Increasing Session-to-Session Transfer in Brain Computer Interface with On-site Background Noise Acquisition

Hohyun Cho,1 Minkyu Ahn,2 Kiwoong Kim,3,4 and Sung Chan Jun1
1School of Information and Communications, Gwangju Institute of Science and Technology, Gwangju, South Korea 500-701
2Department of Neuroscience, Brown University, Providence, RI, U.S.A. 02912
3Center for Biosignals, Korea Research Institute of Standards and Science (KRISS), Daejeon, South Korea 305-340
4Department of Medical Physics, University of Science and Technology (UST), Daejeon, South Korea, 305-333

E-mail: scjun@gist.ac.kr

Abstract

Objective. A brain-computer interface (BCI) usually requires a time-consuming training phase during which data are collected and used to generate a classifier. Because brain signals vary dynamically over time (and even over sessions), this training phase may be necessary each time the BCI system is used, which is impractical. However, the variability in background noise, which is less dependent on a control signal, may dominate the dynamics of brain signals. Therefore, we hypothesized that an understanding of variations in background noise may allow existing data to be reused by incorporating the noise characteristics into the feature extraction framework; in this way, new session data are not required each time and this increases the feasibility of the BCI systems.

Approach. In this work, we collected background noise during a single, brief on-site acquisition session (approximately 3 minutes) immediately before a new session, and we found that variations in background noise were predictable to some extent. Then we implemented this simple session-to-session transfer strategy with a regularized spatiotemporal filter (RSTF), and we tested it with a total of 20 cross-session datasets collected over multiple days from 12 subjects. We also proposed and tested a bias correction in the RSTF. Main Results. We found that our proposed session-to-session transfer strategy yielded a slightly less or comparable performance to the conventional paradigm (each session training phase is needed with an on-site training dataset). Furthermore, using an RSTF only and an RSTF with a bias correction outperformed existing approaches in session-to-session transfers. Significance. We inferred from our results that, with an on-site background noise suppression feature extractor and pre-existing training data, further training time may be unnecessary.

1. Introduction

Zero-training, or zero-calibration in a brain-computer interface (BCI), is a strategic approach to reduce the total training costs as it consists of collecting training data (experimental cost) and generating a classifier (computational cost). This is one of the challenging issues in decoding cognitive functions, as in BCI [1–7], neuroergonomics [8–10], and others. An ideal zero-training system may be achieved by developing session-to-session (or subject-to-subject) transfer methods (subject/session independent algorithms) and
applying them to the BCI system. A session-to-session (subject-to-subject) transfer is used to train patterns from a previous session (subject) and transfer them to a new session (subject). Thus, when a pre-trained algorithm from one session (subject) works well in another session (subject) without additional training, then the transfer from one session (subject) is achieved.

It is well known that several critical factors vary across subjects, and this makes it challenging to develop methods to decode cognitive functions. For example, in motor imagery (MI) BCI, the source location of the MI, specifically the sensorimotor cortex; habits of MI, an individual’s way of thinking and imagining; and specific spectral/temporal windows are associated strongly with event-related desynchronization (ERD) [11–15]. These factors are aspects of inter-subject variability [15]. In addition, the nonstationarity of neural oscillations can yield inherent within-subject variability on a moderate time scale (session or day). This intra-subject variability [15] may originate from changes in the subject’s background state of alertness and wakefulness [12–16], evaporation and conductivity of the electrode gel, and small changes in the electrodes’ positions between sessions [12–16]. Another form of intra-subject variability—biosignal variability between the calibration and feedback phases—is also observed commonly. Eye blinking or movements, head movements, electromyograms (EMG) from facial movements, event-related potentials (ERP), and evoked potential (EP) signals are some of the main causes of intra-subject variability. Overall, inter-session, and inter-trial and intra-subject variability are common in BCI. However, this variability violates the assumption of most statistical learning algorithms and can cause covariate shifts of those results into an incorrect classification between subjects, sessions, and trials. Therefore, many methods to overcome these forms of variability have been proposed.

<table>
<thead>
<tr>
<th>Table 1. Two strategic approaches to inter-subject and intra-subject (inter-trial, inter-session) variability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inter-subject variability</strong></td>
</tr>
<tr>
<td>[17–21]</td>
</tr>
<tr>
<td><strong>Adaptive algorithm</strong></td>
</tr>
<tr>
<td>[22–27]</td>
</tr>
<tr>
<td><strong>Intra-subject variability</strong></td>
</tr>
<tr>
<td><strong>Inter-trial variability</strong></td>
</tr>
<tr>
<td>[28–33]</td>
</tr>
<tr>
<td><strong>Inter-session variability</strong></td>
</tr>
<tr>
<td>[29–31, 37]</td>
</tr>
<tr>
<td>[34, 38–41]</td>
</tr>
</tbody>
</table>

Figure 1. Venn diagram for session-to-session model in information space

Regarding inter-subject variability, two elegant categorical approaches have been proposed: robust feature extraction and an adaptive approach. Most of these strategic approaches to overcome intra- and inter-subject variability are tabulated in Table 1. The robust feature extraction approach extracts invariant or robust features from the existing subject or session data, and the adaptive approach consists of first training subject/session independent algorithms within the existing data and then adapting to new data on a real-time basis. Robust feature extraction approaches [17–21] to reduce inter-subject variability were
introduced for the purpose of increasing subject-to-subject transfer by taking into account the proper tradeoff between generality and performance. In general, BCI systems (generality) require training with a sufficiently large pool of users, but this may result in relatively lower performance by an individual user than a user-specific BCI system. Thus, increasing both generality and performance is very challenging. Alternatively, operant conditioning (allowing the user to learn) [3] or adaptive approaches [22–27] have been considered as follows:

- Establish a database.
- For new BCI users, the best similar pattern in the database is identified, and then an adaptive algorithm is applied to data from the new user.

For these approaches, a sufficiently large database that contains as many individual patterns as possible, as well as a fast-searching algorithm, should be constructed to achieve near-zero-training or zero-training. Without a variety of data, tremendous effort is required by the users, and the adaptive algorithm may be extremely intense, thereby making it difficult to achieve a fast adaption to new data.

![Spatial feature extraction](image)

### Spatial feature extraction

\[
\begin{bmatrix}
  w_1 \\
  w_2 
\end{bmatrix}^T X^{(k)} = \begin{bmatrix}
  z_1 \\
  z_2
\end{bmatrix}^{(k)} \rightarrow \log \begin{bmatrix}
  z_1 \\
  z_2
\end{bmatrix}^{(k)} = \begin{bmatrix}
  f_1 \\
  f_2
\end{bmatrix} \rightarrow \omega \begin{bmatrix}
  f_1 \\
  f_2
\end{bmatrix} + \omega_{offset} = \omega^{(k)}
\]

(a) (b) (c) (d) (e)

**Figure 2.** Conventional common spatial pattern (CSP) and Fisher’s linear discriminant analysis (FLDA) model: (a) CSP filters; (b) projected signal variance: green indicates left-hand MI and red indicates right-hand MI; (c) log variance of projected feature; (d) FLDA’s discrimination line; and (e) distributions of classifier outputs for two classes.

In this paper, we focused primarily on session-to-session transfer to achieve zero-training for an individual user. As with inter-subject variability, similar approaches can be applied to inter-session variability. To describe inter-session variability in a more systematic way, we illustrate our session-to-session model in Figure 1. Let \( F_n \) and \( B_n \) be task-related feature information and background noise information at Session \( n \) \((n = 1, 2, 3 \ldots)\), respectively. For simplicity, consider only two sessions. It may
be assumed that the two pieces of feature information (F₁ and F₂) are quite similar, given that the same MI tasks are conducted during both sessions and user-specific motor areas and their associated spectra of sensorimotor rhythms will not change substantially [11–15]. Thus, if F₁ and F₂ were substantially different, none of the statistical learning theories on session-to-session transfer would be applicable. In general, under the assumption that feature information from two sessions is similar (that is, F₁≈F₂), if background noise information is also similar (that is, B₁≈B₂), invariant feature extraction methods will work well. However, if background noise information from two sessions, B₁ and B₂, is quite different, then we expect that a classifier trained in Session 1 may yield a notably biased classification for Session 2. Thus, the session-to-session transfer will be low. We tested this hypothesis with simulated data (see the next paragraph). In contrast, adaptive algorithms try to adapt B₂ information while minimizing B₁ information and updating the algorithms for every new trial (Session 2). However, it is time-consuming to use adaptive algorithms and they do not guarantee improved performance until the adaptation is complete. Therefore, our hypothesis was that varying background noise over sessions might be a critical factor that causes a covariate shift and bias in classifier outputs in online sessions.

Here, using some typical MI data [13], we conducted a simulation study of a session-to-session transfer with a common spatial pattern (CSP) [42, 43] and Fisher’s linear discriminant analysis (FLDA), as illustrated in Figure 2. This simulation study was similar to that in Blankertz’s work [28]. We assumed that there exists background noise information (B₁) and feature information (F₁) that may be separated well in the 1st session. Thus, a classifier was constructed with the 1st session data. Then, the 2nd session data were tested with this classifier. In order to investigate a session-to-session transfer easily from the 1st to 2nd sessions, 2nd session data were generated by adding artificial noise information to the 1st session data. This yielded 2nd session data that consisted of feature information (F₂=F₁) and background noise information (B₂=B₁+αΣ), where Σ is artificial noise information (some that possibly overlaps with F₁) and α is a scaling parameter to control the amount of noise information added. Finally, 1st session data (F₁ + B₁) were training data and 2nd session (F₂ + B₂), data were testing data, where F₂=F₁ and B₂=B₁+αΣ. Figure 3 shows the results of 2nd session data over variations in α. In Figure 3, topographies in the 1st column show three types of noise information, Σ. Some Σ may overlap with feature information (F₁), which is illustrated in Figure 2(a).

Figure 3. Session-to-session transfer simulation with three different noise activation patterns for CSP and FLDA models.
Interpreting the weight vectors of linear models [44, 45], CSP filters are backwards or discriminative models used to extract latent factors in MI (Figure 2), while the principal components of three different noise realizations are forward, generative models, or activation patterns that factor the data into principal latent factors (Figure 3). This simulation demonstrated clearly that the extent of the overlapping of the spatial filter (corresponding to feature information, F_2) and noise pattern (part of the background noise information, B_2), i.e., the extent of the noise pattern passed by the spatial filter, is the key factor that produces the covariate shift and, finally, the bias in classifier outputs. The first noise signal pattern (1st row in Figure 3) overlapped with the second spatial filter w_2 (Figure 2(a)), such that the second spatial filter passes the majority of the first noise pattern. This caused the feature distribution to shift to the y-axis, which is the projected space of the second filter. Regarding the second noise pattern (2nd row in Figure 3), this overlapped with both spatial filters (w_1 and w_2 in Figure 2(a)), and its classification results shifted along both the x and y axes, and then converged. The classification results of the third noise pattern yielded only a slight shift towards the origin because the second filter passed a relatively small amount of the third noise pattern. This demonstrates that the extent of overlap of the spatial filter and the noise pattern (the extent of the noise pattern passed by the spatial filter) might provide a crucial clue about the covariate shift.

Figure 4. Comparison of conventional and proposed closed loop in BCI paradigm. A conventional feature extraction strategy is used with an offline dataset to train the algorithms and is tested in online EEG in the conventional paradigm. Our proposed strategy used a noise suppression method and on-site background noise. After the initial session, daily sessions related to on-site background noise only were used to update the filters and classifiers, thereby reducing the bias in the classification outputs.

Covariate shift was introduced by Sugiyama et al. [38]. They tried to reduce the gap between different session data distributions using an importance-weighted, cross-validation method. Unlabeled samples were collected from the first half of the online data to train the transfer algorithm. Recently, Li et al. [34] proposed a strategy to incorporate the new online session data into the training procedure in an adaptive way, which may be a method to obtain changing background noise information (B_2) from new online session data. These two approaches require subjects to undergo several training/feedback procedures.
during the adaptation phase. Adaptive algorithms [34–36, 38–41] adjust to this varying background noise (B2=B1+αΣ) in order to extract F1 and to achieve reliable classification accuracy, but this adaptation takes time, either during the adaptation phase or the test session. In contrast, invariant feature extraction methods [28–33, 37] may extract robust features from the B1+F1 successfully and yield a higher classification accuracy (e.g., 95%), as shown in Figure 2(d). However, invariant feature extraction methods were unable to guarantee a similarly high accuracy in the session-to-session transfer simulation due to undesirable noise components that appeared commonly in the online tests.

In this paper, we focused on the background noise effect in an effort to overcome session-to-session variability. Specifically, new online session data were not used for training; instead, existing data and on-site noise alone were used for this purpose. Although most studies have tried to avoid background noise, we used it here to acquire information useful in estimating the covariate shift and bias of classifier outputs. Our proposed approach was as follows: first, collect initial session data or good qualitative data to extract feature information (F1); second, collect on-site noise to obtain some of the background noise information (a part of B2) before a new session; third, regularize the algorithm using on-site noise covariance before a new session, and fourth, begin a new session with the updated classifier. Furthermore, it may be possible to estimate the potential bias of classifier outputs and correct the hyper-plane offset in the classification domain, which we address in the subsequent section. Our proposed paradigm and the conventional paradigm are illustrated in Figure 4.

![Session Diagram](image)

**Figure 5.** Detailed experimental paradigm for MI experiment. The same paradigm was used for both offline and feedback (online) phases. Subjects stared at a red dot during MI to avoid eye movement artifacts. The yellow ball always hits the gray target in the offline phase. The ball moves randomly in the background phase. In the online phase, the ball is controlled by the data collected from 0 to 2 seconds after the cue (blue).

2. **Methods**

2.1. **Experiment and materials**

Twelve healthy, male subjects (right-handed, average age = 25.6±2.6 years) participated over multiple days in two-class MI experiments (a pair of right-hand (R), left-hand (L) or foot (F) tasks). Except for one subject (s5), all were novices in MI. Sixty-four EEG electrodes (Biosemi ActiveTwo) were attached to the
scalp according to the 10-20 international system, and signals were digitized at a 512 Hz sampling rate. This experiment and the data were approved by the Institutional Review Board at Gwangju Institute of Science and Technology (GIST) and all subjects signed a written informed consent.

To implement our proposed strategy, we designed a slightly unconventional experimental paradigm for MI [42]. Our experiment consisted of 3 phases and the experimental paradigm, as illustrated in Figure 5, was applied in all phases:

- In the first phase, MI EEG data were collected in order to train filters and classifiers. During this training phase, the ball moved in one of two directions; however, the subject’s intention was not reflected in the movement of the ball. We collected a total of 60 trials for each class.
- In the second phase, on-site background noise was acquired. Before a subject proceeded to the online (feedback) phase, we recorded online background noise under conditions similar to the online phase, but the ball moved in an arbitrary direction, and we instructed the subject not to imagine limb movement. We collected background noise under nearly the same conditions as in the online MI experiment.
- In the third phase, 75 trials of online feedback data were recorded for each class (only 60 trials for s1-1 and 50 trials for s6 and s7).

Ultimately, a total of 60 offline trials, 10 background noise trials, and 75 online trials per class were collected for each session. Each subject participated in multiple sessions over the two days. Table 2 shows the subject and session numbers as, for example, s1-1 and s1-2.

**Figure 6.** Mechanism of covariate shift for each axis. Extent of overlapping area (element-wise product) between spatial filter and noise pattern indicates the covariate shift.

2.2. Covariate shift estimation and bias correction using on-site noise covariance

Based on the simulation, we found that the area of overlap between the spatial filter and noise pattern, i.e., the extent of the noise pattern passed by the spatial filter, contributed to the covariate shift for each filter, as shown in Figure 3. For the simulation study in the previous section, the extent of the overlap between
each noise pattern and spatial filter is depicted in Figure 6. Some covariate shifts may be estimated even from on-site noise before starting a new online session. $w_i^T \Xi w_i$ in Equation (1) is an ensemble average of the sum of squares of the overlapping area described in the 3rd columns of Figures 6(a) and 6(b). Thus, transforming it into the feature space with a logarithmic function yields:

$$\phi_i = \log(w_i^T \Xi w_i),$$

where $\phi_i$ is an expected shift associated with $i$th filter (i.e., along the $i$th axis) in the feature space, $w_i$ is a $i$th feature extraction filter, and $\Xi$ is on-site noise covariance that contains background information from the on-site session; thus, on-site noise information may explain a part of the overlapping area (with $i$th spatial filter) related implicitly to the covariate shift. Furthermore, a bias measure of classifier output (assuming FLDA classifier is applied) is estimated as follows:

$$\varepsilon = w_{FLDA}^T \Phi + w_{FLDA\_offset}^T,$$

where $\varepsilon$ is the expectation of bias measures of classifier outputs using on-site noise, and $\Phi$ is a vector composed of expected shifts, $\phi_i$. In general, $\varepsilon$ is expected to be approximately 0 if a classifier has no bias due to on-site noise (that is, noise and spatial filters do not overlap notably). Accordingly, $w_{FLDA\_offset}$ may be updated heuristically depending on the value of $\varepsilon$ to reduce the expected classification bias; thus, this bias-corrected classification is conducted for session-to-session transfer. We observed occasionally that the use of this bias correction value, $\varepsilon$, yielded far more biased results. Thus, to reduce this unexpected bias in the classifier outputs, rather than using $-\varepsilon$, a scaled $-C \varepsilon$ was added to $w_{FLDA\_offset}$, where $C$ is a scaling factor ranging from 0-1 and is estimated as follows:

$$C = \frac{\tilde{D}_{KL}(\Xi_{day1}, \Xi_{day2})}{\max(D_{KL}(\Xi_{day1}, I), D_{KL}(\Xi_{day2}, I))}.$$  

Here, $I$ is identity matrix, and $\Xi_{day1}$ and $\Xi_{day2}$ is the background noise covariance for each day, respectively. $\tilde{D}_{KL}(\Xi_{day1}, \Xi_{day2})$ is a symmetric Kullback Leibler (KL) distance between $\Xi_{day1}$ and $\Xi_{day2}$. When two background noise datasets from different days were similar, a small scaling value $C$ was produced, and a relatively small bias correction was applied.

### 2.3. Regularized spatio-temporal filter to minimize covariate shift

As addressed in the previous section, $w_i^T \Xi w_i$ can be interpreted as an ensemble average of a squared sum of element-wise products (if an element-wise product is visualized in the sensor space, it appears as an area of overlap (as in Figure 6) of the $i$th filter and noise; therefore, we wanted to minimize $w_i^T \Xi w_i$, the overlapping area of the $i$th filter and the noise. To do so, we proposed to solve the following two simultaneous optimization problems:

$$\max_{w_i} \left( \frac{w_i^T C_i w_i}{w_i^T C_i w_i} \right) \quad \text{and} \quad \min_{w_i} \left( w_i^T \Xi w_i \right).$$

Here, $C_i$ is the spatial covariance matrix of EEG signals belonging to class $i$, and $\{i\}^C$ is a complement of $\{i\}$. The first optimization problem (left of Equation (4)) comes to the CSP method [43]. However, in this strategy, we considered simultaneously an additional problem in minimizing the overlapping area...
associated with the covariate shift. This simultaneous optimization (in Equation (4)) is not easy to solve, or is not well defined. Thus, we adopted an idea of Blankertz et al. [28] here and incorporated the parameterized 2nd optimization problem (right of Equation (4)) into the denominator of the 1st optimization problem to yield the following optimization problem:

\[
\max_{w_{[i]}} \left\{ \frac{w_{[i]}^T C_{[i]} w_{[i]}}{(1-\xi)w_{[i]}^T C_{[i]} w_{[i]} + \xi P(w_{[i]})} \right\} \quad \text{with} \quad P(w) = w^T \Xi w.
\] (5)

Here, \( P(w) \) is a quadratic penalty function dependent on the structure of matrix \( \Xi \) and spatial filter \( w \). If the structure of \( \Xi \) and the covariance of another class \( (C_{[2]} \text{ when } i=1, \text{ or } C_{[1]} \text{ when } i=2) \) are very similar, or \( \xi \) reaches zero, then it comes to the conventional CSP formulation. This optimization (5) can be converted equivalently into a generalized eigenvalue problem:

\[
C_{[i]} w_{[i]} = \lambda ((1-\xi) C_{[i]} + \xi \Xi) w_{[i]}.
\] (6)

Unlike the conventional CSP, (6) should be solved twice (when \( i=1 \), and when \( i=2 \)). In any case, both invariant CSP (iCSP) [28] and stationary CSP (sCSP) [31] belong to this regularized framework [29].

For invariant CSP, the matrix \( \Xi \) becomes a sham feedback noise covariance to penalize the spatial filter. The sham feedback noise consists of all non-task related signals from the brain and facial movements, for example, resting state, eye blinking, eyeball movements, head movements, and so on. Therefore, matrix \( \Xi \) may be considered as a summation of all (or some) types of noise covariance matrices. In contrast, the matrix \( \Xi \) becomes on-site background noise covariance to penalize the filter.

For stationary CSP, redundant covariance matrix information in a trial from averaged covariance matrix over all trials [31] is considered to be non-stationary information. The redundant covariance can penalize the non-stationary part of the filter over all trials. Thus, it is possible that matrix \( \Xi \) is a summation of redundant information for each class of data, as below:

\[
\Xi = \overline{P}_{[1]} + \overline{P}_{[2]} \quad \text{with} \quad P^{(k)}_{[i]} = \Psi(C^{(k)}_{[i]} - C_{[i]}),
\] (7)

where \( \overline{P}_{[i]} \) is the average of the redundant information for each class, \( i=1 \text{ or } 2 \). \( C^{(k)}_{[i]} \) and \( C_{[i]} \) are the \( k \)th trial’s covariance of class \( i \) and the covariance of class \( i \), respectively. An operator, \( \Psi \), is defined as \( \Psi(M) = E \cdot \text{diag}(d_1, d_2, \ldots) \cdot E^T \) when a symmetric matrix \( M \) is decomposed into \( E \cdot \text{diag}(d_1, d_2, \ldots) \cdot E^T \). Thus, \( \Psi \) is an operator that makes the symmetric matrix positive definite.

We proposed an extension of the regularized CSP (rCSP) to consider the spatio-spectral feature space or regularized spatiotemporal filter (RSTF). The purpose of the RSTF algorithm is to suppress the spatio-spectral noise and tune the spatio-spectral filters by penalizing the spatio-spectral noise covariance.

At first, the basic concept of finite impulse response (FIR) filtering was adopted to express spectral filters. For a discrete-time FIR filter, the output is a weighted sum of the current value and a finite number of previous values of the input. Thus, the output sequence, \( y(n) \), is defined in terms of its input sequence, \( x(n) \), as follows:

\[
y(n) = b_0 x(n) + b_1 x(n-1) + \cdots + b_N x(n-N)\).
\] (8)

Here, \( x(n) \) is the input signal, \( y(n) \) is the output signal, and \( b_i \) are the filter coefficients (known as tap weights) that make up the impulse response. \( N \) is the filter order. For an \( N \)th-order filter, output can be
expressed by \((N + 1)\) terms. Note that the \(x(n-j)\) \((j=0,1,2,...,N)\) are referred to commonly as taps, which are based on the structure of a tapped delay line that provides the delayed inputs to the multiplication operations in block diagrams. Therefore, for the case of \(N=5\), it may be called an 8th order filter or 9-tap filter.

For \(d\) channels, the following expressions are obtained from (8):

\[
y_i(n) = b_{0,i}x_i(n) + b_{1,i}x_i(n-1) + \cdots + b_{N,i}x_i(n-N), \quad (i = 1, 2, \ldots, d)
\]  

(9)

Here, \(x_i(n)\) and \(y_i(n)\) are input and output values for channel \(i\), respectively. \(b_{k,i} \ (k=0,1,\ldots,N, \ i=1,2,\ldots,d)\) are coefficients of FIR filters. Assuming a spatial filter \(\Gamma = \{\gamma_{i}\}_{i=1}^{d}\) is given, applying this spatial filter \(\omega\) to (9) yields:

\[
\begin{align*}
&w_1y_1(n) = w_1b_{0,1}x_1(n) + w_1b_{1,1}x_1(n-1) + \cdots + w_1b_{N,1}x_1(n-N) \\
&w_2y_2(n) = w_2b_{0,2}x_2(n) + w_2b_{1,2}x_2(n-1) + \cdots + w_2b_{N,2}x_2(n-N) \\
&\vdots \\
&w_dy_d(n) = w_db_{0,d}x_d(n) + w_db_{1,d}x_d(n-1) + \cdots + w_db_{N,d}x_d(n-N).
\end{align*}
\]  

(10)

Letting \(\gamma_i(k) = w_i\beta_{k,i} \ (k=0,1,\ldots,N, \ i=1,2,\ldots,d)\), (9) can be re-expressed in matrix form, as follows:

\[
w^T Y(n) = \Gamma(0)^T X(n) + \Gamma(1)^T X(n-1) + \cdots + \Gamma(N)^T X(n-N). \tag{11}
\]

Here, \(Y(n) = [y_1(n) \ y_2(n) \ \cdots \ y_d(n)]^T\) and \(X(k) = [x_1(k) \ x_2(k) \ \cdots \ x_d(k)]^T\) represent the vectors of output and input signals, respectively. \(\Gamma(k) = [\gamma_1(k) \ \gamma_2(k) \ \cdots \ \gamma_d(k)]^T\) represents a vector of coefficients \(\gamma(k)\). Concatenating vectors \(\Gamma(k)\) into a single vector, \(\Gamma\), (11) can be represented as:

\[
w^T Y(n) = \Gamma^T \begin{pmatrix} X(n) \\ X(n-1) \\ \vdots \\ X(n-N) \end{pmatrix}, \quad \Gamma = \begin{pmatrix} \Gamma(0) \\ \Gamma(1) \\ \vdots \\ \Gamma(N) \end{pmatrix}. \tag{12}
\]

From (12), \(\Gamma\) (a time embedding of \(N\)th order to the original signal) can be recognized as a spatio-spectral filter. Application of this idea to the rCSP framework can then be formulated as:

\[
\max_{\gamma^{(1,2)}} \left( \Gamma_{(i)}^T \hat{\mathcal{C}}_{(i)} \Gamma_{(i)} \right) \left( (1-\xi) \Gamma_{(i)}^T \hat{\mathcal{C}}_{(i)} \Gamma_{(i)} + \xi P(\Gamma_{(i)}) \right) \quad \text{with} \quad P(\Gamma) = \Gamma^T \hat{\Sigma} \Gamma \quad \text{and} \quad i \in \{1,2\} \tag{13}
\]

Here, \(\hat{\mathcal{C}}_{(i)}\) is a spatio-spectral covariance matrix of EEG signals, \(\hat{X}_{(i)}\), belonging to class \(I\), where \(\hat{X}_{(i)}\) is defined as:

\[
\hat{X}_{(i)} = \begin{pmatrix} X_{(i)}(n) \\ X_{(i)}(n-1) \\ \vdots \\ X_{(i)}(n-N) \end{pmatrix} = \begin{pmatrix} X_{(i)} \\ \delta^1 X_{(i)} \\ \vdots \\ \delta^N X_{(i)} \end{pmatrix}, \quad \text{where} \quad \delta^k X := X(n-k) \tag{14}
\]

Furthermore, \(\hat{\Sigma}\) can be given as the spatial-spectral noise covariance, which is defined similarly to \(\hat{\mathcal{C}}_{(i)}\),
and thus the optimization problem (13) can be transformed into a generalized eigenvalue problem:

$$
\hat{C}_i^r \Gamma_i^p = \lambda ((1-\xi)\hat{C}_i^r + \hat{\Xi}) \Gamma_i^p, \quad i \in \{1,2\}.
$$

(15)

Similar to (15), spatio-spectral filtering approaches using an FIR filter under the CSP optimization framework were proposed, such as the common spatio-spectral pattern (CSSP) algorithm [46], and the common sparse spatio-spectral pattern (CSSSP) algorithm [47]. CSSP considers limited FIR coefficients ($\Gamma(0)$ and $\Gamma(N)$) only, without a penalty function $P(\Gamma)$. CSSSP considers general FIR coefficients without a penalty function, but yields the same spectral filter for all channels. Our proposed algorithm (RSTF) is a unified framework that considers both CSSP and CSSSP as special cases.

Our goal in this work was to propose a new experimental paradigm and to investigate its feasibility and efficacy relative to the conventional paradigm. Therefore, an RSTF was introduced as a feature extractor. For simplicity, the RSTF was confined to the following case:

- Consider FIR coefficients ($\Gamma(0)$ and $\Gamma(N)$) only by setting $\Gamma(k)=0$ for $k=1,2,\ldots,N-1$.
- Penalty function $P(\Gamma)$ is defined by the spatio-spectral on-site background noise covariance $\hat{\Xi}$.

2.4. Evaluations

2.4.1. Comparison between the conventional and proposed paradigms

Whole data were filtered both spectrally (8–30 Hz) and temporally (0–2 seconds after cue onset). Next, an RSTF trained with pre-chosen training data and session-related background noise was applied to other sessions, after which the online and offline feedback performances were measured. For our proposed paradigm, the noise covariance in formula (15) was calculated with 20 trials of background noise.

To evaluate our proposed session-to-session transfer concept, we used an RSTF, FLDA, and bias correction and compared the classification accuracies of the conventional and proposed paradigms (Figure 4). For the conventional paradigm, the online data were classified by the pre-generated classifier (RSFT feature extractor and FLDA classifier) trained with offline data from the same session. However, our proposed paradigm tested the online data using the algorithms trained with the offline data from the initial session and the on-site background noise data. For example, in the conventional paradigm, online data from s1-2 were classified by the algorithm that was trained with s1-2’s offline data. However, for our proposed paradigm, online data from s1-2 were classified by the algorithm that was trained with s1-1’s (initial session) offline data and on-site background noise collected just prior to s1-2’s online data. Comparative classification results of the conventional and proposed paradigms over data from multiple sessions with 12 subjects (N = 20 sessions) are tabulated in Table 2. Session data for each subject were clustered depending on the class of condition used in the experiments. For example, s3-1 to s3-3 were session data for LH vs. both feet (BF) MI for subject 3, and s2-4 to s2-5 were for RH vs. BF MI.

To select the hyper-parameters ($N$ or $\xi$), we down-sampled EEG data to a 128-Hz sampling rate and then performed 120 iterations of cross-validation. We divided the 60 trials of training data for each class into 10 subsets of 6 trials each. Seven subsets were chosen randomly and used to train algorithms and the remaining 3 subsets were used to test them. This procedure was repeated 120 times by choosing 7 among 10 subsets randomly. Finally, 120 classification accuracies were estimated and averaged for each hyper-parameter. $\xi$ was optimal for the set $\{0.01, 0.02, \ldots, 0.1\}$, and maintained a minimal noise effect in the RSTF feature extractor. For the RSTF, we selected a filter order $N$ among the set $\{1, 3, 5, \ldots, 15\}$ by referring to the literature [46], and then selected $\xi$ for a given optimal $N$.

2.4.2. Comparative study of session-to-session transfers

To investigate the session-to-session transfers for various existing feature extractors, as well as our proposed RSTF, we compared them with CSP, sCSP, and ICSP. Although CSP and sCSP feature extractors
do not consider noise information, we tested them to investigate whether or not feature extractors using noise information show relatively effective performance. For session-to-session transfers, all feature extractors were trained with offline (initial) data and tested on different online session data. In addition, experimental paradigms that used either on-site background noise or offline background noise were applied to iCSP and the RSTF. From these, we expected to see that on-site background noise plays an important role in increasing session-to-session transfer. The paradigms used were as follows:

- **CSP (baseline):** CSP was used for feature extraction. A classifier was trained with offline data from the training session and tested with online data from the testing session.
- **iCSP with offline:** iCSP was used for feature extraction with offline background noise. A classifier was trained with offline data from the training session and tested with online data from the testing session.
- **iCSP with on-site:** iCSP was used for feature extraction with on-site background noise. A classifier was trained with offline data from the training session and tested with online data from the testing session.
- **sCSP:** sCSP was used for feature extraction. A classifier was trained with offline data from the training session and tested with online data from the testing session.
- **RSTF with offline:** An RSTF was used for feature extraction with offline background noise. A classifier was trained with offline data from the training session and tested with online data from the testing session.
- **RSTF with on-site:** An RSTF was used for feature extraction with on-site background noise. A classifier was trained with offline data from the testing session and tested with online data from the testing session.
- **RSTF with on-site and BC:** A bias correction (BC) was added to the RSTF with on-site background noise. Using the on-site background noise covariance and the KL distance between offline and on-site background noise, a scaled bias was expected and the FLDA offset of the RSTF with on-site background noise was corrected.

All comparative performance results for the 7 types of paradigms (according to feature extractor and offline/on-site noise) over a total of 20 sessions of data are tabulated in Table 3. Scatter plots for comparison of paired paradigms are illustrated in Figure 7. One-way paired student’s t-tests were conducted and p-values are shown for each plot. No correction was performed here. Statistically significant pairs ($p < 0.01$ marked with ** or $p < 0.05$ with *) are indicated in red stars.
Table 2. Comparison between conventional and proposed experimental paradigms. The first column is subject and session index, the second column indicates class pair, the third column shows intervals between 1st and other sessions, and the fourth column indicates online feedback results (actual feedback) from the conventional paradigm using an RSTF and FLDA in the initial session: (a) conventional paradigm using the same session data only, and (b) our proposed paradigm (session-to-session transfer) using an RSTF with session-related on-site noise and bias correction. LH=left hand, RH=right hand, BF=both feet.

<table>
<thead>
<tr>
<th>Subject-Session</th>
<th>Class</th>
<th>Day intervals from 1st session</th>
<th>Performance of initial session</th>
<th>(a) Conventional paradigm</th>
<th>(b) Proposed paradigm</th>
</tr>
</thead>
<tbody>
<tr>
<td>s1-1</td>
<td>(LH,RH)</td>
<td>0</td>
<td>82.0%</td>
<td>86.4%</td>
<td>91.4%</td>
</tr>
<tr>
<td>s1-2</td>
<td>(LH,RH)</td>
<td>7</td>
<td></td>
<td>84.7%</td>
<td>89.3%</td>
</tr>
<tr>
<td>s1-3</td>
<td>(LH,RH)</td>
<td>155</td>
<td></td>
<td>92.7%</td>
<td>92.7%</td>
</tr>
<tr>
<td>s1-4</td>
<td>(LH,RH)</td>
<td>157</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s2-1</td>
<td>(RH,BF)</td>
<td>0</td>
<td>96.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s2-2</td>
<td>(RH,BF)</td>
<td>15</td>
<td></td>
<td>83.3%</td>
<td>85.3%</td>
</tr>
<tr>
<td>s2-3</td>
<td>(RH,BF)</td>
<td>163</td>
<td></td>
<td>88.7%</td>
<td>84.0%</td>
</tr>
<tr>
<td>s2-4</td>
<td>(LH,BF)</td>
<td>0</td>
<td>94.7%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s2-5</td>
<td>(LH,BF)</td>
<td>155</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s3-1</td>
<td>(RH,BF)</td>
<td>0</td>
<td>90.7%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s3-2</td>
<td>(RH,BF)</td>
<td>15</td>
<td></td>
<td>70.7%</td>
<td>68.0%</td>
</tr>
<tr>
<td>s3-3</td>
<td>(RH,BF)</td>
<td>16</td>
<td></td>
<td>57.3%</td>
<td>43.3%</td>
</tr>
<tr>
<td>s3-4</td>
<td>(LH,BF)</td>
<td>0</td>
<td>85.3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s3-5</td>
<td>(LH,BF)</td>
<td>3</td>
<td></td>
<td>85.3%</td>
<td>68.0%</td>
</tr>
<tr>
<td>s4-1</td>
<td>(LH,BF)</td>
<td>0</td>
<td>88.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s4-2</td>
<td>(LH,BF)</td>
<td>3</td>
<td></td>
<td>98.0%</td>
<td>80.7%</td>
</tr>
<tr>
<td>s4-3</td>
<td>(LH,BF)</td>
<td>148</td>
<td></td>
<td>98.7%</td>
<td>92.7%</td>
</tr>
<tr>
<td>s4-4</td>
<td>(LH,BF)</td>
<td>149</td>
<td></td>
<td>99.3%</td>
<td>89.3%</td>
</tr>
<tr>
<td>s5-1</td>
<td>(LH,BF)</td>
<td>0</td>
<td>90.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s5-2</td>
<td>(LH,BF)</td>
<td>20</td>
<td></td>
<td>96.7%</td>
<td>88.7%</td>
</tr>
<tr>
<td>s6-1</td>
<td>(LH,BF)</td>
<td>0</td>
<td>88.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s6-2</td>
<td>(LH,BF)</td>
<td>1</td>
<td></td>
<td>73.0%</td>
<td>89.0%</td>
</tr>
<tr>
<td>s7-1</td>
<td>(RH,BF)</td>
<td>0</td>
<td>59.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s7-2</td>
<td>(RH,BF)</td>
<td>9</td>
<td></td>
<td>87.0%</td>
<td>60.0%</td>
</tr>
<tr>
<td>s8-1</td>
<td>(RH,BF)</td>
<td>0</td>
<td>77.3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s8-2</td>
<td>(RH,BF)</td>
<td>1</td>
<td></td>
<td>76.7%</td>
<td>63.3%</td>
</tr>
<tr>
<td>s9-1</td>
<td>(LH,BF)</td>
<td>0</td>
<td>62.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s9-2</td>
<td>(LH,BF)</td>
<td>7</td>
<td></td>
<td>61.3%</td>
<td>66.7%</td>
</tr>
<tr>
<td>s10-1</td>
<td>(LH,BF)</td>
<td>0</td>
<td>92.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s10-2</td>
<td>(LH,BF)</td>
<td>7</td>
<td></td>
<td>99.3%</td>
<td>98.7%</td>
</tr>
<tr>
<td>s11-1</td>
<td>(LH,BF)</td>
<td>0</td>
<td>95.3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s11-2</td>
<td>(LH,BF)</td>
<td>1</td>
<td></td>
<td>69.3%</td>
<td>80.0%</td>
</tr>
<tr>
<td>s12-1</td>
<td>(RH,BF)</td>
<td>0</td>
<td>93.3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s12-2</td>
<td>(RH,BF)</td>
<td>1</td>
<td></td>
<td>62.7%</td>
<td>82.7%</td>
</tr>
</tbody>
</table>

Mean 83.0% 79.5%
Median 85.9% 83.3%
Standard deviation 13.4% 13.9%
Table 3. Comparison between conventional methods (CSP, iCSP, and sCSP) and proposed RSTF and RSTF with bias correction in terms of session-to-session transfer performance. To show the noise suppression effect, offline (training session) and on-site background noise were compared as well.

<table>
<thead>
<tr>
<th>Training session</th>
<th>Testing session (baseline)</th>
<th>CSP (with offline)</th>
<th>iCSP (with offline)</th>
<th>iCSP (with on-site)</th>
<th>sCSP</th>
<th>RSTF (with offline)</th>
<th>RSTF (with on-site)</th>
<th>RSTF with on-site and BC</th>
</tr>
</thead>
<tbody>
<tr>
<td>s1-1</td>
<td>s1-2</td>
<td>80.0%</td>
<td>60.0%</td>
<td>91.4%</td>
<td>60.0%</td>
<td>50.7%</td>
<td>89.3%</td>
<td>91.4%</td>
</tr>
<tr>
<td>s1-3</td>
<td>61.3%</td>
<td>59.3%</td>
<td>81.3%</td>
<td>52.7%</td>
<td>70.7%</td>
<td>50.0%</td>
<td>82.7%</td>
<td>89.3%</td>
</tr>
<tr>
<td>s1-4</td>
<td>78.7%</td>
<td>74.7%</td>
<td>88.7%</td>
<td>80.7%</td>
<td>50.0%</td>
<td>93.3%</td>
<td>92.7%</td>
<td></td>
</tr>
<tr>
<td>s2-1</td>
<td>s2-2</td>
<td>50.0%</td>
<td>50.0%</td>
<td>79.3%</td>
<td>72.7%</td>
<td>50.0%</td>
<td>86.7%</td>
<td>85.3%</td>
</tr>
<tr>
<td>s2-3</td>
<td>50.0%</td>
<td>50.0%</td>
<td>82.7%</td>
<td>53.3%</td>
<td>50.0%</td>
<td>80.7%</td>
<td>84.0%</td>
<td></td>
</tr>
<tr>
<td>s2-4</td>
<td>s2-5</td>
<td>50.0%</td>
<td>50.0%</td>
<td>65.3%</td>
<td>70.0%</td>
<td>54.0%</td>
<td>70.0%</td>
<td>76.7%</td>
</tr>
<tr>
<td>s3-1</td>
<td>s3-2</td>
<td>64.7%</td>
<td>76.0%</td>
<td>70.7%</td>
<td>75.3%</td>
<td>56.7%</td>
<td>67.3%</td>
<td>68.0%</td>
</tr>
<tr>
<td>s3-3</td>
<td>50.7%</td>
<td>54.7%</td>
<td>50.0%</td>
<td>48.7%</td>
<td>57.3%</td>
<td>49.3%</td>
<td>43.3%</td>
<td></td>
</tr>
<tr>
<td>s3-4</td>
<td>s3-5</td>
<td>67.3%</td>
<td>56.7%</td>
<td>60.0%</td>
<td>68.0%</td>
<td>66.7%</td>
<td>64.7%</td>
<td>68.0%</td>
</tr>
<tr>
<td>s4-1</td>
<td>s4-2</td>
<td>95.3%</td>
<td>84.7%</td>
<td>92.7%</td>
<td>87.3%</td>
<td>82.0%</td>
<td>80.0%</td>
<td>80.7%</td>
</tr>
<tr>
<td>s4-2</td>
<td>92.7%</td>
<td>92.7%</td>
<td>79.3%</td>
<td>91.3%</td>
<td>83.3%</td>
<td>90.7%</td>
<td>92.7%</td>
<td></td>
</tr>
<tr>
<td>s4-4</td>
<td>65.3%</td>
<td>90.0%</td>
<td>93.3%</td>
<td>87.3%</td>
<td>51.3%</td>
<td>88.7%</td>
<td>89.3%</td>
<td></td>
</tr>
<tr>
<td>s5-1</td>
<td>s5-2</td>
<td>51.3%</td>
<td>50.7%</td>
<td>87.3%</td>
<td>62.7%</td>
<td>52.0%</td>
<td>92.7%</td>
<td>88.7%</td>
</tr>
<tr>
<td>s6-1</td>
<td>s6-2</td>
<td>81.0%</td>
<td>84.0%</td>
<td>84.0%</td>
<td>79.0%</td>
<td>87.0%</td>
<td>88.0%</td>
<td>89.0%</td>
</tr>
<tr>
<td>s7-1</td>
<td>s7-2</td>
<td>62.0%</td>
<td>59.0%</td>
<td>54.0%</td>
<td>62.0%</td>
<td>62.0%</td>
<td>63.0%</td>
<td>60.0%</td>
</tr>
<tr>
<td>s8-1</td>
<td>s8-2</td>
<td>60.0%</td>
<td>61.3%</td>
<td>59.3%</td>
<td>64.0%</td>
<td>60.0%</td>
<td>63.3%</td>
<td>63.3%</td>
</tr>
<tr>
<td>s9-1</td>
<td>s9-2</td>
<td>52.0%</td>
<td>66.7%</td>
<td>50.0%</td>
<td>56.7%</td>
<td>60.7%</td>
<td>66.0%</td>
<td>66.7%</td>
</tr>
<tr>
<td>s10-1</td>
<td>s10-2</td>
<td>96.0%</td>
<td>96.0%</td>
<td>92.7%</td>
<td>97.3%</td>
<td>95.3%</td>
<td>96.7%</td>
<td>98.7%</td>
</tr>
<tr>
<td>s11-1</td>
<td>s11-2</td>
<td>78.7%</td>
<td>80.7%</td>
<td>79.3%</td>
<td>82.7%</td>
<td>79.3%</td>
<td>80.0%</td>
<td>80.0%</td>
</tr>
<tr>
<td>s12-1</td>
<td>s12-2</td>
<td>66.7%</td>
<td>72.7%</td>
<td>70.7%</td>
<td>77.3%</td>
<td>80.0%</td>
<td>81.3%</td>
<td>82.7%</td>
</tr>
<tr>
<td>Mean</td>
<td>67.7%</td>
<td>68.5%</td>
<td>75.6%</td>
<td>71.5%</td>
<td>65.0%</td>
<td>78.7%</td>
<td>79.5%</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>65.0%</td>
<td>64.0%</td>
<td>79.3%</td>
<td>71.3%</td>
<td>60.3%</td>
<td>81.0%</td>
<td>83.3%</td>
<td></td>
</tr>
<tr>
<td>Standard deviation</td>
<td>15.6%</td>
<td>15.5%</td>
<td>14.7%</td>
<td>13.9%</td>
<td>14.6%</td>
<td>12.9%</td>
<td>13.9%</td>
<td></td>
</tr>
</tbody>
</table>

3. Results

As shown in Table 2, the average of the initial session performance (4th column) was 85.3% (SD = 11.7%); the majority of the initial performances using an RSTF were far higher than chance level [48]. For comparison of the conventional and proposed paradigms, the mean performances were 83.0% and 79.5% (medians = 85.9% and 83.3%, SDs = 13.4% and 13.9%), respectively. We expected that the conventional approach would yield uniformly better performance than ours. For some session data, however, the accuracy of our approach exceeded that of the conventional (e.g., s1-2, s1-3, s2-2, s6-2, s9-2, s11-2, and s12-2). Online data from s3-3 were at chance levels, and thus, we did find lower performance in some cases. Although the overall performance of the conventional paradigm was better than that of our proposed paradigm, the differences were not statistically significant. Thus, we concluded that the conventional and proposed paradigms showed comparable performance.
For comparison of session-to-session transfers over various feature extractors (Table 3 and Figure 7), the “RSTF with on-site and BC” outperformed other methods in classification accuracy. Specifically, both feature extractors (iCSP, RSTF) that used on-site background noise showed substantially higher session-to-session transfers (> 4%) than when offline background noise was used. With the exception of iCSP with offline background noise, CSP and sCSP, feature extractors used widely in BCI systems, yielded relatively inferior session-to-session transfer than feature extractors using our proposed paradigm (iCSP with on-site noise, an RSTF with on-site noise, and an RSTF with on-site and BC). Furthermore, it was notable that the noise suppression effect with offline background noise did not yield any significant improvement in session-to-session transfers.

**Figure 7.** Scatter plots for session-to-session transfer performance in classification accuracy. A total 21 pairs from 7 approaches were compared. Statistically significant pairs are marked with * (p < 0.05) and ** (p < 0.01).
With offline background noise update

(a) With offline background noise update

(b) With on-site background noise update

(C) With on-site background noise update and bias correction

Figure 8. To investigate the effect of updating the pre-trained filters and classifiers and bias correction with on-site noise data, we plotted the classifier outputs for all subjects. The x-axis indicates the whole number of online trials. Red and green dots indicate classifier outputs of different classes. The classification accuracy of entire online trials was estimated as in the titles of the plots. Dotted blue lines in each plot indicate boundaries in different subjects’ online data. The sequence of each subject’s data is as in the second column in Table 3. Degrees of bias are represented for each data point and calculated as the percentage of classification points of both red and green classifier outputs and 100 (y-axis) values. Red-shaded values indicate the mean of the degrees of bias defined in Equation (16). (a) RSTF and FLDA classifiers were updated by offline background noise; (b) RSTF and FLDA classifiers were updated with on-site noise using optimal $\xi_0$; (c) RSTF and FLDA classifiers were updated with on-site noise and a bias correction was applied to adjust FLDA offset.

4. Discussion

4.1. Two conventional paradigms and the proposed experimental paradigm

Conventional and proposed paradigms can be considered as intra- and inter-session paradigms. Generally, decrements in classification accuracy exist in inter-session performance [20, 49]; thus, the majority of cases decrease in session-to-session transfer, i.e., on the whole, our proposed paradigm yielded inferior accuracy to the conventional paradigm. However, as stated above, some cases yielded increased session-to-session transfers (Table 2). We found from simple performance analysis classification that the accuracies of offline data in s11-2 and s12-2 were below 70%, which is relatively low for BCI performance; thus, it may be that the offline data (training data) for s11-2 and s12-2 lacked discriminative
information, and thus failed to train good filters. However, initial offline data (e.g., s11-1, s12-1) may be quite good, in that they yielded reasonable session-to-session transfers.

For s3-3 and s9-2, neither the conventional nor the proposed paradigms produced good results. In these cases, simple analysis revealed that classification performances for their online data were at around chance levels, which tells us that the data did not discriminate well and that even good spatial filters could not extract discriminative features from these data.

4.2. Reduction in covariate shift using background noise
To investigate how session-related (on-site) background noise is important in transferring offline data to online data (inter-session analysis), we compared two RSTF cases (with offline and on-site noise) in the classification domain. Figure 8 shows a scatter plot of the classification domain of two conditioned trials (one is denoted by red, the other by green crosses) over 20 online sessions. To measure the degree of bias of the classifier outputs, the distance between the discriminative line of the classifier (horizontal line y=0) and discriminating center of clustering (dcc) was estimated for each session of data. Here, dcc was estimated as follows:

$$\text{dcc} = \frac{m_2 - m_1}{\sigma_1 + \sigma_2} \sigma_1 + m_1$$

where $m_{[1,2]}$ and $\sigma_{[1,2]}$ are the mean and standard deviation of classifier outputs corresponding to each condition (condition 1 or 2: red or green). This distance represents the way in which a classifier yields biased results. We depicted the degrees of classification bias as dcc values, as shown in Figure 8. The averages of the classification accuracies were 65.0%, 78.7%, and 79.5%, respectively (Table 3), and the averages of the dcc values were 20.6, 5.4, and 5.1, respectively. Even in the comparison of dcc values, an RSTF with on-site noise and bias correction demonstrated better performance in session-to-session transfers.

4.3. Bias correction for classifier outputs
To reduce the possible significant covariate shift in the feature domain, a bias correction approach can be introduced using $\varepsilon$ in equation (2). By adding $-C\varepsilon$ to the FLDA offset, a discriminative line in the classification domain is corrected. We applied this bias correction strategy to our proposed RSTF with on-site noise to test whether or not classification performance would improve significantly, as in Figure 8(c). Unfortunately, this bias correction strategy produced a minor improvement in performance; however, it was not statistically significant. We are now investigating a more finely tuned bias correction. As in Figure 1 (session-to-session model), we wanted to estimate the covariance of $B_2$, but actually the covariance estimated was derived from $B_2$ and nonstationary noise. We thought this non-stationary noise or small number of samples might yield the biased covariance of $B_2$, which could give incorrect results. The estimation of a well-conditioned covariance matrix of $B_2$ is quite important. Shrinkage techniques [29, 50, 51] may be applicable to resolve this problem, which is also under investigation.

4.4. Toward zero-training
We demonstrated that performance with the conventional and proposed paradigms different. The conventional paradigm usually requires a lengthy amount of time (approximately 10~30 minutes) to collect training data for every session [52]; in contrast, our proposed paradigm requires less than 3 minutes to collect on-site background noise, after only a single set of training data is collected. This is comparable to a co-adaptive framework, which requires approximately 15 minutes adapting the subject-independent classifier to the individual user [27]. Furthermore, even a single set of training data may be unnecessary if pre-existing data are available. In our proposed paradigm, a classifier that simply updates with regularized feature extractors (such as iCSP, RSTF, etc.) using on-site background noise (session-
related background noise) may improve the ability to reuse data, while controlling classification performance. Thus, we believe that our proposed paradigm will have advantages over the conventional paradigm in that training time is reduced significantly, thereby increasing the practicality of the BCI system. Ultimately, our strategy may achieve nearly zero training in some respects. Initial collection of good quality data or the availability of good pre-existing data is of great importance to our proposed paradigm, as without the use of qualitative data, it is likely to yield inferior performance in session-to-session transfers. Furthermore, because background noise changes more rapidly, even on-site background noise may not provide reasonable noise characteristics with online data; in this case, our proposed paradigm may yield a little better-than-expected performance.

4.5. Session-to-session transfers using on-site noise information

Our proposed paradigm can introduce any feature extractors that use background noise information within their optimization frame. For a simple comparative study of feature extractors with our proposed paradigm, we compared the session-wise transfer performances of methods that use on-site background noise (Table 3). Figure 7 shows the distribution of our data with respect to classification accuracy. The “RSTF with on-site and BC” showed the highest accuracy (median) and was significantly different ($p < 0.05$ in Student’s t-test) from other existing methods (except for an RSTF with on-site background noise). This demonstrates that session-to-session transfers can be achieved moderately well with on-site background noise and initial session data.

Comparing sCSP and iCSP with our concept provided a pure comparison of session-to-session transfers between conventional robust feature extraction techniques and our proposed paradigm. First, sCSP and iCSP are variants of the regularized CSP algorithm, which is a type of spatial filter. Second, sCSP filters are trained with offline data from the source session and tested with online data from the target session (Table 3), and iCSP filters are trained with offline data from the source session and background noise data from the target session, and then are tested with online data from the target session. The results of the comparison between sCSP and iCSP according to our concept showed that there was no significant difference between the two methods ($p > 0.05$); however, iCSP with our concept showed both better average and median performances. This suggests that our proposed experimental paradigm provides a very simple solution for zero training that may work as well as more complex methods, such as sCSP and other variants [21, 29]. Interestingly, for cases s1-2 and s1-3, iCSP and an RSTF with on-site noise, which has a background noise attenuation term, showed improved performance by comparison to sCSP, which indicates that it is important to consider background noise.

Although the RSTF method with bias correction was comparable to the conventional paradigm, some limitations remain. For example, s2-5, s3-5, s4-2, s7-2, and s8-2 showed over a 10% decrement in performance by comparison to the conventional paradigm. We believe that our proposed paradigm may be embedded in state-of-the-art methods, such as divCSP frameworks [21, 53], 1-norm CSP [32], and sparse representation approaches [33], among others. This may yield improved session-to-session transfers. We will continue to investigate which method is the best in which to embed our approach.

4.6. Generative model of RSTF

As shown in Table 3, the RSTF variants yielded significantly higher performance than did the iCSP variants ($p < 0.05$). We know that an RSTF is inherently able to extract spectral, as well as temporal information, while iCSP can extract only temporal information. Accordingly, we inferred that spectral information may help improve classification accuracy and an increase in the rate of session-to-session transfers. In previous work by our research team, the performance of the generative model (as formulated in (13)) was evaluated from 52 MI datasets [52]. We observed that the addition of more than two FIR coefficients did not yield statistically significant improvement in classification performance and the session-to-session transfer rate. MI-related ERD patterns in temporal or frequency space are quite simple
and are seen easily in the user-specific band. Thus, a higher-order RSTF (more than two coefficients) may produce an overfitting problem in optimization of the spectral filters of MI data, so that, in addition to being the simplest case, consideration of two FIR coefficients ($\Gamma(0)$, $\Gamma(N)$) also may be optimal in MI data. An RSTF using FIR coefficients (two or more) may be useful in other applications, for example, steady state somatosensory evoked potential (SSSEP) [54] and auditory steady state response (ASSR) [55]. These data may contain multi-modal spectral features and the potential capability of our generalized RSTF would be revealed.

5. Conclusions
We proposed an experimental paradigm (updating a filter and classifier with pre-existing data and on-site background noise only) to reduce training time and overcome inter-session variability. We investigated the feasibility of our proposed paradigm with 20 cross sessions of MI data collected on multiple days from 12 subjects. Our findings may be summarized as follows:

- Our proposed paradigm, which replaced the time-consuming training phase with brief collection of on-site background noise, was comparable in performance to the conventional paradigm (updating a classifier with training data collected in every session); this is quite promising in the sense that near-zero training may be possible with this strategy.
- An RSTF that considered both spatial and temporal information and attenuated noise effects outperformed other feature extractors in reducing the bias of classifier output for data from different days. Thus, it improved the session-to-session transfer rate.
- A bias correction strategy, in which the extent of bias in classifier outputs estimated from on-site background noise was applied to correct the hype-plane offset, reduced performance variation over sessions moderately well. A more finely tuned strategy may be highly beneficial in BCI systems.

6. Acknowledgments
This work is supported by the National Research Foundation of Korea (NRF-2013R1A1A2009029), the Korea Research Institute of Standards and Science (KRISS-14011024), and Ministry of Culture, Sports and Tourism (MCST) and Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program 2015.

7. References


Summary (New York, NY; Springer)


[16] Sanei S and Chambers J A 2013 EEG signal processing (John Wiley & Sons)


[18] Lotte F and Guan C 2010 Learning from other subjects helps reducing brain-computer interface calibration time Int. Conf. Audio, Speech and Signal Processing (ICASSP’2010) pp 614–7


[22] Fazli S, Popescu F, Danóczy M, Blankertz B, Müller K-R and Grozea C 2009 Subject-independent mental state classification in single trials *Neural Netw.* 22 1305–12


[25] Tu W and Sun S 2012 A subject transfer framework for EEG classification *Neurocomputing* 82 109–16


[34] Li X, Guan C, Zhang H, Ang K K and Ong S H 2014 Adaptation of motor imagery EEG classification model based on tensor decomposition *J. Neural Eng.* 11 056020


[52] Ahn M, Hong J H and Jun S C 2012 Feasibility of approaches combining sensor and source features

