



Spectral Transfer Learning Using Information Geometry for a User-Independent Brain-Computer Interface

Nicholas R. Waytowich^{1,2*†}, Vernon J. Lawhern^{1,3†}, Addison W. Bohannon^{1,4}, Kenneth R. Ball¹ and Brent J. Lance¹

¹ Human Research and Engineering Directorate, US Army Research Laboratory, Aberdeen Proving Ground, MD, USA, ² Department of Biomedical Engineering, Columbia University, New York, NY, USA, ³ Department of Computer Science, University of Texas at San Antonio, San Antonio, TX, USA, ⁴ Applied Mathematics, Statistics, and Scientific Computation Program, University of Maryland at College Park, College Park, MD, USA

OPEN ACCESS

Edited by:

Mikhail Lebedev,
Duke University, USA

Reviewed by:

Chadwick Boulay,
Keio University, Japan
David Thomas Bundy,
University of Kansas Medical Center,
USA
Anirban Dutta,
Leibniz-Institut für Arbeitsforschung an
der TU Dortmund, Germany

*Correspondence:

Nicholas R. Waytowich
nrw2106@columbia.edu;
nick.waytowich@gmail.com

[†]These authors have contributed
equally to this work.

Specialty section:

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

Received: 29 March 2016

Accepted: 06 September 2016

Published: 22 September 2016

Citation:

Waytowich NR, Lawhern VJ,
Bohannon AW, Ball KR and Lance BJ
(2016) Spectral Transfer Learning
Using Information Geometry for a
User-Independent Brain-Computer
Interface. *Front. Neurosci.* 10:430.
doi: 10.3389/fnins.2016.00430

Recent advances in signal processing and machine learning techniques have enabled the application of Brain-Computer Interface (BCI) technologies to fields such as medicine, industry, and recreation; however, BCIs still suffer from the requirement of frequent calibration sessions due to the intra- and inter-individual variability of brain-signals, which makes calibration suppression through transfer learning an area of increasing interest for the development of practical BCI systems. In this paper, we present an unsupervised transfer method (spectral transfer using information geometry, STIG), which ranks and combines unlabeled predictions from an ensemble of information geometry classifiers built on data from individual training subjects. The STIG method is validated in both off-line and real-time feedback analysis during a rapid serial visual presentation task (RSVP). For detection of single-trial, event-related potentials (ERPs), the proposed method can significantly outperform existing calibration-free techniques as well as outperform traditional within-subject calibration techniques when limited data is available. This method demonstrates that unsupervised transfer learning for single-trial detection in ERP-based BCIs can be achieved without the requirement of costly training data, representing a step-forward in the overall goal of achieving a practical user-independent BCI system.

Keywords: unsupervised learning, ensemble learning, calibration-free BCI, P300, RSVP

1. INTRODUCTION

Brain-Computer Interfaces (BCIs) are augmentative devices that decode user intent directly from the brain (Wolpaw et al., 2002). Recent advances in signal processing and machine learning techniques have enabled the application of BCI technologies to fields such as medicine, industry, and recreation (Blankertz et al., 2010; Lance et al., 2012; van Erp et al., 2012). Despite the potential of this recent progress, performance decrements due to intra- and inter-individual variability require most BCI systems to employ time-consuming and costly calibration sessions for each new user and

session thus decreasing the overall utility of current BCI systems (Johnson et al., 2011). The use of advanced signal processing and machine learning techniques to minimize or eliminate this need for calibration data is an area of on-going interest for the development of practical BCI systems (Lotte, 2015).

Information geometry is one of many promising techniques for robust brain signal detection and classification. The overarching idea being to map the signal into a Riemannian manifold, such as the cone of symmetric positive definite matrices, in which non-Euclidean statistical inference can be performed (Harandi et al., 2012; Jayasumana et al., 2015). With the development of a computational framework for statistical inference (Pennec et al., 2006), information geometry has been successfully applied to radar signal processing (Barbaresco, 2008), diffusion tensor imaging (Fletcher and Joshi, 2004), and computer vision (Tuzel et al., 2007). Recently, information geometry has been applied successfully to BCI paradigms such as motor imagery (Barachant et al., 2012, 2013) and event related potential (Barachant and Congedo, 2014).

In parallel to this development, advanced machine learning methods such as active learning and transfer learning have sought to reduce or eliminate the need for calibration in BCI systems. Transfer learning, for example, addresses how models and parameters can be used in domains for which they were not initially learned, enabling robust models with little or no new training data. Although there has been much recent work in BCI transfer (see Lotte, 2015) very few general purpose methods for zero-calibration BCI currently exist in the literature. Two successful implementations require task-specific structure and do not generalize to other paradigms. Kindermans et al. achieve zero-calibration BCI on a P300 Speller using a language model to simultaneously inform a traditional Bayesian inference classifier, but this method does not generalize to paradigms which do not lend themselves to hidden Markov process modeling or multiple trial classification (Kindermans et al., 2014). Additionally, zero-calibration methods for Steady-State Visually Evoked Potential (SSVEP) BCIs have been developed that correlate multichannel EEG with sinusoidal reference signals using Canonical Correlation Analysis (CCA) (Bin et al., 2009); however, the CCA method is mainly utilized for SSVEP BCIs due to the well-characterized steady-state frequency response and thus does not generalize well to other BCIs.

A recent review categorizes general purpose machine learning approaches to address calibration in BCI systems as either zero-calibration (methods that utilize zero prior training data) or calibration reduction (methods that utilize a limited or reduced amount of training data) (Lotte, 2015). Zero-calibration methods utilize existing models built from previously collected data to classify new data. Thus, zero-calibration methods are able to immediately classify data from a new subject without requiring a calibration session to collect labeled training data. Zero-calibration methods are further classified as either ensemble or pooled methods based on how they use existing training data. Ensemble methods group data into session and subject-specific classifiers which can make

independent classification on a new test subject. These classifiers are treated as bases for classifying a new subject, and thus methods differ in how the independent classifications are combined. Fazli et al. presented a zero-calibration ensemble method which uses an ℓ_2 regression with an ℓ_1 penalty to enforce sparsity in the linear combination of ensemble classifiers (Fazli et al., 2011). This method is shown to out-perform an unweighted sum of the classifiers. Pooled methods combine all existing data in a single training set and learn a classification model from the aggregated data, and Lotte et al. compared various zero-calibration pooled methods in Lotte et al. (2009).

Unlike the zero-calibration methods, there has been more progress in calibration reduction with the development of machine learning approaches that utilize minimal training data. These calibration reduction techniques achieve performance comparable with conventional calibration techniques. Lotte and Guan demonstrated P300 Speller classification with minimal calibration data by simply applying regularization to a common calibration-based BCI algorithm (Lotte and Guan, 2009). Metzen et al. present a user-to-user ensemble method in which they augment the method from Fazli et al. (2011) with calibration data (Metzen et al., 2011). Dalhoumi et al. (2015) propose an ensemble method which weights the individual classifiers according to their performance accuracy on a small set of calibration data (Dalhoumi et al., 2014, 2015), and this method is shown to out-perform another recent ensemble transfer method from Tu and Sun which uses not only a static ensemble of classifiers but also adapts a dynamic ensemble based on calibration data (Tu and Sun, 2012). In a pooled transfer approach, Jayaram et al. used a multi-task learning method to learn Gaussian priors to estimate optimal ensemble weights for classification (Jayaram et al., 2016). While all of these approaches offer improvements in the field of BCI transfer learning, all still require the use of calibration data, and those that do not are either limited to a specific BCI paradigm or suffer from large performance decreases.

This paper aims to eliminate the need for calibration and develop a user-independent BCI by proposing a new method, Spectral Transfer with Information Geometry (STIG), which leverages an ensemble of information geometric classifiers coupled with spectral-meta learning (SML), an unsupervised ensemble method for inferring the appropriate weights for a linear combination of classifiers (Barachant et al., 2012; Parisi et al., 2014). The STIG method uses an ensemble of subjects trained with Minimum Distance to Riemannian Mean classifiers (MDRM) to make predictions on the test data of a new subject which are linearly combined with spectral meta-learning. We compare competing methods and show that the STIG method out-performs existing zero-calibration methods such as unweighted sum or ℓ_1 -regularized regression as well as several recent calibration reduction methods with limited calibration data. The STIG method is then validated in an on-line rapid serial visual presentation (RSVP) experiment with single-trial ERP detection where real-time feedback is given to the user.

2. UNSUPERVISED ENSEMBLE TRANSFER LEARNING

STIG is a real-time, unsupervised ensemble transfer method which dynamically combines independent models based on information geometry to classify streaming data for BCI applications. For our application, we have paired the spectral ensemble method introduced in Parisi et al. (2014) with an information geometric classifier, MDRM, based on its superior performance on ERP-based BCIs (Barachant and Congedo, 2014), but this framework could work with any feature extraction and classification method. Off-line, models are trained using existing data from different subjects and sessions. These models make real-time predictions on the trials of a new subject, and the decisions of each model are dynamically combined according to the inferred accuracy of each model.

2.1. Classifier

MDRM learns a map, $f : X \rightarrow Y$, which assigns a target or non-target label, $y \in \{-1, 1\}$, to each EEG epoch, $\mathbf{X} \in \mathbb{R}^{C \times N}$ (C channels and N time samples). The classification works in feature space and thus, MDRM comprises a feature extraction step, a map $H : X \rightarrow \Sigma$, and a classification step, a map $G : \Sigma \rightarrow Y$, such that the overall classification algorithm is the composition of feature extraction and classification, $f = G \circ H$.

2.1.1. Information Geometric Features of Single Trial P300 ERP

Let $S = \{(\mathbf{X}_i, y_i)\}_{i=1}^n$ be a labeled training set of n trials for a single subject. Define an extended trial,

$$\tilde{\mathbf{X}}_i = \begin{bmatrix} \mathbf{P} \\ \mathbf{X}_i \end{bmatrix} = [\tilde{\mathbf{x}}_i^{(1)} | \tilde{\mathbf{x}}_i^{(2)} | \dots | \tilde{\mathbf{x}}_i^{(N)}], \quad (1)$$

in which $\mathbf{P} = E[\mathbf{X}|y = 1]$ is a prototypical target response. The covariance across time of the extended trial yields the Extended Sample Covariance Matrix (ESCM):

$$\Sigma_i = \frac{1}{N-1} \sum_{k=1}^N (\tilde{\mathbf{x}}_i^{(k)} - \hat{\mu}_{\tilde{\mathbf{x}}})(\tilde{\mathbf{x}}_i^{(k)} - \hat{\mu}_{\tilde{\mathbf{x}}})^T, \quad (2)$$

where T is the usual matrix transpose and $\hat{\mu}_{\tilde{\mathbf{x}}} = E[\tilde{\mathbf{x}}]$ is the sample extended trial column mean. This signal processing map is defined by the prototypical target response, $H = H(\mathbf{P})$.

2.1.2. Minimum Distance to Riemannian Mean

The Minimum Distance to Riemannian Mean classification algorithm (MDRM) assigns each trial to the nearest class mean. With the ESCM, embedded in a differentiable manifold, the cone of symmetric positive definite matrices, $S_+ \subset \mathbb{R}^{2C \times 2C}$, we can induce a notion of distance by selecting a metric in the associated tangent space (Moakher and Batchelor, 2006; Pennec et al., 2006). As in Barachant and Congedo (2014), we will use the Affine Invariant Riemannian Metric (AIRM) to define distance between any two ESCM:

$$\delta(\Sigma_i, \Sigma_j) = \|\log(\Sigma_i^{-1/2} \Sigma_j \Sigma_i^{-1/2})\|_F, \quad (3)$$

where F indicates the Frobenius norm. Since a covariance matrix is a translation-invariant transformation, the AIRM, which is invariant to linear transformations, provides an invariant measure of distance over all affine transformations of EEG epochs. This facilitates the robustness of the learning algorithm to rotations, scaling, and translations. With this distance function, the geometric mean of each class, ω_ℓ , on S_+ is

$$\Sigma^{(\ell)} = \arg \min_{\Sigma \in S_+} \sum_{i \in I_\ell} \delta^2(\Sigma, \Sigma_i), \quad (4)$$

where $I_\ell = \{i \in 1, \dots, n | y_i = \omega_\ell\}$. Therefore, each ESCM, Σ , can be classified according to a mapping defined by its class means $G = G(\Sigma^{(-1)}, \Sigma^{(+1)})$:

$$g(\Sigma) = \arg \min_{\ell \in \{-1, 1\}} \delta(\Sigma, \Sigma^{(\ell)}). \quad (5)$$

2.2. Unsupervised Ensemble Learning

As introduced in Parisi et al. (2014), SML provides an unsupervised method to combine the classification decisions of independent models to classify a previously collected and unlabeled data set. It attempts to weight each classifier's decision according to the inferred accuracy of the direct classification. Here, it is applied dynamically to streaming data for BCI applications where the independent models are trained with existing data.

2.2.1. Spectral Transfer

Let $\{f_i\}_{i=1}^m$ be an ensemble of m conditionally independent classifiers trained on m different training subjects and $\{\mathbf{X}_j\}_{j=1}^n$ be independent and identically distributed test trials for classification. Then, $F(\mathbf{X}_j) = (f_1(\mathbf{X}_j), f_2(\mathbf{X}_j), \dots, f_m(\mathbf{X}_j))^T$. Also, define the balanced accuracy of classifier i , $\pi_i = \frac{1}{2}(\psi_i + \eta_i)$, which accounts for class imbalance by giving equal weight to sensitivity, ψ , and specificity, η . The sample covariance matrix, $\mathbf{Q} \in \mathbb{R}^{m \times m}$,

$$\mathbf{Q} = \frac{1}{n-1} \sum_{j=1}^n (F(\mathbf{X}_j) - E[F(\mathbf{X})])(F(\mathbf{X}_j) - E[F(\mathbf{X})])^T, \quad (6)$$

approximates a rank-one matrix, $\mathbf{Q} \approx \lambda \mathbf{v} \mathbf{v}^T$, for which $v_i \propto (2\pi_i - 1)$ (Parisi et al., 2014). This implies that the entries of the principal eigenvector of \mathbf{Q} will be proportional to the balanced accuracy of the associated classifier.

Under a Maximum Likelihood Estimation (MLE), the most likely label, \hat{y} , for a trial, \mathbf{X} , is then

$$\hat{y} = \text{sign} \left(\sum_{i=1}^m f_i(\mathbf{X}) \left(\log \left(\frac{\psi_i \eta_i}{(1 - \psi_i)(1 - \eta_i)} \right) + \log \left(\frac{\psi_i(1 - \psi_i)}{\eta_i(1 - \eta_i)} \right) \right) \right), \quad (7)$$

This provides a natural Expectation-Maximization (EM) extension (Dawid and Skene, 1979). The expectation step consists of calculating the respective sensitivity, $\psi = P(f(\mathbf{X}) = 1 | y = 1)$, and specificity, $\eta = P(f(\mathbf{X}) = -1 | y = -1)$ of each classifier with

the MLE estimates of Equation (7). The maximization step then re-calculates the MLE estimates according to Equation (7). For a first order solution, and/or an initialization of EM, a Taylor Series approximation about $(\psi, \eta) = (1/2, 1/2)$ yields

$$\hat{y} \approx \text{sign} \left(\sum_{i=1}^m f_i(\mathbf{X}) \cdot v_i \right). \quad (8)$$

2.2.2. Real-Time Implementation

For streaming data, the weights of the ensemble classifiers can be dynamically updated with the addition of each trial. Naively, \mathbf{Q} can be updated according to Equation (6) with each new trial, or for efficient on-line implementation, just \mathbf{v} can be updated according to the rank-one update to \mathbf{Q} of a new trial (Gu and Eisenstat, 1994). See Algorithm 1 for the real-time implementation of STIG.

3. EXPERIMENTS

For all experiments, the voluntary, fully informed, written consent of the persons used in this research was obtained as required by federal and Army regulations (U.S. Department of Defense Office of the Secretary of Defense, 1999; U.S. Department of the Army, 1990). The investigators adhered to Army policies for the protection of human subject (U.S. Department of the Army, 1990). All human subjects testing was approved by the Institutional Review Board of the United States Army Research Laboratory.

3.1. Rapid Serial Visual Presentation

RSVP experiments leverage the oddball paradigm to elicit a well-characterized P300 visually evoked neural response. In BCI uses, images are shown to a user at a high rate of speed (2–10 Hz), and the user is instructed to anticipate a pre-defined target class which appears infrequently in the images (Gerson et al., 2005; Sajda et al., 2010). When the target class image appears, the resultant P300 response provides a discernible signal for recognition by machine learning methods, allowing a BCI system to discriminate between target and non-target.

3.2. Experiment 1: Off-Line

The data sets used in Experiment 1 were recorded and analyzed previously; a brief description of the original experiment and collection is given here.

3.2.1. Paradigm

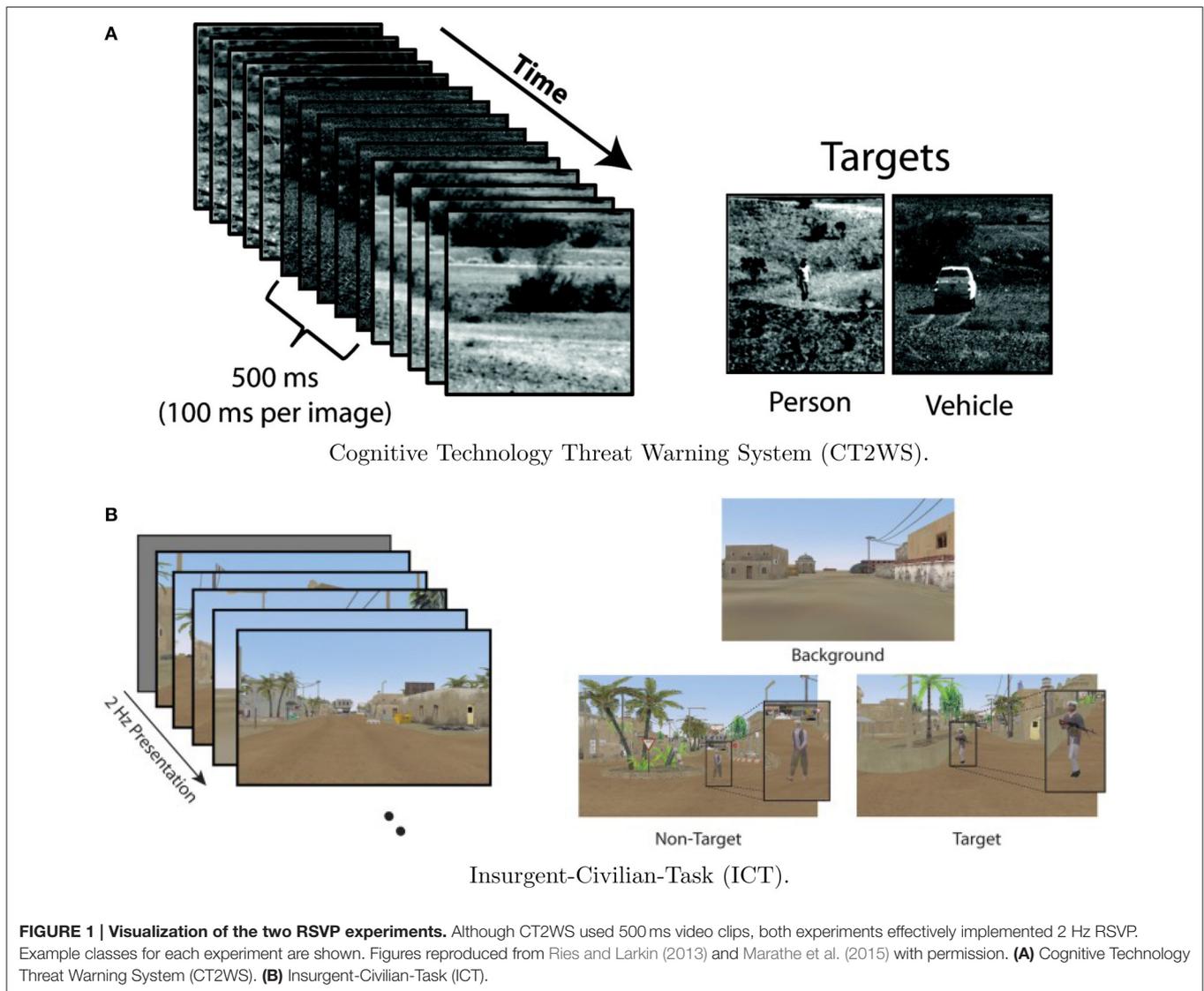
Data from Experiment 1 is comprised of two previously collected datasets. The first dataset utilized EEG data from the Cognitive Technology Threat Warning System (CT2WS) DARPA project in which subjects participated in a RSVP task comprised of short movie clips of five images at 10 Hz rate with no inter-stimulus interval (ISI) such that the effective rate between successive movie clips was 2 Hz. Targets consisted of either a person or a vehicle in natural settings. See **Figure 1A** for a visualization. The target class to background ratio was 1:9. Participants were instructed to manually press a button with their dominant hand when a target

Algorithm 1: Spectral Transfer with Information Geometry (STIG)

Data: $\{S_j\}_{j=1}^m, \{\mathbf{X}_i\}_{i=1}^n$
Result: $\{\hat{y}_i\}_{i=1}^n$
// Off-line: Calculate f_j for each subject, S_j , in the ensemble
for $j = 1, \dots, m$ **do**
 $f_j = G(\Sigma_j^{(-1)}, \Sigma_j^{(+1)}) \circ H(\mathbf{P}_j)$
end
// On-line: Classify each trial, \mathbf{x}_k , in real-time using spectral transfer
case $k \in 1, \dots, m-1$ **do**
 $\hat{y}_k = \text{mode}(\{f_j(\mathbf{X}_k)\}_{j=1}^m)$
end
case $k \in m, \dots, n$ **do**
 (i) $\mathbf{v} = \{\mathbf{v}_\ell \mid \arg \max_\ell \lambda_\ell\}$ s.t. $\mathbf{Q}(\{\mathbf{X}_i\}_{i=1}^k, \{f_j\}_{j=1}^m) \mathbf{v}_\ell = \lambda_\ell \mathbf{v}_\ell$
 (ii) $\hat{y}_k^{(0)} = \text{sign} \left(\sum_{j=1}^m f_j(\mathbf{X}_k) \cdot v_j \right)$
 (iii) $q = 0$
 repeat
 // E-step:
 for $j = 1, \dots, m$ **do**
 $\psi_j^{(q+1)} = P(f_j(\mathbf{X}) = 1 \mid \hat{y}^{(q)} = 1)$
 $\eta_j^{(q+1)} = P(f_j(\mathbf{X}) = -1 \mid \hat{y}^{(q)} = -1)$
 end
 // M-step:
 for $i = 1, \dots, k$ **do**
 $\hat{y}_i^{(q+1)} =$
 $\text{sign} \left(\sum_{j=1}^m f_j(\mathbf{X}_i) \left(\log \left(\frac{\psi_j^{(q+1)} \eta_j^{(q+1)}}{(1 - \psi_j^{(q+1)})(1 - \eta_j^{(q+1)})} \right) \right) \right)$
 $+ \log \left(\frac{\psi_j^{(q+1)}(1 - \psi_j^{(q+1)})}{\eta_j^{(q+1)}(1 - \eta_j^{(q+1)})} \right)$
 end
 $q = q + 1$
 until convergence;
 end
end

movie clip was shown. Participants conducted 25 2-min blocks with a short break between each block (Ries and Larkin, 2013).

For the second data set, Insurgent-Civilian-Task (ICT), subjects participated in a RSVP task with still images of a simulated desert metropolitan environment. Target images contained a person carrying a weapon. Non-target distractor images contained a person without a weapon, and background images contained neither. There were two experimental conditions: Target Only (TO) and Target-Non-target (TN). In the TO condition, only target images and background images were shown. In the TN condition, non-target distractor images were shown in addition to target and background images. Images were shown at 2 Hz with no ISI. See **Figure 1B** for a visualization. In the TO condition, the



ratio of target to background class images was 1:20, and for the TN condition, the ratio of target to distractor to background images was 1:1:14. Participants were instructed to manually press a button with their dominant hand when a target image was shown. Participants conducted six 2-min blocks of the TO condition and six 2-min blocks of the TN condition with a self-paced rest between each block (Marathe et al., 2015). Although there were three stimuli types (target, non-target, and distractor), data from only the target and non-target types were used in the present analysis.

3.2.2. Data Collection

Fifteen subjects (9 male, average age 39.5 years) participated in CT2WS. EEG data was collected with a BioSemi Active Two system (Amsterdam, Netherlands) at 512 Hz from 64 scalp electrodes arranged in a 10–10 montage and referenced off-line to the average of the left and right earlobes. All signals were digitally filtered from 0.1 to 55 Hz using an FIR filter. Data was collected using E-Prime software (Psychology Software Tools, Pittsburgh, PA; Ries and Larkin, 2013).

Seventeen different subjects (14 male, average age 34.9 years) participated in ICT. EEG data was collected with a BioSemi Active Two system (Amsterdam, Netherlands) at 1024 Hz from 64 scalp electrodes arranged in a 10–10 montage and referenced off-line to the average of the left and right earlobes. All signals were digitally filtered with an FIR filter from 0.1 to 50 Hz. Data was collected using E-Prime software (Psychology Software Tools, Pittsburgh, PA; Marathe et al., 2015).

3.2.3. Analysis

For single trial classification, all data was high-pass filtered at 1 Hz, down-sampled to 256 Hz, and Independent Component Analysis was performed in which eye movement and muscle activity were manually identified and removed. The data was divided into 1 s epochs time-locked to the stimulus onset, [0, 1s], and 18 channels (Fz, Cz, C3, C4, CPz, Pz, P3, P4, P7, P8, POz, PO3, PO4, PO7, PO8, Oz, O1, O2) were selected for classification according to Krusienski et al. (2008), $\mathbf{X} \in \mathbb{R}^{18 \times 256}$. Whereas, Krusienski et al. (2008) used 19 channels, our electrode montage did not have FCz and therefore only 18 of the 19 channels

from Krusienski et al. (2008) were used. For each of the 17 IC subjects, all distractor trials were removed (i.e., trials with a civilian shown). Pre-processing was performed with EEGLAB (Delorme and Makeig, 2004).

For Experiment 1, subject data from both paradigms were combined for a total of 32 subjects. Off-line analysis of the transfer method was conducted by leave-one-subject-out cross-validation. First, a subject-specific MDRM classifier was trained on each of the 32 subjects. For a test subject, all 31 other subject-specific MDRM classifiers comprised the STIG ensemble.

We compared our method against several other calibration suppression and calibration reduction methods which include both ensemble and pooled methods. For all ensemble methods, we use the MDRM classifier as the base classifier, and transfer information in decision space, $\{-1, 1\}$. For all pooled methods, we combine information at the feature level using information geometric features. First, we considered the following alternative methods which do not require a dedicated calibration session:

- **Max Subject-Specific Classifier (MSS)**—We first compared our method against the best direct subject-subject transfer of all 31 possible subjects. Although not possible to implement in reality, it serves as a hypothetical ceiling if we had omniscient knowledge of the available training subjects and could simply pick the best subject-specific classifier for subject-to-subject transfer.
- **Majority Vote (MV)**—This method uses a simple uniform weighted average of classifications from the ensemble of classifiers. By combining in decision space, an un-weighted linear combination is equivalent to a majority vote.
- **ℓ_1 -Regularized Cross-Validation (L1)**—This method pools all available training data and learns a weighted linear combination of classifier decisions according to a ℓ_2 regression with a ℓ_1 penalty (Fazli et al., 2011). We implemented this method using the MDRM classifier.
- **Pooled MDRM (PMDRM)**—We implemented a pooled version of the MDRM method in which we combined all available training data into a single MDRM classifier.

Additionally, we compared our STIG method to several of the most recent and highest-performing calibration reduction methods in the BCI literature. Unlike zero-calibration methods, calibration reduction methods require a dedicated calibration session. Thus, to compare against our zero-calibration method, we tested each calibration reduction method with varying amounts of calibration data. As with the zero-calibration methods, we used the MDRM as the base classifier for all calibration reduction methods.

- **Accuracy Weighted Ensemble (AWE)**—We used the weighted linear combination approach from Dalhoumi et al. (2014), Dalhoumi et al. (2015) to linearly combine classifiers to minimize a Mean-Squared Error (MSE) loss function on a set of calibration data.
- **Multi-Task Learning (MT)**—This method treats classification as a linear regression problem where feature weights are modeled as a multivariate Gaussian distribution whose unknown mean and co-variance matrix are jointly estimated

from multiple training subjects (Jayaram et al., 2016). Once these Gaussian priors are learned, calibration data from the test subject can be used to estimate optimal weights for classification. Since this method utilizes linear regression for classification, it is not possible to implement the MDRM classifier directly. Instead, in an effort to achieve consistency with the other methods, we implement the MT method with the same information geometric features that are used in the MDRM classifier.

- **Within-Subject Calibration (CALIB)**—This is the standard within subject calibration method that is traditionally used in the BCI literature. Data from the test subject is split into training and testing. In this case, an MDRM classifier, which is constructed using training data from a test subject, is used to classify the remaining test data from the same subject. This traditional classification scheme of calibrating the algorithm to the unique subject during a single session provides the performance ceiling of any calibration reduction or suppression method.

Due to the inherent class imbalance of the RSVP task, the performance of all implemented methods are reported in terms of the balanced accuracy, $\pi = \frac{1}{2}(\psi + \eta)$, which equally weights accuracy on both target and non-target trials.

3.2.4. Results

3.2.4.1. Comparison of zero-calibration algorithms

Figure 2 presents the balanced accuracy of each zero-calibration method averaged across all subjects in Experiment 1. A one-way ANOVA showed a significant difference between the five methods tested [$F_{(4, 159)} = 9.73, p < 0.001$].

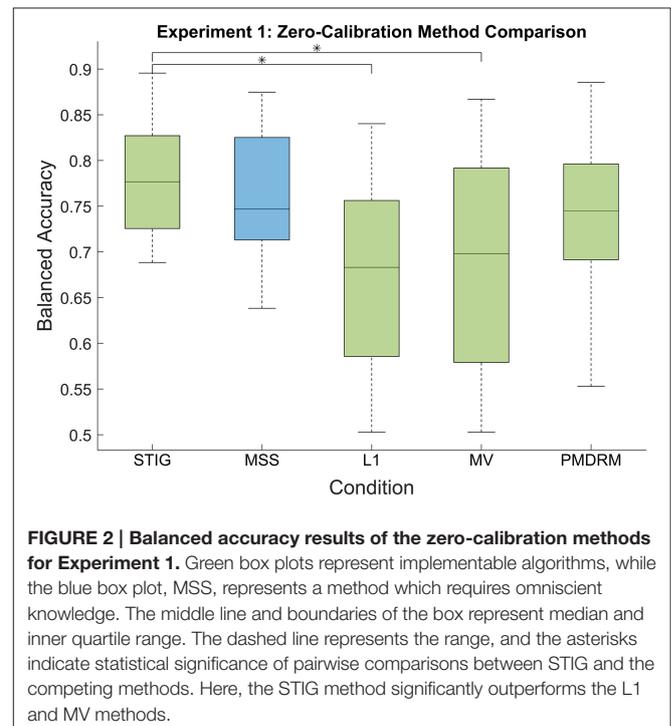


FIGURE 2 | Balanced accuracy results of the zero-calibration methods for Experiment 1. Green box plots represent implementable algorithms, while the blue box plot, MSS, represents a method which requires omniscient knowledge. The middle line and boundaries of the box represent median and inner quartile range. The dashed line represents the range, and the asterisks indicate statistical significance of pairwise comparisons between STIG and the competing methods. Here, the STIG method significantly outperforms the L1 and MV methods.

Multiple comparisons were performed for each pair using the Tukey-Kramer method. The proposed STIG method (0.78 ± 0.07) (mean \pm standard deviation) achieved comparable performances with the MSS classifier (0.77 ± 0.07 , $p = 0.97$) and significantly outperformed the L1 (0.67 ± 0.10 , $p < 0.001$) and MV (0.69 ± 0.11 , $p < 0.001$) classifiers, illustrating a benefit to be gained from the spectral-based ensemble learner over the competing zero-calibration methods. Additionally, The STIG outperformed our own implemented PMDRM classifier (0.74 ± 0.08), although not significantly ($p = 0.39$).

3.2.4.2. Comparison of calibration reduction algorithms

Figure 3 shows the performance of the STIG method compared to the results of the calibration reductions methods (AWE, MT, and CALIB) as a function of available trials. For testing, the calculation of balanced accuracy for each method was performed on a fixed number of hold-out trials (600). Each of the calibration reduction methods trained on an increasing number of trials (50–1000 trials at 50 trial intervals). Five hundred bootstrap iterations were performed in which training and testing trials were drawn randomly for each iteration to provide a more robust estimate of the performance distribution. STIG was applied to the sampled testing set in its entirety but did not incorporate any training trials. As shown in **Figure 3**, each of the calibration-based methods shows a general increase in performance for increasing number of training trials used in the classification.

Statistical analysis was performed between the four methods shown in **Figure 3** in a similar fashion as described in the previous section. Since additional ANOVAS were performed for each of the 20 calibration size conditions, the resulting p -values were corrected using the false-discovery rate (Benjamini and Yekutieli, 2001). Results show that the STIG method significantly outperforms the AWE method across all numbers of calibration trials used ($p < 0.05$). The STIG method

significantly outperforms the MT for the 50–400 calibration trial conditions ($p < 0.05$), and achieves comparable performances for the 450–1000 calibration trial conditions. Similarly, when compared to the traditional within-subject calibration method, STIG significantly outperforms CALIB when 400 calibration trials or less are available ($p < 0.05$). From 450 to 1000 calibration trials, STIG achieves statistically similar performance to the CALIB method ($p > 0.05$).

3.2.4.3. Effect of ensemble size on transfer accuracy

An obvious question that can be asked about the effectiveness of STIG is the effect of ensemble size on the accuracy of the transfer. **Figure 4** shows the effect on performance when various numbers of training subjects are included in the STIG method. For each of the 32 subjects in Experiment 1, 500 bootstrap runs were conducted for each ensemble size (5–31) in which a subset of training subjects were randomly chosen for the ensemble. **Figure 4** shows that the performance of the STIG method asymptotically increases with increasing ensemble size.

3.2.4.4. STIG estimation of subject-subject transfer accuracy

SML attempts to infer the balanced accuracy of each individual classifier in the ensemble and weights the classification provided by each subject accordingly. In order to analyze the effectiveness of STIG's ability to infer the accuracy of each ensemble subject from unsupervised data, the learned ensemble weights from each test subject were plotted against the true individual balanced accuracy of each direct subject-to-subject transfer of each subject in the ensemble. As shown in **Figure 5**, a clear correlation can be seen between the balanced accuracy of the direct subject-subject transfer with the weights assigned according to Equation (8).

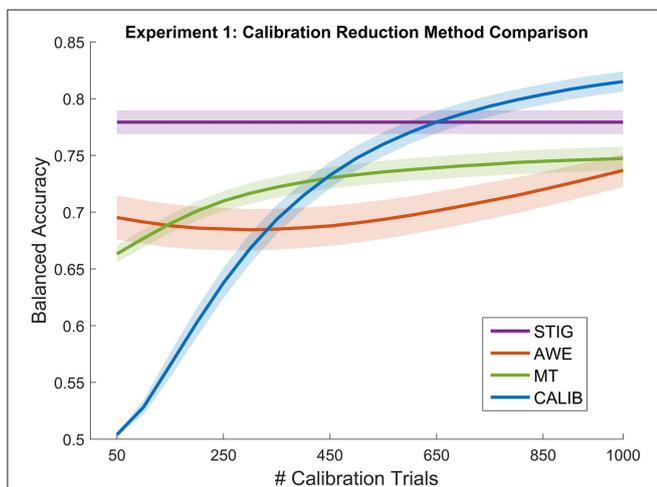


FIGURE 3 | Balanced Accuracy as a function of number of calibration trials for the calibration reduction methods (AWE, MT, and CALIB) tested in Experiment 1. The proposed zero-calibration method, STIG, which is independent of the number of calibration trials, is plotted for comparison. Lines represent means averaged across all subjects ($N = 32$) while shaded regions represent standard error.

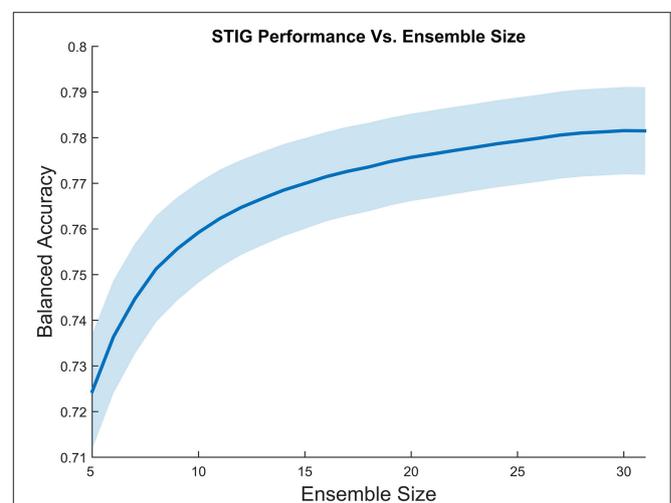
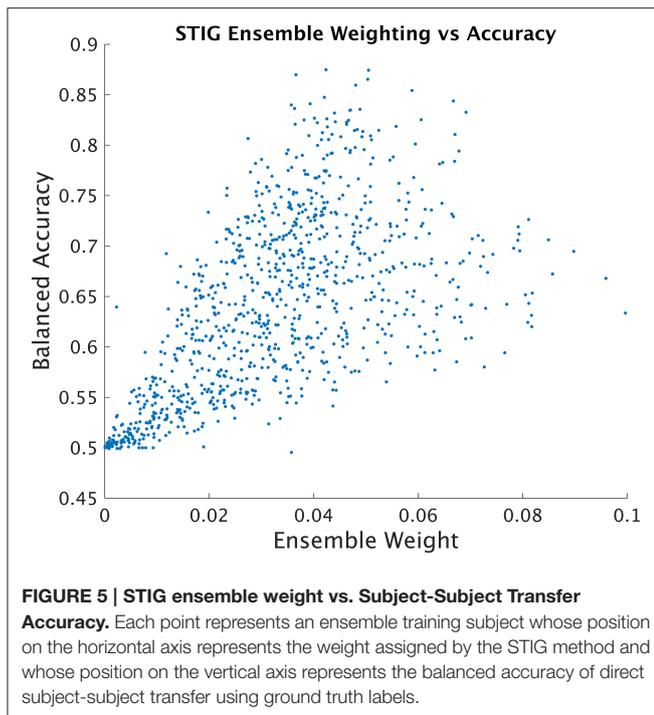


FIGURE 4 | Grand average balanced accuracy as a function of ensemble size for the STIG method. The balanced accuracy averaged over 1000 bootstrap runs is reported for each ensemble size. The ensemble size represents the number of training subjects selected for ensemble learning. The shaded region represents standard error.



3.3. Experiment 2: Real-Time Feedback

3.3.1. Paradigm

The ICT paradigm was modified for the real-time feedback experiment, resulting in the ICT2 paradigm. Subjects participated in a similar RSVP task in which only target and background images were presented (see **Figure 6**). Images were shown at 2 Hz with no inter-stimulus interval. The ratio of target to background images was 1:9. Subjects sat approximately 60 cm away from a 21" LCD monitor which displayed RSVP images that spanned the entire screen. Unlike Experiment 1, subjects in Experiment 2 did not perform any manual button pressing and were instead instructed to inaudibly count the number of targets observed during a single block. Participants conducted 15 blocks of 1 min duration in which 120 images were presented in each block. Participants were given a self-paced rest interval between each block to mitigate any fatigue effects.

Each image was classified as target or non-target in real-time from a 1-s epoch of EEG data immediately following the image presentation. A separate 21" LCD monitor immediately to the left of the image display monitor presented the feedback results in real-time. Although feedback was continuously displayed, subjects were instructed to only observe the feedback screen during the rest intervals between blocks as to maintain focus on the RSVP task. The feedback window, as shown in **Figure 7**, displays the 10 targets with the highest score and 10 non-targets with the highest score, where score is the un-signed value of Equation (7). This provided a running measure of how accurately the system classified the images. At the conclusion of each block, the feedback window also displayed balanced accuracy from a single block for odd numbered blocks or the combined accuracy

of the last two blocks for even blocks. The feedback window was reset for each odd block.

3.3.2. Data Collection

Ten new subjects (9 male, average age 32.7 years) participated in ICT2. EEG data was collected with a BioSemi Active Two system (Amsterdam, Netherlands) at 256 Hz from 64 scalp electrodes arranged in a 10–10 montage and referenced off-line to the average of the left and right mastoids. All signals were digitally filtered from 0.1 to 50 Hz. BCI2000 was used for data collection and on-line classification (Schalk et al., 2004).

3.3.3. Analysis

The data was divided into 1 s epochs time-locked to the stimulus onset, [0, 1s]. The same 18 channels used in Experiment 1 were also used for classification in Experiment 2. For classification of the online subjects in Experiment 2, MDRM classifiers were generated from all 32 subjects from Experiment 1. Although the MDRM classifiers were trained using data cleaned of artifacts, ICA was not performed on the streaming data from subjects of Experiment 2. For the first 31 trials of the first RSVP block, while $\text{rank}(\hat{Q}) < 32$, MV was used for real-time classification. Only steps (i) and (ii) of Algorithm 1 were implemented in real-time to classify images and provide feedback. Those results are reported in **Table 1**. The results reported in **Figures 8–12** were generated in *post-hoc* analysis by implementing steps (i)-(iii) of Algorithm 1 causally on the streamed data. These *post-hoc* results were compared against the competing methods described in Experiment 1, where for zero-calibration methods, we compared STIG to MSS, L1, MV, and PMDRM, and for calibration reduction methods, we compared STIG to AWE and MT and CALIB.

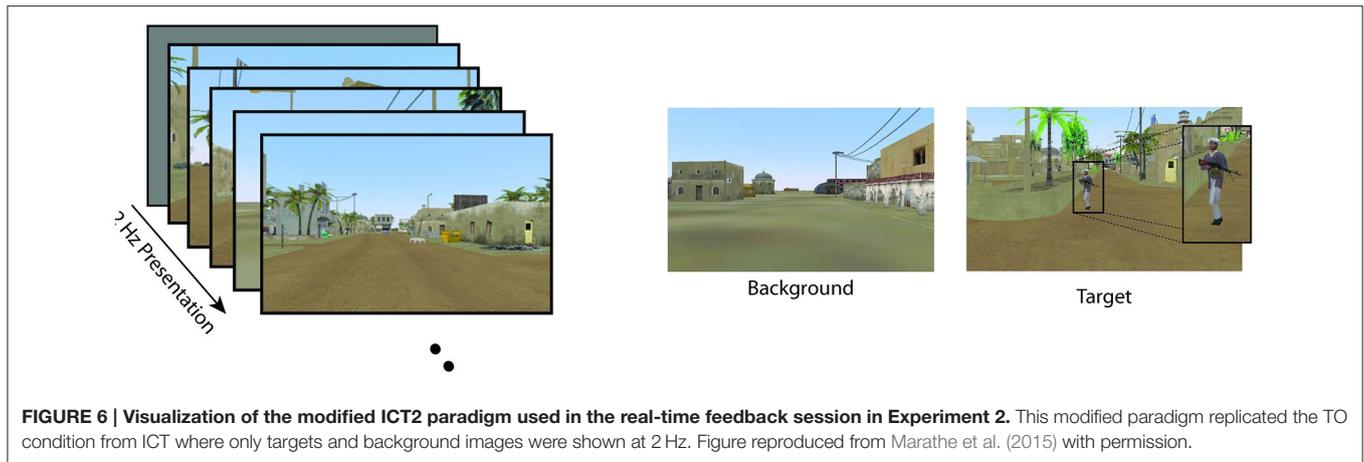
3.3.4. Results

3.3.4.1. Real-time feedback results

Results of the real-time feedback session in Experiment 2 are reported for each subject in **Table 1**, with an average balanced accuracy of 0.62 ± 0.06 . These reflect results from analysis performed during the on-line experiment that was used to generate the real-time RSVP feedback. Two subjects, S2 and S6, perform more than one standard deviation above the mean and two subjects, S1 and S9, perform more than one standard deviation below the mean. Notice that the majority of subjects achieve a maximum online block performance of at least 70% and several subjects are able to achieve online block performances of over 80%.

3.3.4.2. Comparison of zero-calibration algorithms

Figure 8 shows the balanced accuracy results of each zero-calibration method averaged across all subjects in Experiment 2. As with Experiment 1, a one-way ANOVA showed a significant difference between the five methods tested [$F_{(4, 49)} = 4.92$; $p < 0.01$]. Using the Tukey-Kramer method for correction of multiple comparisons, the STIG method (0.64 ± 0.07) achieved statistically similar results to the omniscient MSS classifier (0.65 ± 0.06 , $p = 0.9$) and significantly out-performed the MV classifier (0.54 ± 0.06 , $p = 0.012$). The STIG method achieves



slightly higher performance results compared to the L1 (0.61 ± 0.07) and PMDRM (0.55 ± 0.07) classifiers although these were not found to be significant ($p = 0.83$ and $p = 0.12$, respectively).

3.3.4.3. Comparison of calibration reduction algorithms

Figure 9 shows the results of STIG compared to the calibration reduction methods (AWE and MT) and within-subject performance ceiling (CALIB). Unlike the bootstrap analysis performed in Experiment 1, results for the calibration methods in Experiment 2 were calculated causally. The data was partitioned in sequence into calibration and testing. This was done so

that the *post-hoc* analysis would simulate online analysis from the real-time feedback experiment. The calibration reduction methods were tested using varying numbers of calibration trials from 50 to 1000 trials.

Similar to Experiment 1, statistical analysis was performed where one-way ANOVAS were computed for each of the 20 calibration size conditions, the resulting p -values were corrected using the false-discovery rate (Benjamini and Yekutieli, 2001). As shown in **Figure 9**, the STIG method significantly outperforms the AWE method for when 50–600 calibration trials are used for AWE training ($p < 0.05$), and comparable performance

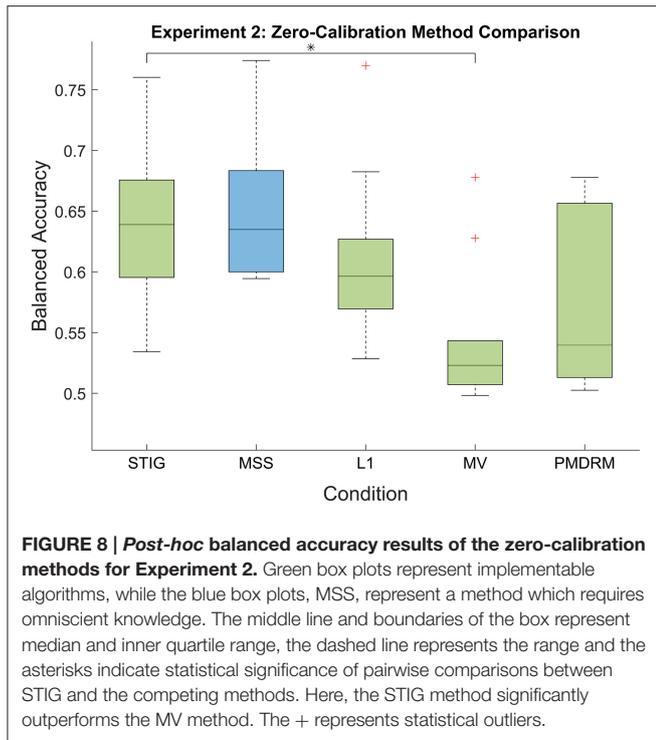


FIGURE 8 | Post-hoc balanced accuracy results of the zero-calibration methods for Experiment 2. Green box plots represent implementable algorithms, while the blue box plots, MSS, represent a method which requires omniscient knowledge. The middle line and boundaries of the box represent median and inner quartile range, the dashed line represents the range and the asterisks indicate statistical significance of pairwise comparisons between STIG and the competing methods. Here, the STIG method significantly outperforms the MV method. The + represents statistical outliers.

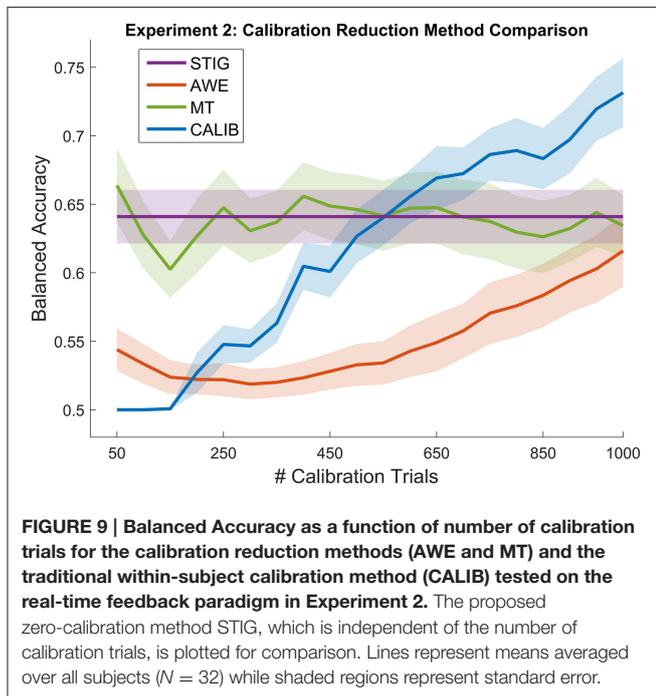


FIGURE 9 | Balanced Accuracy as a function of number of calibration trials for the calibration reduction methods (AWE and MT) and the traditional within-subject calibration method (CALIB) tested on the real-time feedback paradigm in Experiment 2. The proposed zero-calibration method STIG, which is independent of the number of calibration trials, is plotted for comparison. Lines represent means averaged over all subjects ($N = 32$) while shaded regions represent standard error.

is attained when 650 or more calibration trials are used. The STIG method achieves similar performance compared to the MT method ($p > 0.05$) across all numbers of calibration trials used. Unlike the performance shown in Experiment 1 (Figure 3), the MT method does not appear to show a gradual increase in performance with increasing number of calibration trials, as

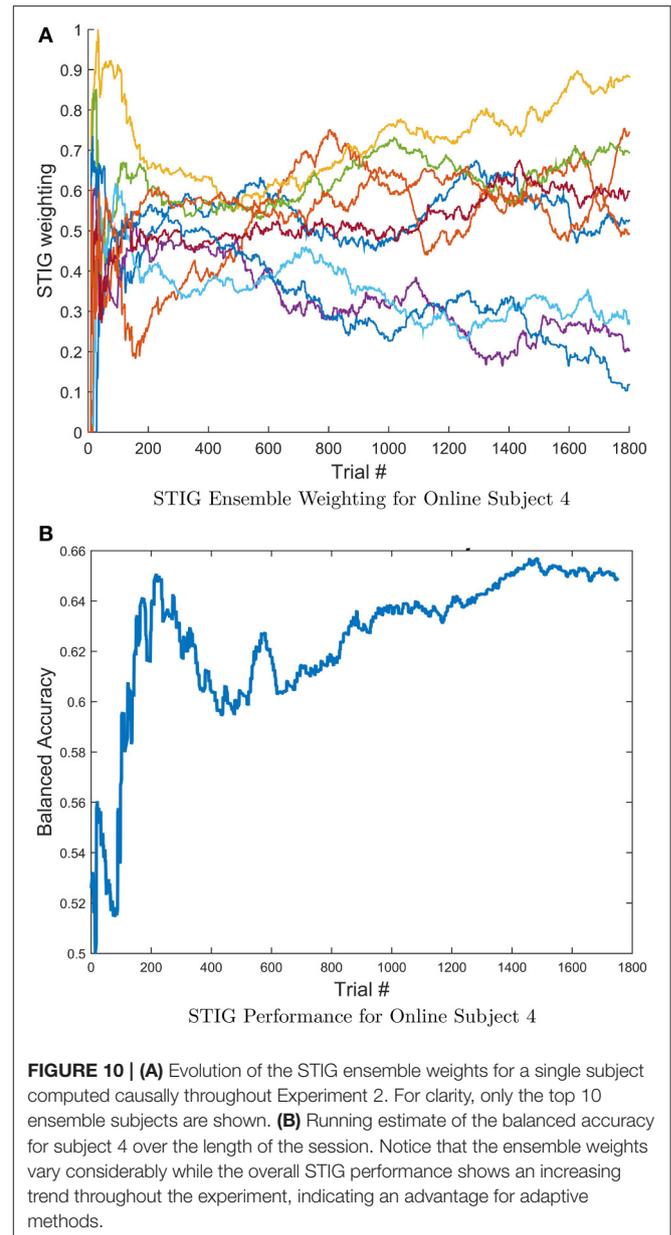


FIGURE 10 | (A) Evolution of the STIG ensemble weights for a single subject computed causally throughout Experiment 2. For clarity, only the top 10 ensemble subjects are shown. **(B)** Running estimate of the balanced accuracy for subject 4 over the length of the session. Notice that the ensemble weights vary considerably while the overall STIG performance shows an increasing trend throughout the experiment, indicating an advantage for adaptive methods.

would be expected. Compared to the standard within-subject calibration method (CALIB), STIG achieves significantly higher balanced accuracies when 50–350 calibration trials are available for the CALIB method ($p < 0.05$). For the 400–1000 calibration size conditions, STIG achieves statistically similar performance to the CALIB method ($p > 0.05$).

4. STIG CHARACTERIZATION

4.1. Evolution of SML Weights Over Experiment

To further characterize the transfer properties of the ensemble learner in the STIG method, the ensemble weights from STIG were plotted over time. The estimated weights are tracked as

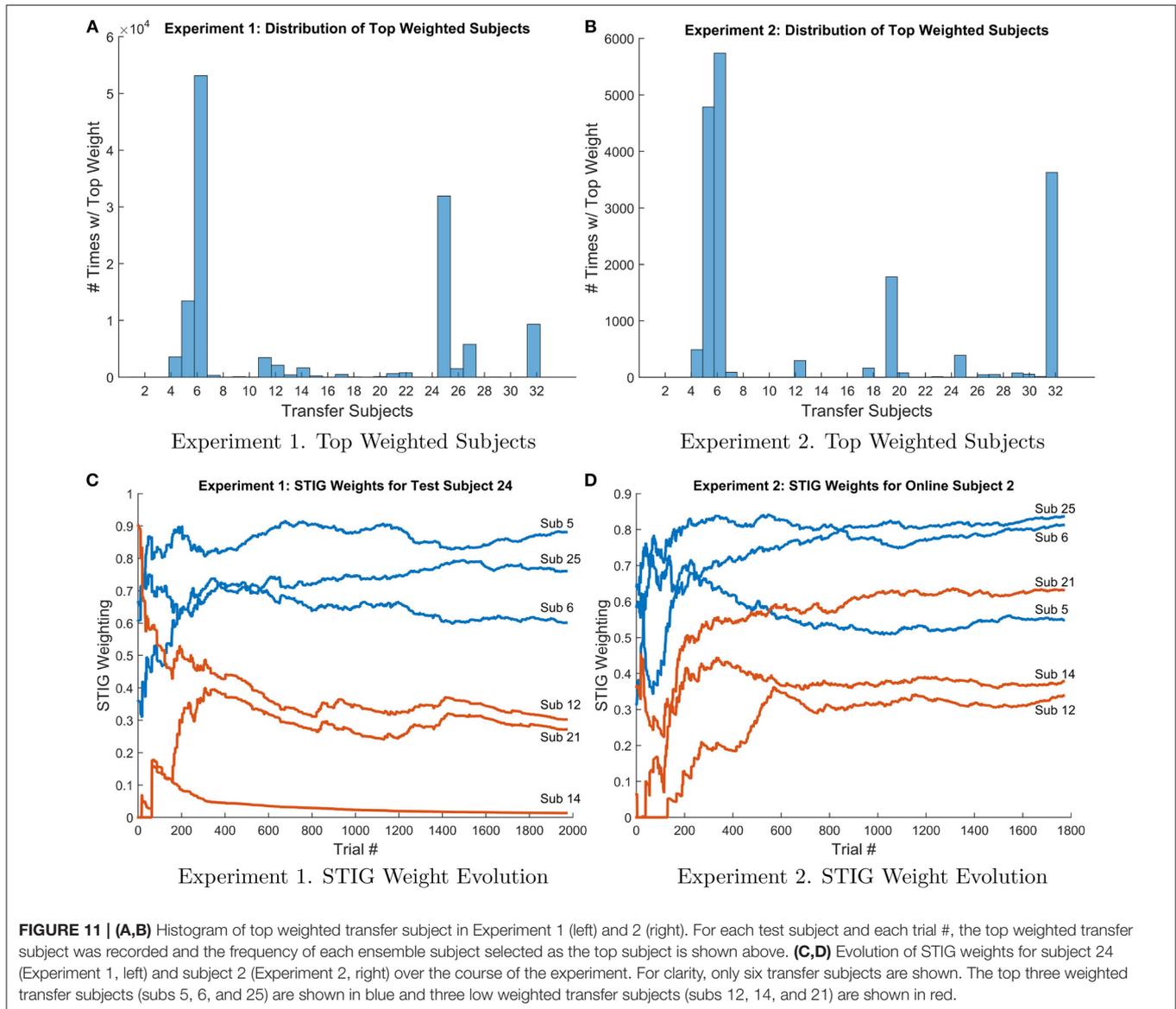
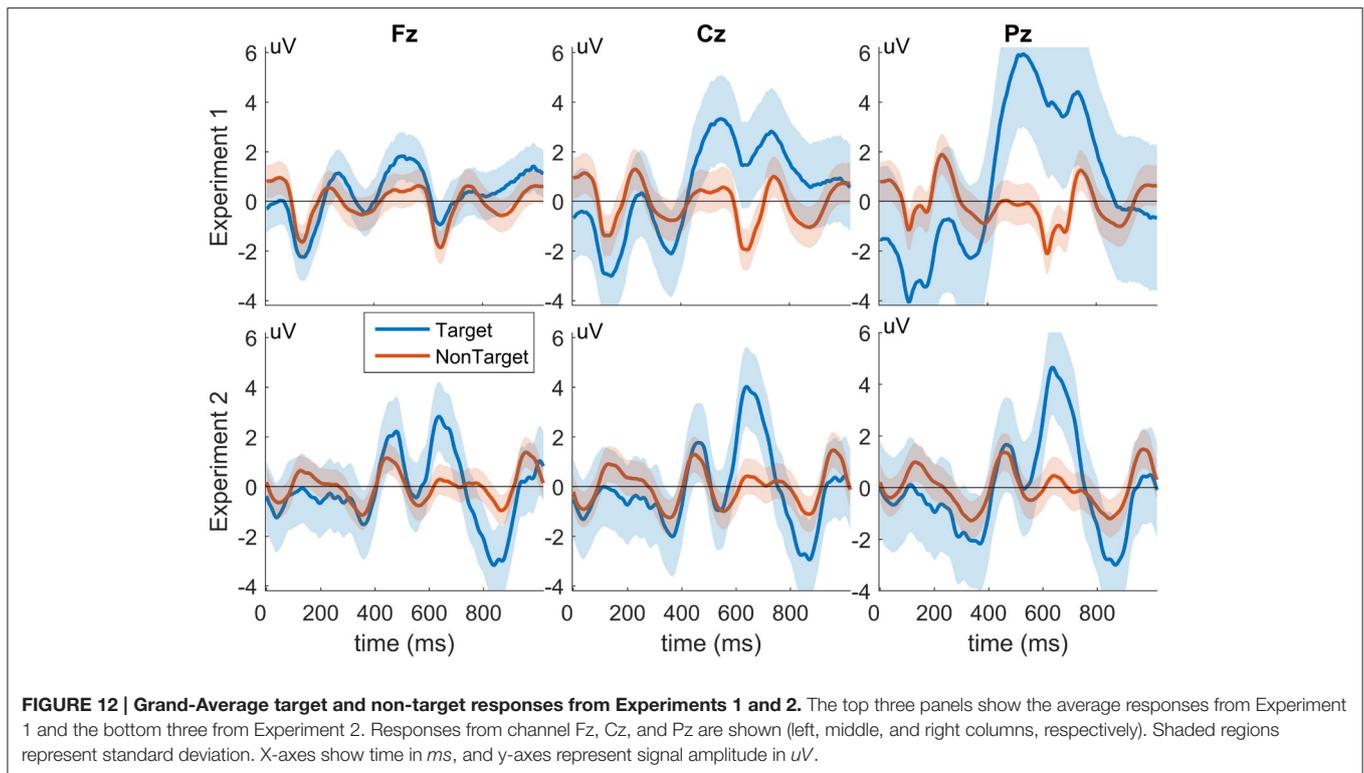


FIGURE 11 | (A,B) Histogram of top weighted transfer subject in Experiment 1 (left) and 2 (right). For each test subject and each trial #, the top weighted transfer subject was recorded and the frequency of each ensemble subject selected as the top subject is shown above. **(C,D)** Evolution of STIG weights for subject 24 (Experiment 1, left) and subject 2 (Experiment 2, right) over the course of the experiment. For clarity, only six transfer subjects are shown. The top three weighted transfer subjects (subs 5, 6, and 25) are shown in blue and three low weighted transfer subjects (subs 12, 14, and 21) are shown in red.

STIG classifies each trial in Experiment 2 causally. Note that in the real-time feedback experiment, as each new trial becomes available, the ensemble classification from that trial gets added to the window of previous classified trials that is used by the STIG algorithm. Thus, STIG updates its ensemble weights for each trial. **Figure 10A** shows the weights of 10 ensemble subjects for subject 4 as they evolve over the course of the real-time experiment. It is apparent that the relative weight of each subject in the ensemble is not static but rather evolves throughout the experiment, varying by as much as 0.6. **Figure 10B** illustrates a running estimate of the balanced accuracy for subject 4 where the STIG performance shows an increasing trend over the length of the session.

Histogram plots and weight evolution curves of the top weighted transfer subjects were made and are shown in **Figure 11**. **Figures 11A,B** show the distribution of top weighted subjects for experiments 1 and 2 respectively. For each test subject

and for each test trial, the ensemble subject with the top weight assigned by the STIG method was recorded and a histogram was made showing the frequency to which each transfer subject was weighted the highest for a given trial. These histogram plots show which subjects are utilized the most in the ensemble transfer. The top weighted subjects from Experiment 1 were subjects 5, 6, 25, and 32, whereas the top weighted subjects from Experiment 2 were subjects 5, 6, 19, and 32. Notice that subjects 5, 6, and 32 are among the top weighted subjects across both experiments. **Figures 11C,D** show the STIG weights for the top three subjects as well as three lower performing subjects over-time for Experiments 1 and 2, respectively. The top three high-weighted subjects (subjects 5, 6, and 25) are shown in blue and the three low-weighted subjects (12, 14, and 15) are shown in red. Note, only six subjects are shown instead of all 32 for visual clarity. It can be seen in both **Figures 11C,D** that the STIG weights



vary over-time. Additionally, **Figure 11D** shows a lower-ranked ensemble subject (subject 21) eventually overtaking the higher-ranked ensemble subject (subject 5) at around 600 trials.

5. DISCUSSION

The field of transfer learning for BCI is rapidly increasing with a push to develop algorithms that require less and less training data. Although the field has shown more progress in developing algorithms that can reduce calibration, there still remains very little progress on general purpose, zero-calibration algorithms that are capable of classifying brain-signal data from a new subject without any prior training data. This study combines the recent efforts in information geometry and joint classification to present the Spectral Transfer with Information Geometry (STIG) method, an unsupervised ensemble learner that requires zero calibration data to out-perform other state-of-the-art calibration suppression and calibration reduction methods.

The effectiveness of STIG was tested and validated with both off-line and on-line experimentation. For off-line analysis in Experiment 1, STIG significantly outperforms the two competing zero-calibration based ensemble methods (L1 and MV). This demonstrates a benefit gained from a spectral based ensemble learner when applied to BCI transfer. In theory, the SML will provide an ensemble accuracy that is within an additive constant of the best performing subject in the ensemble, and this bound improves with increasing ensemble size. As shown in **Figures 2, 8**, STIG achieves and exceeds the performance of the best performing classifier in the ensemble. Together, this demonstrates that with neither a calibration session nor

any labeled training data, the STIG method should equal the best performing subject in the ensemble and can potentially do better. As compared to the calibration reduction methods for Experiment 1, the STIG method significantly outperforms the AWE method for all numbers of available calibration data and outperforms the MT method and traditional within-subject calibration methods when only a limited amount of training data is available (<400 trials), which approximates to a 5–10 min calibration session. With large amounts of training data available however, the MT and within-subject calibration methods achieve comparable performance to the STIG method. Nevertheless, these methods still require training data and thus necessitate a deliberate calibration session in order to use them. The analyzed data from the real-time paradigm in Experiment 2 show similar conclusions, overall demonstrating that the STIG method has the ability to out-perform other zero-calibration methods and can achieve comparable—and often better—performance to methods which require training data.

More than simply training a single classifier with historical data, STIG provides a unique and adaptable transfer to each test subject. **Figure 10** shows for a representative test subject that the relative weighting of each member of the ensemble is not static with time, and, in fact, varies quite significantly throughout the experiment. According to **Figure 5**, where STIG ensemble weights are shown to be correlated with direct subject-to-subject transfer accuracy, there exist a non-stationarity in the ability of any one ensemble member to classify the test subject. Thus, the adaptive nature of STIG provides a significant advantage over traditional zero-calibration methods which do not adapt after the beginning of the experiment. **Figures 11C,D** also show that this

TABLE 1 | Total session balanced accuracy by subject from real-time feedback experiment.

On-line Feedback Results (Balanced Accuracy)				
Subject	Mean	STD	Min	Max
S1	0.54	0.07	0.46	0.68
S2	0.74	0.06	0.60	0.83
S3	0.59	0.05	0.48	0.68
S4	0.64	0.06	0.54	0.78
S5	0.61	0.07	0.47	0.73
S6	0.71	0.07	0.60	0.83
S7	0.62	0.08	0.52	0.81
S8	0.57	0.05	0.51	0.69
S9	0.54	0.06	0.42	0.62
S10	0.67	0.07	0.55	0.76
Overall	0.62	0.06	0.51	0.74

Shown is the mean accuracy as well as the STD, minimum, and maximum run performance over the 15 online feedback blocks.

non-stationarity exists even with subjects that are consistently top-weighted. Despite this non-stationarity, there exist subjects that tend to be consistently top-weighted throughout all test subjects and throughout all trials as seen in **Figures 11A,B**, where several subjects (notably 5, 6, and 32) have frequency counts that are several orders of magnitude higher than most other subjects. These subjects show high generalizability as their models transfer well to new subjects, achieving high subject-to-subject prediction accuracy, which the STIG method is able to infer.

For on-line implementation, consideration must be given to the computational complexity of the presented algorithm. In the presented experiments, convergence of Algorithm 1 completed well within the half second window before the next EEG epoch was available to classify. Much of the computational work is performed off-line in finding the class means of each of the ensemble subjects. During on-line implementation, Algorithm 1 incurs the following computational costs (m classifiers in the ensemble, T samples in an EEG epoch, C channels used for classification, n trials collected): $4m(T-1)C^2 + \mathcal{O}(mTC)$ from feature extraction (2) for each classifier in the ensemble, $\frac{64}{3}mC^3 + \mathcal{O}(mC)^2$ from the MDRM classification (5), $2(2n-1)m^2 + \mathcal{O}(m)$ from the covariance estimation in (6), $\frac{1}{3}m^3 + \mathcal{O}(m^2)$ from the eigendecomposition of \mathbf{Q} , and $\mathcal{O}(nm)$ for each iteration of (iii) in Algorithm 1. The costly steps of the algorithm take place in the extraction and classification of the information geometric feature. The spectral transfer contributes relatively little to the computational complexity. Regardless, it can be further diminished by simply updating the eigendecomposition based on the rank one update of an additional trial without the re-estimation of \mathbf{Q} (Gu and Eisenstat, 1994). More difficult would be to reduce the cost of the MDRM classification which grows combinatorially with the size of the ensemble, but in practice, this has not proven to be prohibitive for on-line implementation with real-time feedback.

There exists a small difference between the results of the two experiments in which there is a nominal decrease in balanced accuracy from Experiment 2 as compared to Experiment 1. Analysis from Experiment 1 was performed in a similar way most methods use when comparing BCI transfer; using a leave-one-subject-out cross validation transfer where one test subject is extracted from a dataset and the remaining subjects are used as training for BCI transfer. However, leave-one-out cross validation is not indicative of a use-case for practical BCI transfer where data from the same experiment is used for transfer. Experiment 2 attempts to more realistically capture real world use with cross-experiment transfer by using data from Experiment 1 as historical training data to classify subject data in Experiment 2. However, Experiment 2 contains several confounding factors which might explain its decreased performance. Experiment 2, which uses data from Experiment 1 for transfer, contains several paradigmatic differences that can negatively impact the performance of this cross-experiment transfer. First, Experiment 1 displays target, non-target, and distractor images to each subject, whereas Experiment 2 only utilizes target and non-target images. Additionally, subjects in Experiment 1 performed manual button pressing during target images, while subjects in Experiment 2 did not. Finally, Experiment 1 was collected using e-prime software and Experiment 2 was collected using BCI2000, which may have different latencies between the two systems. All of these factors likely contribute to observed differences seen in target response amplitude and latency between the two experiments (**Figure 12**), which shows the grand-average target and non-target responses for Experiments 1 (top) and 2 (bottom) for channels Fz, Cz, and Pz. These morphological differences in target waveforms can negatively impact the performance of the cross experiment transfer and potentially explain the decrease in accuracy from Experiment 1 to Experiment 2.

The STIG method has an initial start-up cost associated with it as it requires at least m predictions from each of its m ensemble classifiers in order to estimate the covariance matrix and generate accurate ensemble weights. As a result, during the beginning of online classification of a new subject, STIG resorts to simple majority voting until m trials have been collected. Additional work will seek to alleviate this issue and allow the STIG method to be applied immediately from the start, for example, one approach could be to use the L1-regularization technique from Fazli et al. (2011) as a warm-start for the STIG weight estimation. Additionally, the STIG method utilizes all previous test data labels for ensemble weight estimation, however, this may not be optimal during long test sessions as information from ensemble predictions from the beginning of a session may not be useful during the end of a session, due to fatigue effects and other sources of non-stationarity (i.e., an ensemble subject that transfers well during one portion of a test session may not transfer well during another portion). One possible solution to this issue would be to utilize an iterative covariance matrix update method together with a decay-rate to put emphasis on more recent classification predictions. These current limitations of the proposed method will be addressed in future work which will potentially improve the transfer capabilities of the STIG method.

6. CONCLUSION

This paper presents a novel method, termed Spectral Transfer with Information Geometry (STIG), to achieve unsupervised BCI transfer learning that outperforms the state-of-the-art calibration reduction and suppression methods when little or no calibration data is present. Additionally, this paper also validates the STIG method for use in real-time feedback BCI systems, representing a step-forward in the overall goal of attaining a general purpose, user-independent BCI system.

AUTHOR CONTRIBUTIONS

NW, VL, AB, and BL designed research; NW, VL, AB, and KB performed research; NW, VL, and AB analyzed data and NW, VL, AB wrote the paper.

REFERENCES

- Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2012). Multiclass brain-computer interface classification by Riemannian geometry. *IEEE Trans. Biomed. Eng.* 59, 920–928. doi: 10.1109/TBME.2011.2172210
- Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2013). Classification of covariance matrices using a Riemannian-based kernel for BCI applications. *Neurocomputing* 112, 172–178. doi: 10.1016/j.neucom.2012.12.039
- Barachant, A., and Congedo, M. (2014). A plug&Play P300 BCI using information geometry. arXiv:1409.0107.
- Barbaresco, F. (2008). “Innovative tools for radar signal processing Based on Cartan’s geometry of SPD matrices and Information Geometry,” in 2008 *IEEE Radar Conference* (Rome), 1–6. doi: 10.1109/RADAR.2008.4720937
- Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29, 1165–1188. doi: 10.1214/aos/1013699998
- Bin, G., Gao, X., Yan, Z., Hong, B., and Gao, S. (2009). An online multi-channel svsep-based brain-computer interface using a canonical correlation analysis method. *J. Neural Eng.* 6:046002. doi: 10.1088/1741-2560/6/4/046002
- Blankertz, B., Tangermann, M., Vidaurre, C., Fazli, S., Sannelli, C., Haufe, S., et al. (2010). The Berlin brain-computer interface: non-medical uses of BCI technology. *Front. Neurosci.* 4:198. doi: 10.3389/fnins.2010.00198
- Dalhousi, S., Dray, G., and Montmain, J. (2014). “Knowledge transfer for reducing calibration time in brain-computer interfacing,” in 2014 *IEEE 26th International Conference on Tools with Artificial Intelligence (ICTAI)*, (Limassol), 634–639.
- Dalhousi, S., Dray, G., Montmain, J., Derosière, G., and Perrey, S. (2015). “An adaptive accuracy-weighted ensemble for inter-subjects classification in brain-computer interfacing,” in 2015 *7th International IEEE/EMBS Conference on Neural Engineering (NER)* (Montpellier), 126–129. doi: 10.1109/NER.2015.7146576
- Dawid, A. P., and Skene, A. M. (1979). Maximum likelihood estimation of observer error-rates using the EM algorithm. *J. R. Stat. Soc. C* 28, 20–28.
- Delorme, A., and Makeig, S. (2004). Eeglab: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009
- Fazli, S., Danóczy, M., Schelldorfer, J., and Müller, K.-R. (2011). ℓ_1 -penalized linear mixed-effects models for high dimensional data with application to BCI. *Neuroimage* 56, 2100–2108. doi: 10.1016/j.neuroimage.2011.03.061
- Fletcher, P. T., and Joshi, S. (2004). “Principal geodesic analysis on symmetric spaces: statistics of diffusion tensors,” in *Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis*, eds M. Sonka, I. A. Kakadiaris, and J. Kybic (Heidelberg: Springer), 87–98.

ACKNOWLEDGMENTS

This project was supported by the Office of the Secretary of Defense Autonomy Research Pilot Initiative program MIPR DWAM31168 and by the U.S. Army Research Laboratory, under Cooperative Agreement Number W911NF-10-2-0022. This research was also supported in part by an appointment to the Student Research Participation Program at the U.S. Army Research Laboratory administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and USARL. The views and the conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

- Gerson, A. D., Parra, L. C., and Sajda, P. (2005). Cortical origins of response time variability during rapid discrimination of visual objects. *Neuroimage* 28, 342–353. doi: 10.1016/j.neuroimage.2005.06.026
- Gu, M., and Eisenstat, S. C. (1994). A stable and efficient algorithm for the rank-one modification of the symmetric eigenproblem. *SIAM J. Matrix Anal. Appl.* 15, 1266–1276. doi: 10.1137/S089547989223924X
- Harandi, M. T., Sanderson, C., Hartley, R., and Lovell, B. C. (2012). “Sparse coding and dictionary learning for symmetric positive definite matrices: a Kernel approach,” in *Computer Vision—ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7–13, 2012, Proceedings, Part II*, eds A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid (Heidelberg: Springer), 216–229.
- Jayaram, V., Alamgir, M., Altun, Y., Schölkopf, B., and Grosse-Wentrup, M. (2016). Transfer learning in brain-computer interfaces. *IEEE Comput. Intell. Mag.* 11, 20–31. doi: 10.1109/MCI.2015.2501545
- Jayasumana, S., Hartley, R., Salzmann, M., Li, H., and Harandi, M. (2015). Kernel methods on riemannian manifolds with gaussian rbf kernels. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 2464–2477. doi: 10.1109/TPAMI.2015.2414422
- Johnson, G., Waytowich, N., and Krusienski, D. J. (2011). “Chapter: The challenges of using scalp-EEG input signals for continuous device control,” in *Foundations of Augmented Cognition. Directing the Future of Adaptive Systems: Proceedings of 6th International Conference, FAC 2011, Held as Part of HCI International 2011, Orlando, FL, USA, July 9–14*, (Heidelberg: Springer), 525–527.
- Kindermans, P.-J., Tangermann, M., Müller, K.-R., and Schrauwen, B. (2014). Integrating dynamic stopping, transfer learning and language models in an adaptive zero-training erp speller. *J. Neural Eng.* 11:035005. doi: 10.1088/1741-2560/11/3/035005
- Krusienski, D. J., Sellers, E. W., McFarland, D. J., Vaughan, T. M., and Wolpaw, J. R. (2008). Toward enhanced p300 speller performance. *J. Neurosci. Methods* 167, 15–21. doi: 10.1016/j.jneumeth.2007.07.017
- Lance, B. J., Kerick, S. E., Ries, A. J., Oie, K. S., and McDowell, K. (2012). Brain-computer interface technologies in the coming decades. *Proc. IEEE* 100, 1585–1599. doi: 10.1109/JPROC.2012.2184830
- Lotte, F. (2015). Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain computer interfaces. *Proc. IEEE* 103, 871–890. doi: 10.1109/JPROC.2015.2404941
- Lotte, F., and Guan, C. (2009). “An efficient P300-based brain-computer interface with minimal calibration time,” in *Assistive Machine Learning for People with Disabilities symposium (NIPS’09 Symposium)* (Vancouver, BC).
- Lotte, F., Guan, C., and Ang, K. K. (2009). “Comparison of designs towards a subject-independent brain-computer interface based on motor imagery,” in 2009 *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Minneapolis, MN), 4543–4546. doi: 10.1109/IEMBS.2009.5334126

- Marathe, A. R., Ries, A. J., Lawhern, V. J., Lance, B. J., Touryan, J., McDowell, K., et al. (2015). The effect of target and non-target similarity on neural classification performance: a boost from confidence. *Front. Neurosci.* 9:270. doi: 10.3389/fnins.2015.00270
- Metzen, J., Kim, S., and Kirchner, E. (2011). "Minimizing calibration time for brain reading," in *Pattern Recognition, Vol. 6835, Lecture Notes in Computer Science*, eds R. Mester and M. Felsberg (Heidelberg: Springer), 366–375.
- Moakher, M., and Batchelor, P. G. (2006). "Symmetric positive-definite matrices: from geometry to applications and visualization," in *Visualization and Processing of Tensor Fields, Mathematics and Visualization*, eds P. J. Weickert and P. H. Hagenpages (Heidelberg: Springer), 285–298.
- Parisi, F., Strino, F., Nadler, B., and Kluger, Y. (2014). Ranking and combining multiple predictors without labeled data. *Proc. Natl. Acad. Sci. U.S.A.* 111, 1253–1258. doi: 10.1073/pnas.1219097111
- Pennec, X., Fillard, P., and Ayache, N. (2006). A riemannian framework for tensor computing. *Int. J. Comput. Vis.* 66, 41–66. doi: 10.1007/s11263-005-3222-z
- Ries, A. J., and Larkin, G. B. (2013). *Stimulus and Response-Locked P3 Activity in a Dynamic Rapid Serial Visual Presentation (RSVP) Task DTIC Document*. Final report, Army Research Laboratory
- Sajda, P., Pohlmeier, E., Wang, J., Parra, L. C., Christoforou, C., Dmochowski, J., et al. (2010). In a blink of an eye and a switch of a transistor: cortically coupled computer vision. *Proc. IEEE* 98, 462–478. doi: 10.1109/JPROC.2009.2038406
- Schalk, G., McFarland, D. J., Hinterberger, T., Birbaumer, N., and Wolpaw, J. R. (2004). Bci2000: a general-purpose brain-computer interface (BCI) system. *IEEE Trans. Biomed. Eng.* 51, 1034–1043. doi: 10.1109/TBME.2004.827072
- Tu, W., and Sun, S. (2012). A subject transfer framework for EEG classification. *Neurocomputing* 82, 109–116. doi: 10.1016/j.neucom.2011.10.024
- Tuzel, O., Porikli, F., and Meer, P. (2007). "Human detection via classification on riemannian manifolds," in *IEEE Conference Computer Vision and Pattern Recognition (CVPR)* (Minneapolis, MN: IEEE Computer Society).
- U.S. Department of the Army (1990). *Use of Volunteers as Subjects of Research*. AR 70–25. Washington, DC: Government Printing Office.
- U.S. Department of Defense Office of the Secretary of Defense (1999). *Code of Federal Regulations, Protection of Human Subjects*. 32 CFR 219. Washington, DC: Government Printing Office.
- van Erp, J., Lotte, F., and Tangermann, M. (2012). Brain-computer interfaces: beyond medical applications. *Computer* 45, 26–34. doi: 10.1109/MC.2012.107
- Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–91. doi: 10.1016/S1388-2457(02)00057-3

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

Copyright © 2016 Waytowich, Lawhern, Bohannon, Ball and Lance. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.