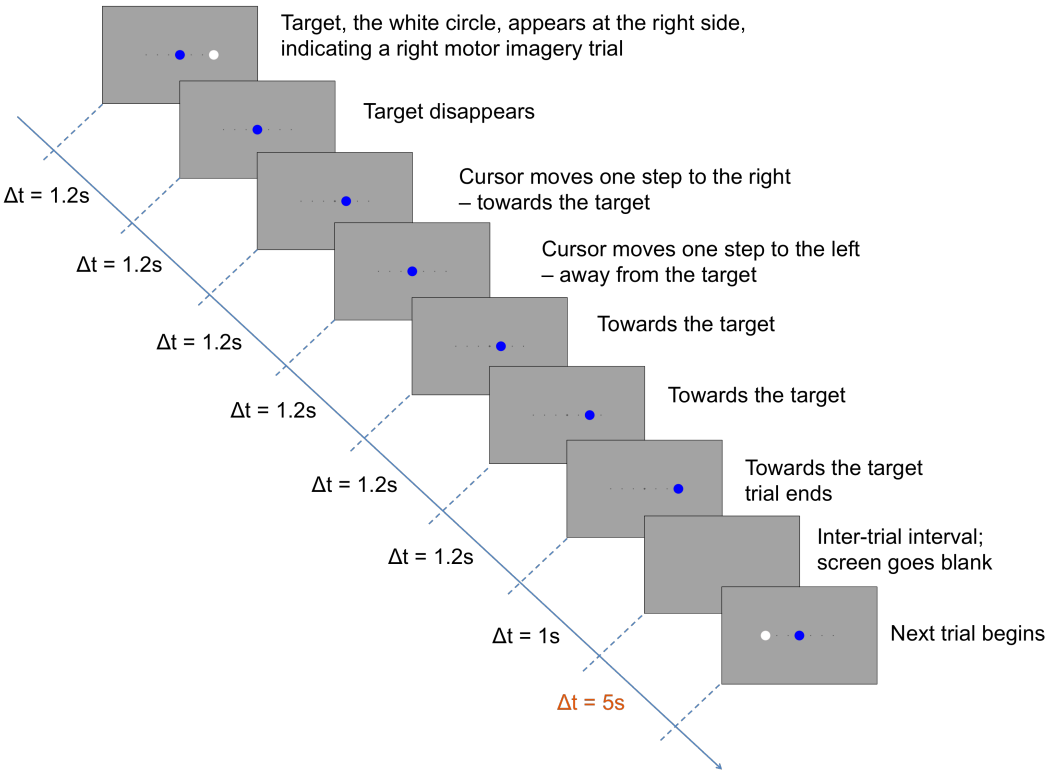
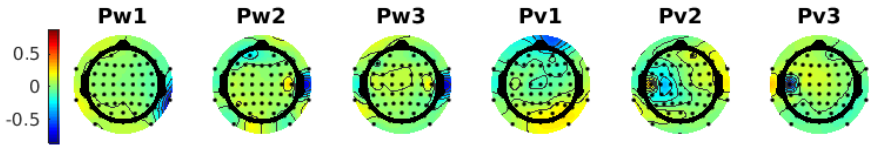
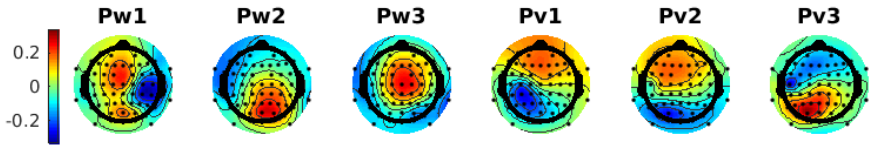


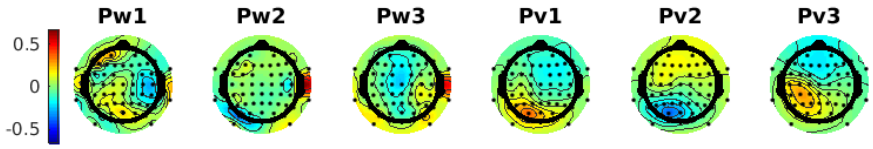
Figure 1: An example of a trial: Each trial began with the cursor (blue circle) at the center and the target (white circle) at either side of the screen, exactly three steps from the cursor. Trial ended when the cursor hit the target location or the corresponding location on the other side or after 12 cursor steps, whichever came first. Participants had 5 seconds to rest before the next trial began. Note that the background of the frames was dark gray in the experiment but is depicted lighter here for easier visualization of the details.



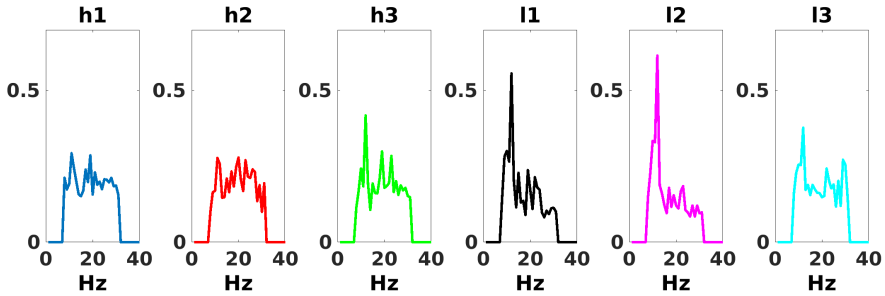
(a) P1 – CSP spatial patterns.



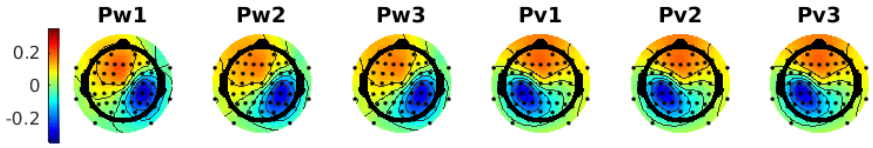
(b) P1 – CCACSP spatial patterns.



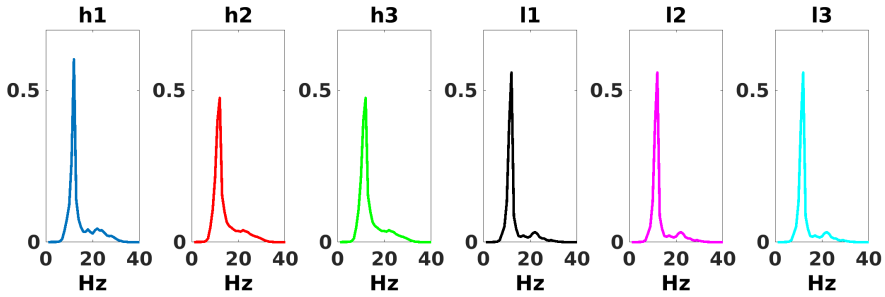
(c) P1 – spec-CSP spatial patterns.



(d) P1 – spec-CSP spectral filters.

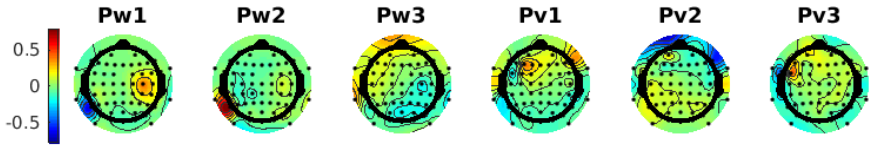


(e) P1 – SACSP spatial patterns.

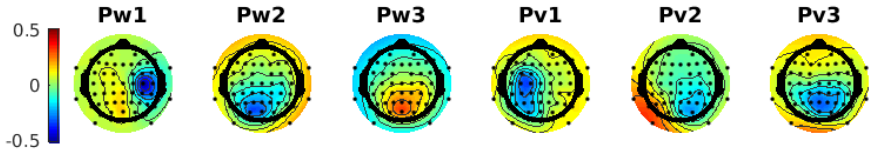


(f) P1 – SACSP spectral filters.

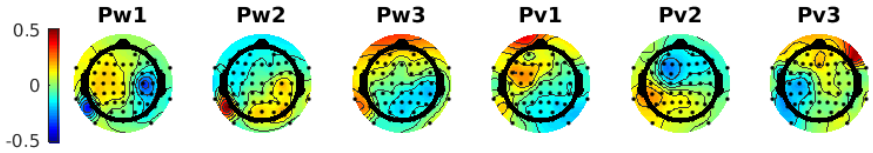
Figure 2: Spatial patterns and spectral filters trained on the calibration data for P1.



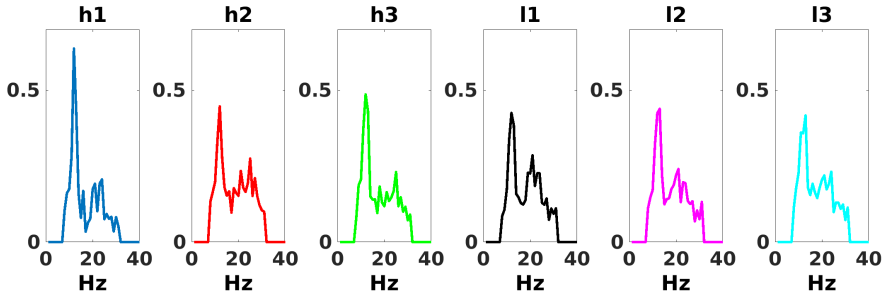
(a) P2 – CSP spatial patterns.



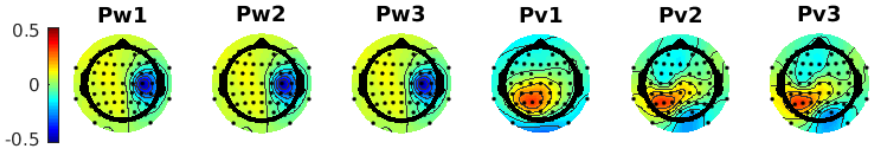
(b) P2 – CCACSP spatial patterns.



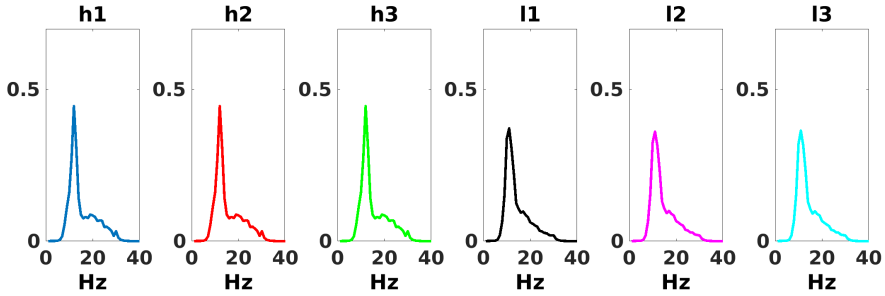
(c) P2 – spec-CSP spatial patterns.



(d) P2 – spec-CSP spectral filters.

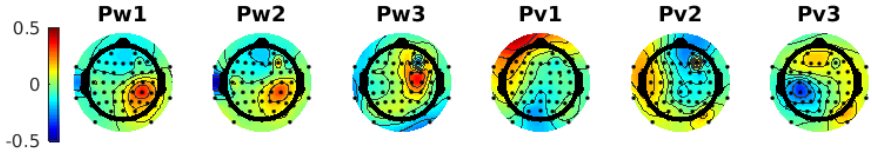


(e) P2 – SACSP spatial patterns.

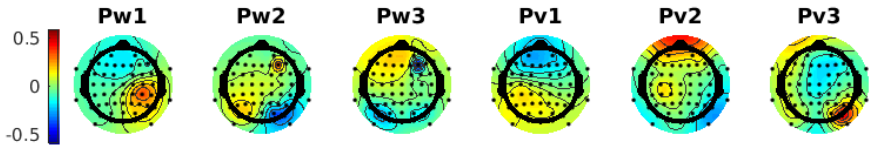


(f) P2 – SACSP spectral filters.

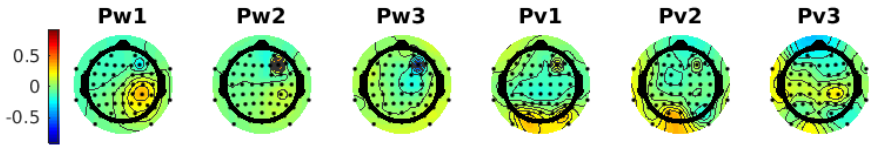
Figure 3: Spatial patterns and spectral filters trained on the calibration data for P2.



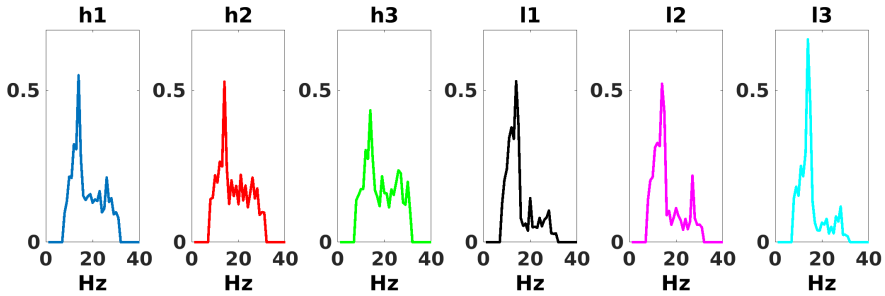
(a) P3 – CSP spatial patterns.



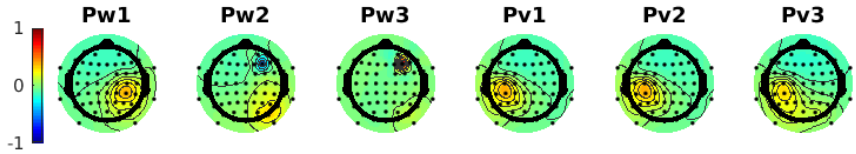
(b) P3 – CCACSP spatial patterns.



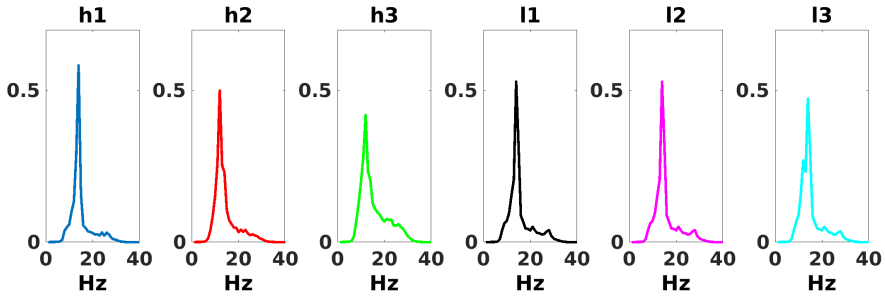
(c) P3 – specCSP spatial patterns.



(d) P3 – spec-CSP spectral filters.

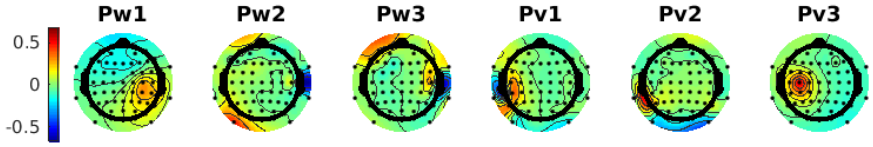


(e) P3 – SACSP spatial patterns.

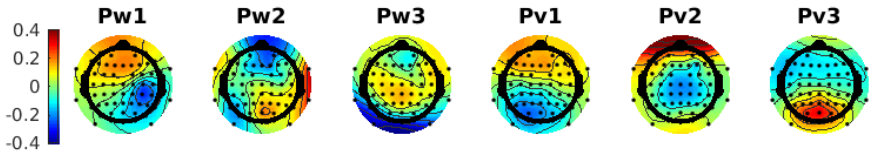


(f) P3 – SACSP spectral filters.

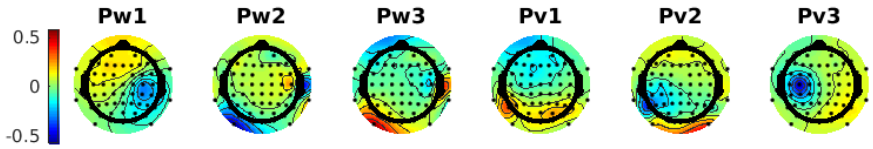
Figure 4: Spatial patterns and spectral filters trained on the calibration data for P3.



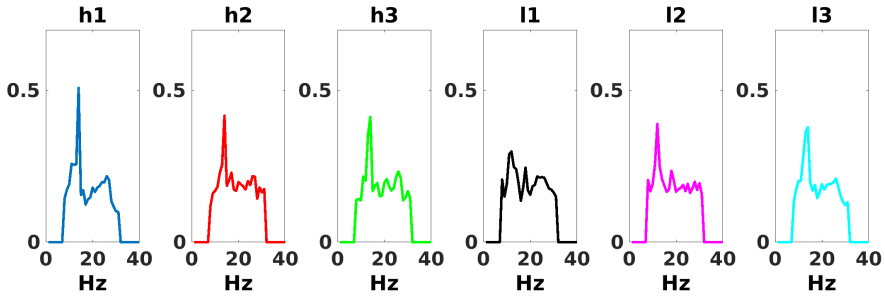
(a) P4 – CSP spatial patterns.



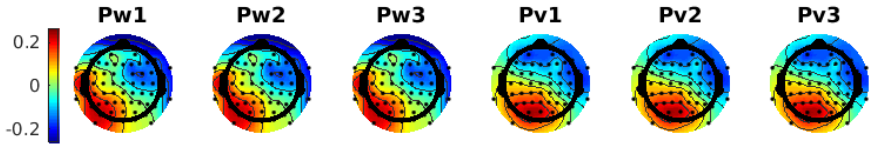
(b) P4 – CCACSP spatial patterns.



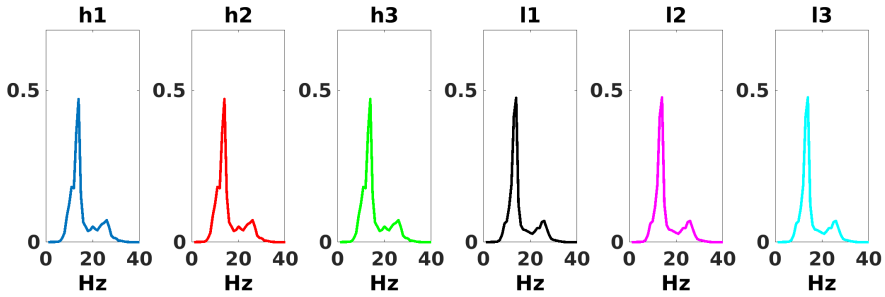
(c) P4 – spec-CSP spatial patterns.



(d) P4 – spec-CSP spectral filters.

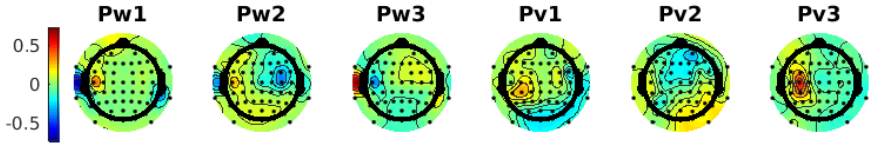


(e) P4 – SACSP spatial patterns.

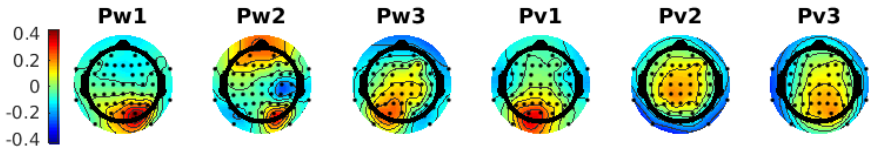


(f) P4 – SACSP spectral filters.

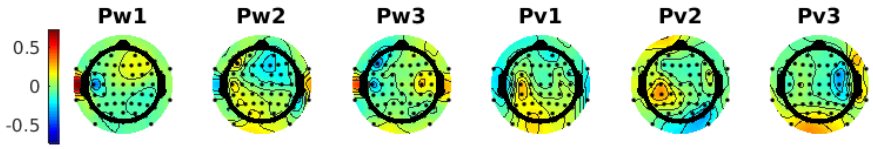
Figure 5: Spatial patterns and spectral filters trained on the calibration data for P4.



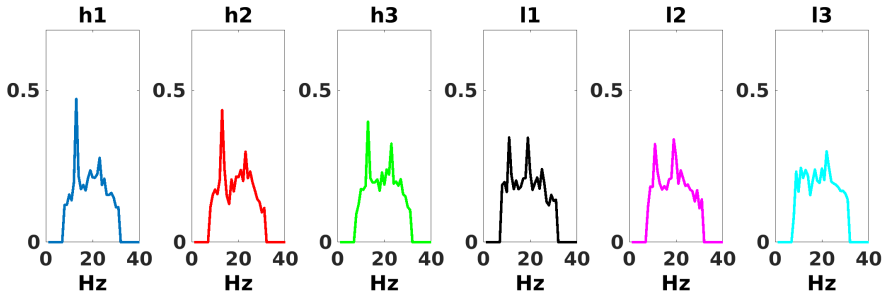
(a) P5 – CSP spatial patterns.



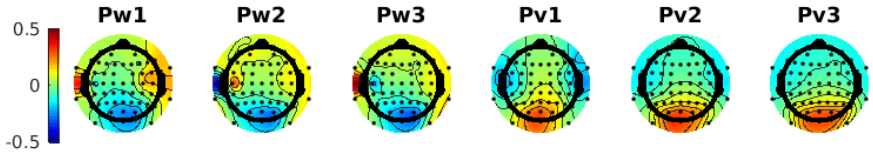
(b) P5 – CCACSP spatial patterns.



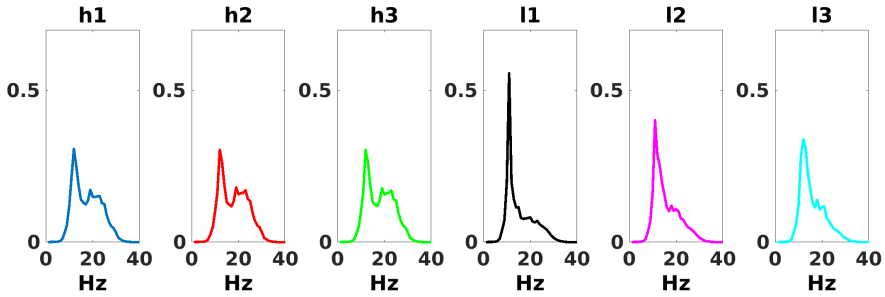
(c) P5 – spec-CSP spatial patterns.



(d) P5 – spec-CSP spectral filters.

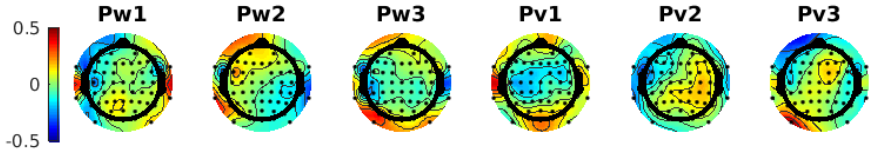


(e) P5 – SACSP spatial patterns.

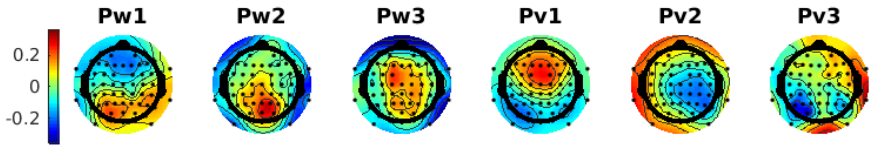


(f) P5 – SACSP spectral filters.

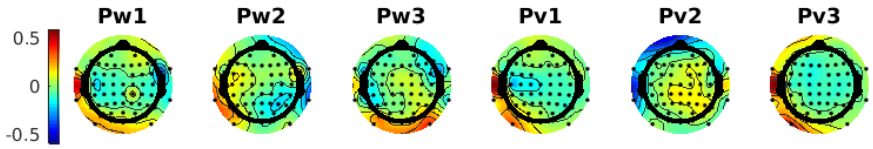
Figure 6: Spatial patterns and spectral filters trained on the calibration data for P5.



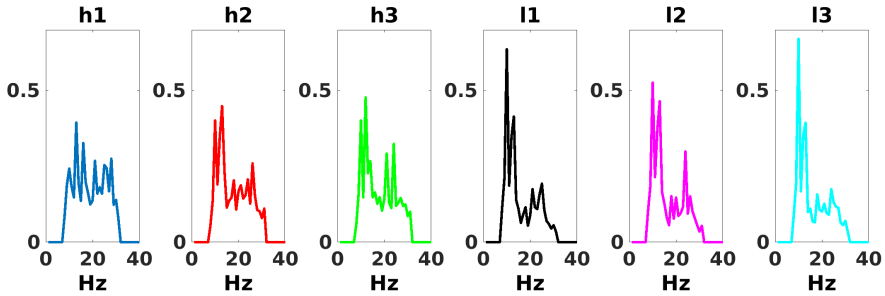
(a) P6 – CSP spatial patterns.



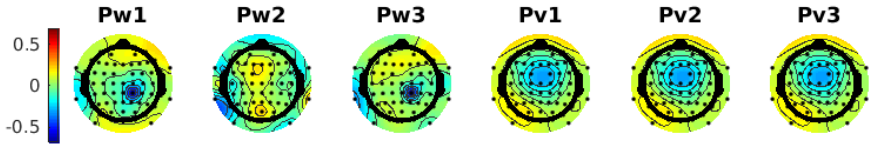
(b) P6 – CCACSP spatial patterns.



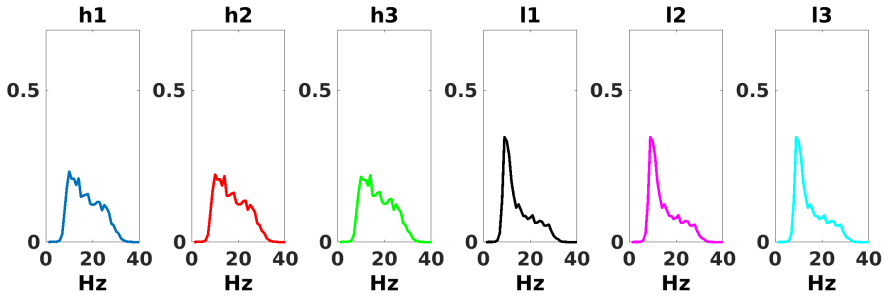
(c) P6 – spec-CSP spatial patterns.



(d) P6 – spec-CSP spectral filters.

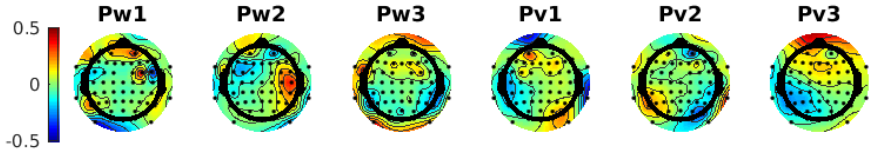


(e) P6 – SACSP spatial patterns.

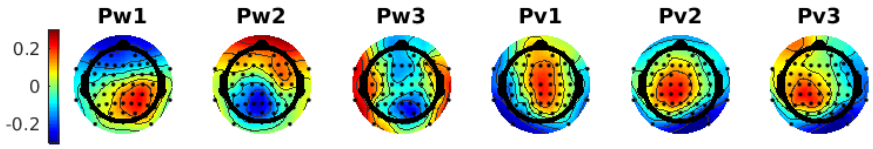


(f) P6 – SACSP spectral filters.

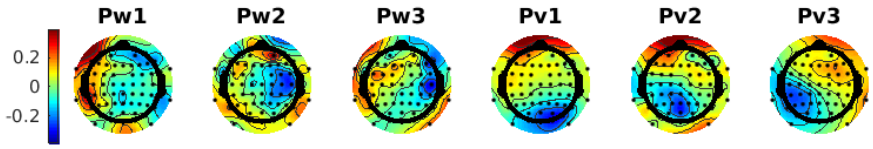
Figure 7: Spatial patterns and spectral filters trained on the calibration data for P6.



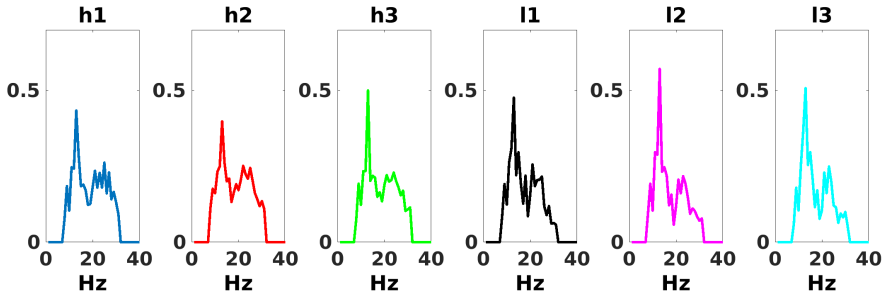
(a) P7 – CSP spatial patterns.



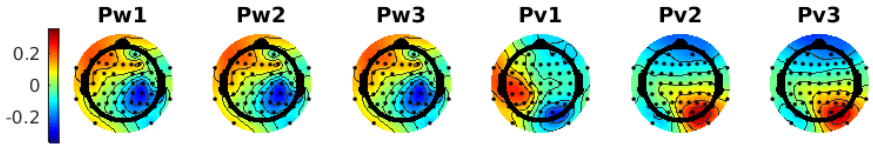
(b) P7 – CCACSP spatial patterns.



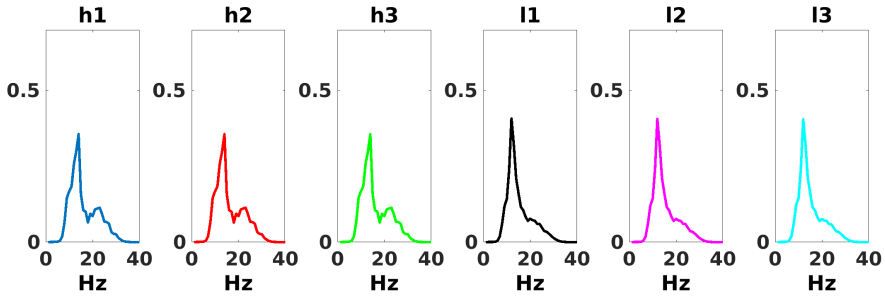
(c) P7 – spec-CSP spatial patterns.



(d) P7 – spec-CSP spectral filters.

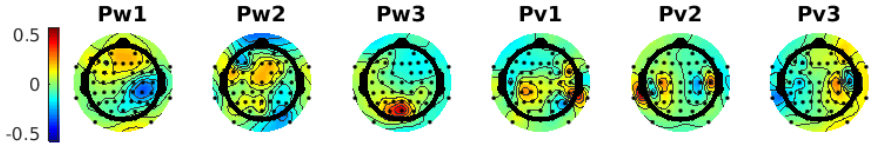


(e) P7 – SACSP spatial patterns.

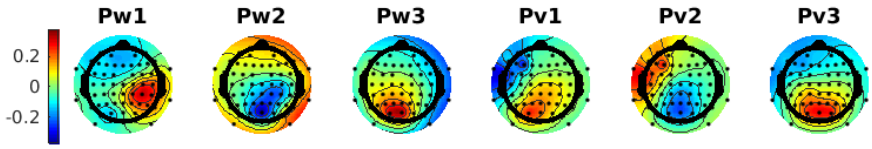


(f) P7 – SACSP spectral filters.

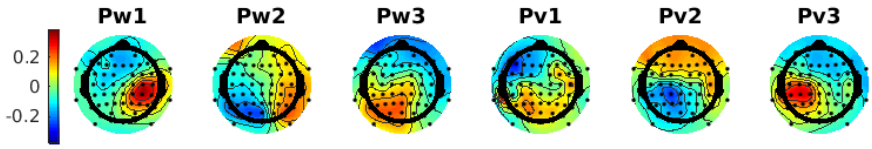
Figure 8: Spatial patterns and spectral filters trained on the calibration data for P7.



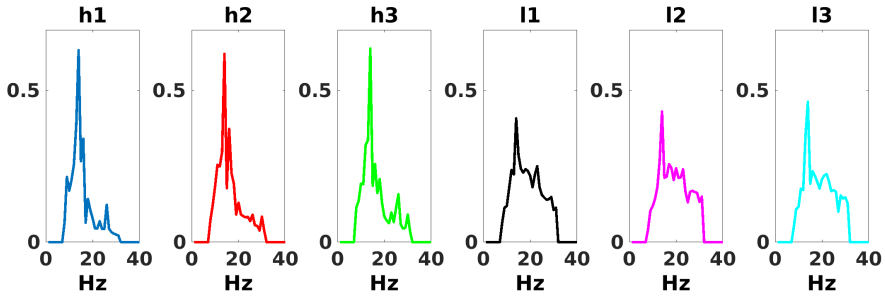
(a) P8 – CSP spatial patterns.



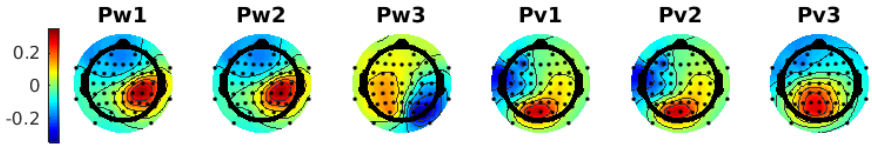
(b) P8 – CCACSP spatial patterns.



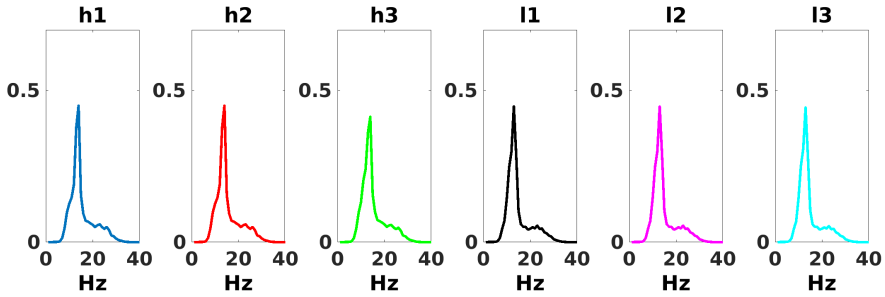
(c) P8 – spec-CSP spatial patterns.



(d) P8 – spec-CSP spectral filters.

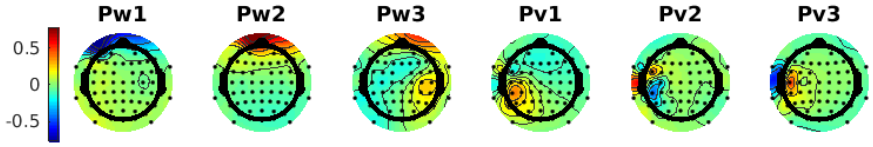


(e) P8 – SACSP spatial patterns.

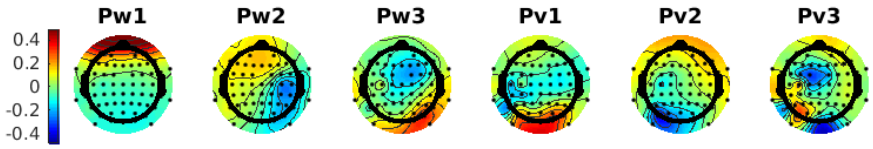


(f) P8 – SACSP spectral filters.

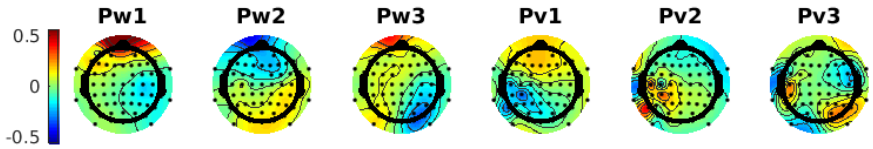
Figure 9: Spatial patterns and spectral filters trained on the calibration data for P8.



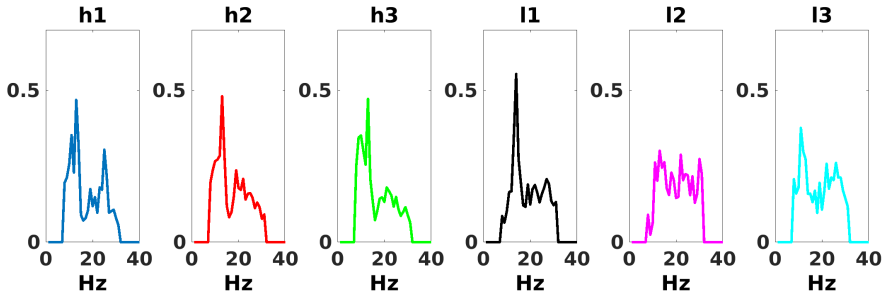
(a) P9 – CSP spatial patterns.



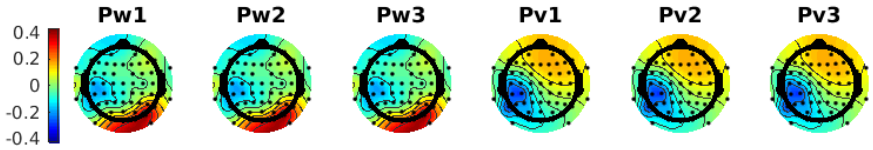
(b) P9 – CCACSP spatial patterns.



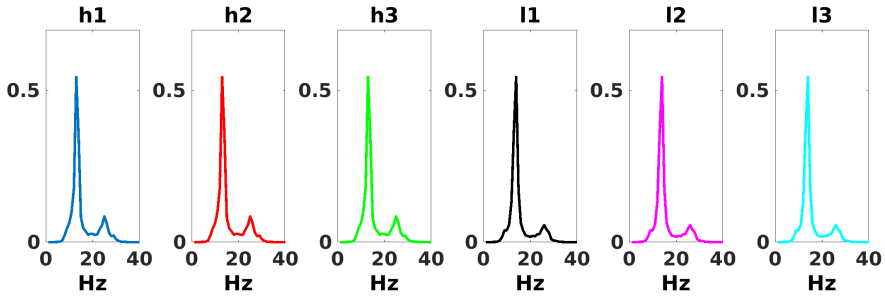
(c) P9 – spec-CSP spatial patterns.



(d) P9 – spec-CSP spectral filters.

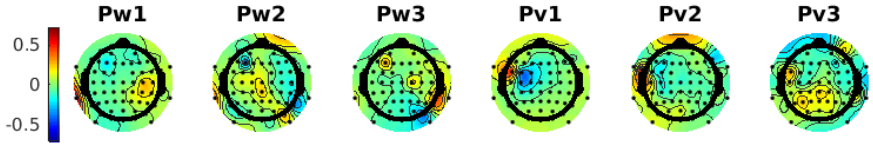


(e) P9 – SACSP spatial patterns.

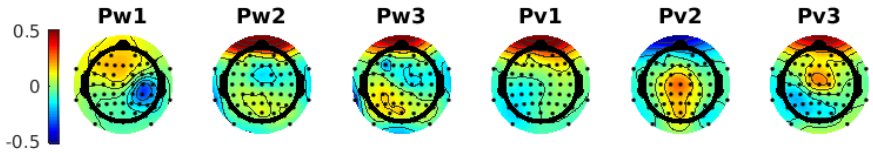


(f) P9 – SACSP spectral filters.

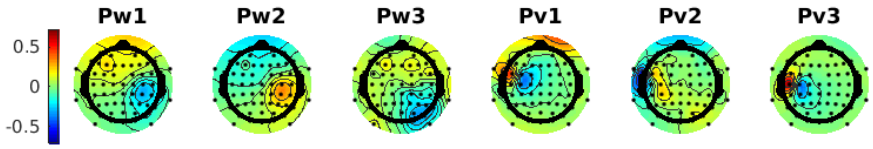
Figure 10: Spatial patterns and spectral filters trained on the calibration data for P9.



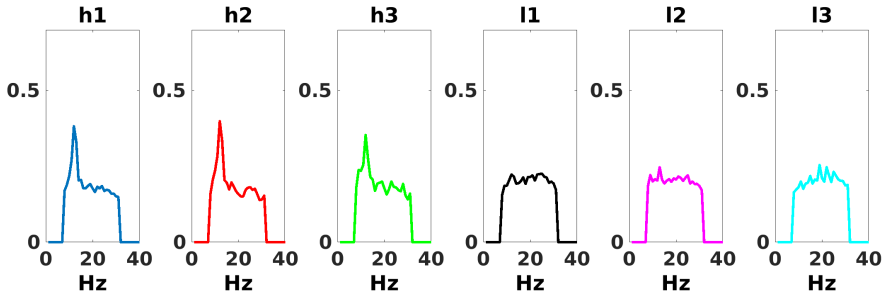
(a) P10 – CSP spatial patterns.



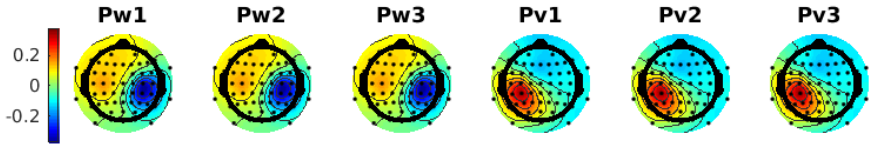
(b) P10 – CCACSP spatial patterns.



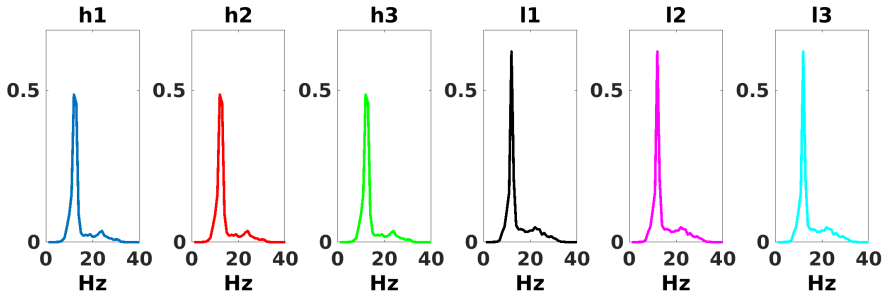
(c) P10 – spec-CSP spatial patterns.



(d) P10 – spec-CSP spectral filters.

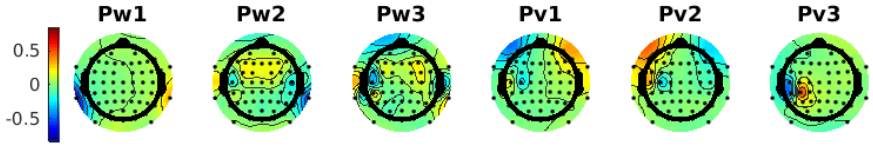


(e) P10 – SACSP spatial patterns.

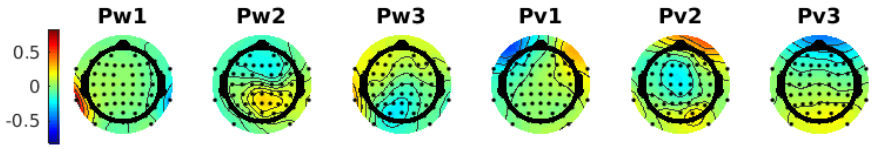


(f) P10 – SACSP spectral filters.

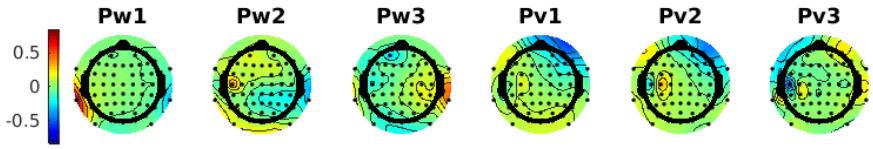
Figure 11: Spatial patterns and spectral filters trained on the calibration data for P10.



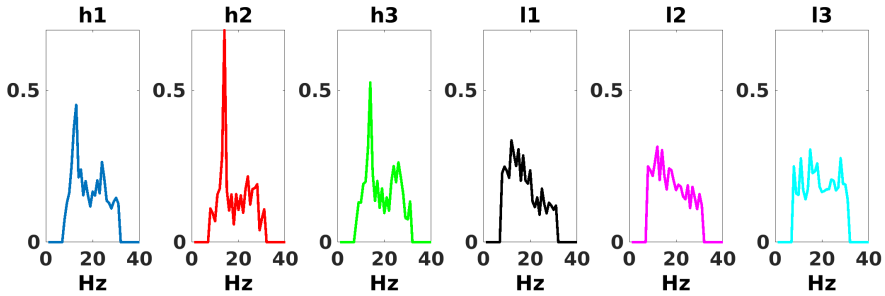
(a) P11 – CSP spatial patterns.



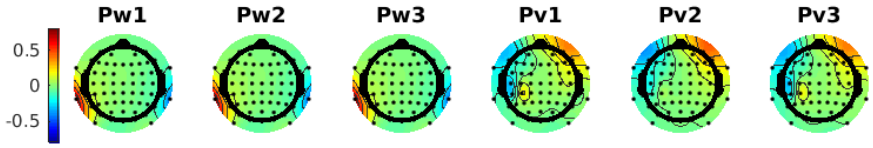
(b) P11 – CCACSP spatial patterns.



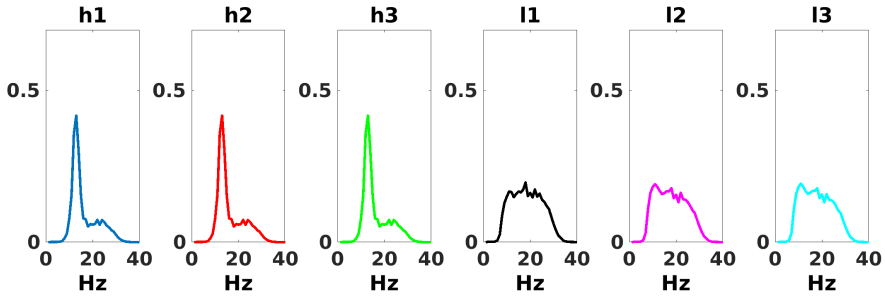
(c) P11 – spec-CSP spatial patterns.



(d) P11 – spec-CSP spectral filters.

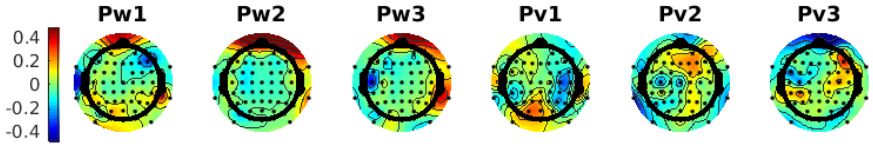


(e) P11 – SACSP spatial patterns.

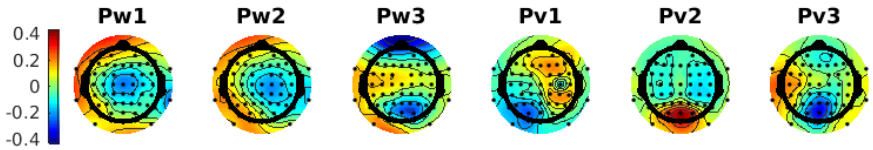


(f) P11 – SACSP spectral filters.

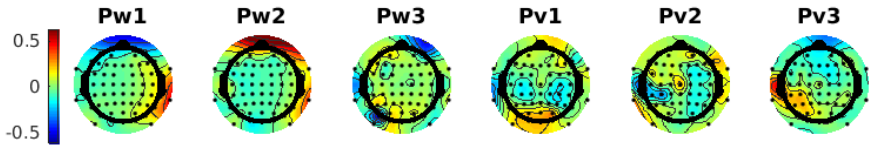
Figure 12: Spatial patterns and spectral filters trained on the calibration data for P11.



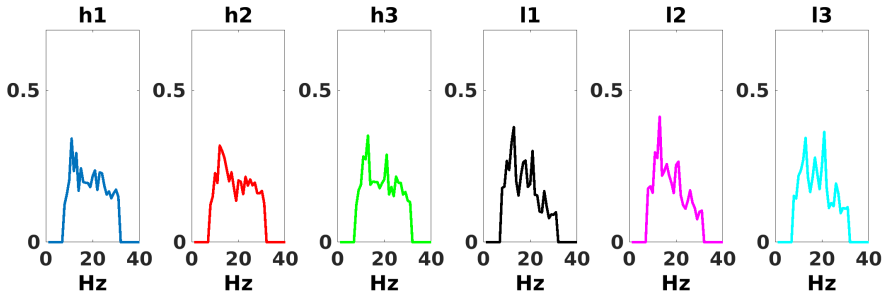
(a) P12 – CSP spatial patterns.



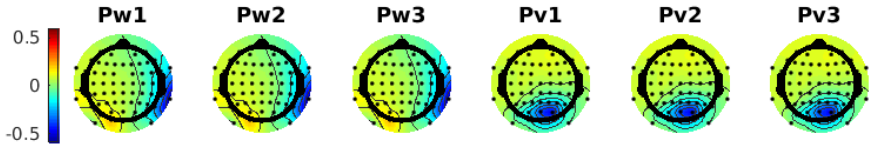
(b) P12 – CCACSP spatial patterns.



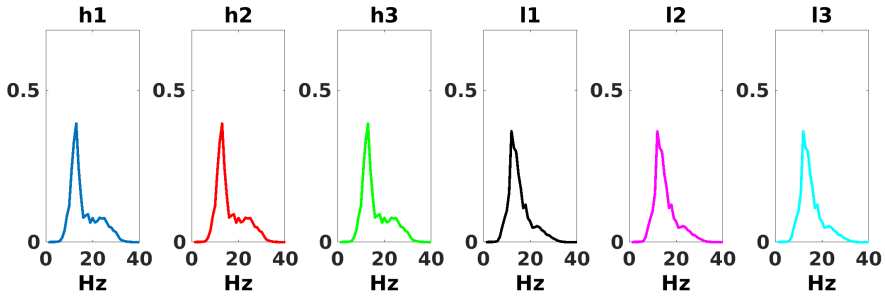
(c) P12 – spec-CSP spatial patterns.



(d) P12 – spec-CSP spectral filters.



(e) P12 – SACSP spatial patterns.



(f) P12 – SACSP spectral filters.

Figure 13: Spatial patterns and spectral filters trained on the calibration data for P12.