Proceedings IEEE International Symposium on Biomedical
Imaging: From nano to macro(2009), pp. 1003–1006.

EEG CLASSIFICATION BY ICA SOURCE SELECTION OF LAPLACIAN-FILTERED DATA

Claudio G. Carvalhaes[*,†], *Marcos Perreau-Guimaraes*[*], *Logan Grosenick*[*], *and Patrick Suppes*[*]

[*] Center for the Study of Language and Information,
Ventura Hall, 200 Panama St., Stanford University, Stanford, CA, 94305-4101, USA

[†]Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro,
Rio de Janeiro, RJ, 20550-900, Brazil

ABSTRACT

We studied the performance of a double-spatial filtering method for classification of single-trial electroencephalography (EEG) data that couples the spherical surface Laplacian (SL) and independent component analysis (ICA). This method was evaluated in the context of a binary classification experiment with brain states driven by mental imagery of auditory and visual stimuli. A statistically significant improvement was achieved with respect to the rates provided by raw data and by data filtered by either SL or ICA.

***Index Terms*—** EEG classification, surface Laplacian, ICA, BCI.

1. INTRODUCTION

The accurate trial-by-trial classification of electroencepalography (EEG) signals is currently a topic of great interest in the field of EEG-based Brain Computer Interfaces (BCIs). Usually high classification rates are only achieved by means of time-intensive multi-stage analyses involving preprocessing, feature reduction, feature extraction, and the classification procedure itself. Such elaborate procedures are necessary in part because each EEG channel is actually measuring a complex mixture of electrical contributions from spatially distinct sources within the brain, and thus the data representation of EEG is non-trivial. Moreover, EEG recordings are affected by volume conduction, limited spatial sampling, reference electrode issues, channel contamination, and various other complications, the net effect of which is a poor spatial resolution that limits classification accuracy.

Modern approaches for EEG classification focus on finding a suitable data representation that maximizes class separation. A special role is played by methods based on linear transformations of the input data, which have advantages in terms of interpretability and computational complexity. Following this line of reasoning, our current work proposes a two-step procedure for EEG classification that combines surface Laplacian (SL) filtering [1] and independent component analysis (ICA) [2]. This combination improves both spatial resolution and signal separation while relying only on linear transformations of the data. Briefly, SL filtering is performed on raw data as a preprocessing step in order to correct for volume conduction effects, thus enhancing the spatial resolution of the data. Next ICA is applied to transform the data into independent sources that are then explored in the classification stage. The performance of this procedure (which we refer to as SL-ICA method) was evaluated using single-trial classification of real data (with basic linear discriminant analysis, or LDA, as the classifier). In developing SL-ICA, we state the rank-deficiency property of spline Laplacians and show how to make SL-transformed data work with ICA.

2. EXPERIMENTAL PROCEDURE

Seven subjects (S1-S7) ranging in age from 21-25 (M = 23) completed 600 two-second trials each involving one of two possible tasks. Participants were instructed to to form mental images of either the English word “go”, uttered by a native male voice digitized at 22 kHz and presented via symmetrically placed computer speakers, or a “stop” sign flashed on a computer monitor. Immediately following the presentation of one of these stimuli for 300 ms, subjects saw 700 ms of blank screen followed by a 300 ms presentation of a small white “+” sign (hereafter referred to as the “fixation-point”). During the fixation-point, they were instructed to imagine as vividly as possible the stimulus (either auditory “go” or visual “stop”) they had just experienced as though they were experiencing it again. Participant imagining was followed by another 700 ms blank screen after which the trial ended.

During each experiment session, the electroencephalogram (EEG) was recorded using a Neuroscan quick-cap with 64 channels (Neurosoft, Inc., Sterling, USA). All EEG channels were referenced to linked mastoid electrodes. Two linked channels were placed at the outer canthi to record the horizontal electrooculogram (HEOG), and two more linked channels on the sub- and supra-orbital ridges of the left eye for vertical electrooculogram (VEOG). The other 62 at scalp locations were from the 5% system [3], not including electrodes Nz,

AF1, AF2, AF5, AF6, T9, T10, P9, P5, P6, P9, P10, PO or I. The analog signal was amplified and band-passed at 0.1–300 Hz and an additional 60 Hz notch filter was used to suppress line noise. The filtered signal was then digitized at a sample rate of 1 kHz. Each session consisted of 30 blocks containing 20 trials per block, and subjects were given regular breaks between blocks (the duration of which they controlled via keypress). The EEG signals were recorded continuously across stimulus presentation and imagination tasks. In this study we examine only the last 1000ms of each trial in which participants vivdly imagined the stimuli. Prior to recording, the subject was given the opportunity to practice over a sample of stimuli.

3. METHODS

3.1. SL spatial filtering

Our computations were carried out using the global method of spherical splines [4, 5]. Spline methods perform the Laplacian differentiation on a spline-constructed analytic function that gives the potential over the whole scalp based on a given set of instantaneous measurements. This analytic function is defined in terms of a set of univariate functions $\phi_1, \cdots, \phi_M$ and a positive definite bivariate function K, according to [6]

$$f(\mathbf{r}) = \sum_{i=1}^{N} c_i \, K(\mathbf{r}, \mathbf{r}_i) + \sum_{i=1}^{M} d_i \, \phi_i(\mathbf{r}). \tag{1}$$

Here N is the number of measurement electrodes, $\mathbf{r}_i$ is the location of the i-th electrode, and $M < N$ is a constant. For convenience, we introduce the time-dependent column vectors $\mathbf{c} = (c_1, \cdots, c_N)^T$ and $\mathbf{d} = (d_1, \cdots, d_M)^T$, the square matrix $\mathbf{K} = \{K(\mathbf{r}_i, \mathbf{r}_j)\}$ and the N-by-M matrix $\mathbf{T} = \{\phi_j(\mathbf{r}_i)\}$. The spline approach assumes that $\mathbf{T}$ is full rank, i.e. $\mathbf{T}$ is rank M. The coefficients $\mathbf{c}$ and $\mathbf{d}$ depend on the measured potentials $\mathbf{v}(t) = (V_1(t), \cdots, V_N(t))^T$ and are determined by [6]

$$(\mathbf{K} + \lambda \mathbf{I})\, \mathbf{c} + \mathbf{T}\, \mathbf{d} = \mathbf{v}, \tag{2a}$$

$$\mathbf{T}^T \mathbf{c} = 0, \tag{2b}$$

where λ is a positive parameter regulating the trade-off between minimizing the square error of the fit and the smoothness. The formal solution for this system can be cast into the form $\mathbf{c} = \mathbf{C}\,\mathbf{v}$ and $\mathbf{d} = \mathbf{D}\,\mathbf{v}$, where $\mathbf{C} \in \mathbb{R}^{N \times N}$ and $\mathbf{D} \in \mathbb{R}^{M \times N}$ are data-independent matrices given by $\mathbf{C} = (\mathbf{K} + \lambda \mathbf{1})^{-1}(\mathbf{1} - \mathbf{TD})$ and $\mathbf{D} = [\mathbf{T}^{\mathrm{T}}(\mathbf{K} + \lambda \mathbf{1})^{-1}\mathbf{T}]^{-1}\mathbf{T}^{\mathrm{T}}(\mathbf{K} + \lambda \mathbf{1})^{-1}$.

The spline SL estimation is computed by applying the SL operator Δ_{surf} to Eq. (1). If the calculation is constrained to the electrode sites only, then one obtains the spline SL expressed as the linear transformation

$$\mathbf{y}(t) = \mathbf{L}_\lambda \, \mathbf{v}(t), \tag{3}$$

where $\mathbf{L}_\lambda$ is a square matrix of N^2 elements, given by

$$\mathbf{L}_\lambda = \tilde{\mathbf{K}}\, \mathbf{C} + \tilde{\mathbf{T}}\, \mathbf{D}. \tag{4}$$

Here $\tilde{\mathbf{K}} \in \mathbb{R}^{N \times N}$ and $\tilde{\mathbf{T}} \in \mathbb{R}^{N \times M}$ are matrices which (i, j)-entries are $\Delta_{\text{surf}} K(\mathbf{r}_i, \mathbf{r}_j)$ and $\Delta_{\text{surf}} \phi_j(\mathbf{r}_i)$, respectively. The Laplacian differentiation of K is computed with respect to its first argument. For convenience, herein we denote by $\mathbf{Y}$ the SL representation of a data matrix $\mathbf{X}$, containing the EEG realization of N channels (rows) and T time samples (columns), i.e.

$$\mathbf{Y} = \mathbf{L}_\lambda \, \mathbf{X}. \tag{5}$$

The computation of $\mathbf{L}_\lambda$ does not involve labels, so from a machine learning standpoint, the spline SL is linear unsupervised learning method for analyzing EEG. Matrix $\mathbf{L}_\lambda$ depends critically on the assumed geometric model and the coordinates of the electrode locations. For a given λ, the same $\mathbf{L}_\lambda$ can be used to transform any signal, provided the electrode setup remains the same. This property may represent an enormous advantage if preprocessing time and memory consumption are concerns. We remark that the linear transformation of Eq. (5) is not a general property of surface Laplacians, but rather a consequence of the spline approach.

For the spherical approach used in this study [4, 5, 6], the SL operator reads

$$\Delta_{\text{surf}} \equiv \frac{1}{r^2 \sin\theta} \frac{\partial}{\partial \theta} \left(\sin\theta \frac{\partial}{\partial \theta} \right) + \frac{1}{r^2 \sin^2\theta} \frac{\partial^2}{\partial \phi^2}, \tag{6}$$

where (r, θ, ϕ) is a point coordinate in the conventional system of spherical coordinates [7]. The kernel K is a symmetric harmonic function given by

$$K(\mathbf{r}, \mathbf{r}_i) = \frac{1}{4\pi} \sum_{n=1}^{\infty} \frac{2n+1}{n^m (n+1)^m} P_n(\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}_i), \tag{7}$$

where carets denote unit vectors, $m > 1$ is a constant (called interpolation order), and P_n is the Legendre polynomial of n-th degree. The summation in (7) converges uniformly, and in practice a few terms are enough to provide a good estimation of $K(\mathbf{r}, \mathbf{r}_i)$, regardless interpolation order. Throughout this work we restricted (7) to its first 50 terms, thus restricting the Legendre polynomials to degree $n \leq 50$. By construction, $M = 1$ and the only function ϕ in the summation in (1) corresponds to the spherical harmonic of degree zero and order zero [6], i.e. $\phi(\mathbf{r}) = 1$. Thus, $\tilde{\mathbf{T}}$ is a null matrix and the spherical SL-transforming matrix reads

$$\mathbf{L}_{m,\lambda} = \tilde{\mathbf{K}}\, \mathbf{C}. \tag{8}$$

The free parameters m and λ are determined in order to optimize the performance of the method. According to [8], the optimal values for these parameters are matched in a small regular grid defined by $m = 2, \cdots, 5$ and $\lambda = 10^{-3}$, 10^{-4},

10^{-5}, 10^{-6}. Based on preliminary computations, all SL estimations throughout this work were carried out using $m = 4$ and $\lambda = 10^{-5}$.

The fact that $\tilde{\mathbf{T}}$ is null entails that $\mathbf{L}_{m,\lambda}$ is a rank deficient matrix. This in turn implies that the generated SL data set $\mathbf{Y}$ is also rank deficient. To clarify this point we evoke the definitions of $\mathbf{C}$ and $\mathbf{D}$ to state that $\mathbf{C}$ is rank $N - 1$ (i.e. $N - M$). Hence, because the rank of a product of two matrices is less than or equal to the rank of either factor (and $\tilde{\mathbf{K}}$ is assumed to be full rank), we obtain from (8) that $\mathbf{L}_{m,\lambda}$ is also rank $N - 1$. Therefore $\mathbf{L}_{m,\lambda}$ and $\mathbf{Y}$ are both rank deficient. It worth to remark that this deficiency is not a particular characteristic of the spherical Laplacian. The same occurs for any spline-based approach [9, 10] for which the set of functions $\{\phi_i\}_{i=1}^{M}$ contains eigenvectors of the SL operator Δ_{surf} associated with a null eigenvalue, thus leading $\tilde{\mathbf{T}}$ to be rank-deficient.

3.2. The SL-ICA combination

ICA performs statistical separation of mixed signals into stationary sources. Numerous studies have shown the usefulness of ICA decompositions in the identification of EEG features, e.g., alpha waves and steady-state responses, for artifact removal, and its applicability in EEG in general. Although ICA methods are widely discussed in the literature, the SL-ICA approach for EEG classification requires additional attention. Standard ICA algorithms do not converge when the input matrix is of defective rank, like those of SL-transformed data. This problem occurs in the sphering step, which produces a decorrelated data matrix in order to speed convergence. Sphering consists in first zero-meaning the rows of the input matrix and then performing a linear transformation by means of the inverse of the principal square root of the covariance matrix. The result is a decorrelated data ensemble whose covariance matrix is diagonal.

Assuming the SL data matrix $\mathbf{Y}$ has zero-mean rows, sphering leads to

$$\mathbf{Y}_{\text{S}} = 2\,[\text{cov}(\mathbf{Y}^T)]^{-1/2}\,\mathbf{Y}. \tag{9}$$

Clearly, this transformation can only be performed if the covariance matrix $\text{cov}(\mathbf{Y}^T)$ is invertible. However, this requirement is not met because $\mathbf{Y}$ is rank deficient and hence $\text{cov}(\mathbf{Y}^T)$ is singular. To solve this problem, we removed the null space of the input matrix prior to sphering. As a consequence, the number of spatial components of $\mathbf{Y}$ was reduced from N to $N - 1$. This was also the maximum number of ICA components in each SL-ICA data matrix.

3.3. Classification

Efficacy of the SL-ICA method was evaluated using single-trial classification of the visual "stop" and auditory "go" imagination data. For each subject, the classification was independently performed over raw data, data filtered by ICA or SL only, and data filtered using SL-ICA. Classification accuracy was used as the criteria to compare the methods.

Single-trials were classified into either "stop" or "go" by using the linear discriminant classification (LDC) procedure described in [11]. The classification accuracy was defined as the ratio of number of trials correctly classified over the total number of trials. A 10×10-fold cross validation was used to avoid overfitting and to estimate extra-sample classification accuracy [12]. The 600 trials of each data set were divided into 10 mutually non-overlapping sets of 60 trials. Then training was repeated 10 times, each time leaving out one set for testing. The test set was classified using the best classifier selected during training. The entire procedure was repeated 10 times, with each set being used as the testing set once, thus providing 10 classification rates for each input matrix.

Feature reduction was performed to decrease the processing time and avoid undersampling. First raw data were downsampled by a ratio of 16:1, cutting off frequencies higher than 31.25 Hz. Then, based on a preliminary search for optimal length, each trial of the downsampled data was reduced to half of its size, keeping only the first segment of 500 ms. With this procedure, the number of spatio-temporal features in the data set corresponded to 1922 features (31 samples per trial $\times$ 62 channels). SL and ICA filtering were processed only after performing these two reductions. The number of features was further reduced in the cross-validation by employing SVD. The goal of SVD was to find a smaller set of k feature vectors providing the highest classification rate for the training set. The optimal k was searched in a loop in the inner session of 10 folds, with k varying in the range from 1 to 200. Thus the training started with the training set reduced to its first 200 SVD feature vectors and following the features were removed one by one until only the most significant one (first) remained. The optimal k was defined based on the maximum rate achieved in the 2000 classifications (10 inner folds $\times$ 200 classifications per fold) performed in the training. The test set was classified using the best classifier and the optimal number of features.

The classifications of ICA and SL-ICA filtered data were performed in two stages. First, the independent sources were rank ordered by means of single-source classifications, following the 10×10-fold cross validation procedure described above. The 10 rates obtained for each source were averaged and used to sort the sources in descending order. Then multi-source classifications were performed with an additional loop in the training to maximize the rates with respect to the number of independent sources. In total, 80,000 classifications were performed in the training, compounded by 10 folds, 200 k-values, and 40 different configurations of ICA sources.

4. RESULTS

The 10 classification rates from the cross-validation were averaged to yield the mean classification rates presented in Table 1. The SL-ICA method improved the classification rates of other methods in all cases. The enhancement in accuracy with respect to the raw data ranged between 0.7% (subject S1) and 11.2% (subject S7).

Table 1. Mean classification rates in %.

Subject	Raw	ICA	SL	SL-ICA
S1	83.0	80.0	81.2	83.7
S2	60.3	63.2	57.7	65.8
S3	65.5	69.3	66.7	69.5
S4	67.8	68.2	67.8	68.7
S5	64.7	68.2	67.5	69.7
S6	54.2	52.8	59.5	60.2
S7	60.3	65.8	69.8	71.5
Mean $\pm$ STD	66.4$\pm$8.4	67.0$\pm$7.3	68.6$\pm$7.8	70.4$\pm$7.3

Paired *t*-tests were carried out to determine whether differences observed from Table 1 are statistically significant or not. Such comparisons have much more statistical power when the difference between methods is small relative to the variation within methods, as in Table 1. The computations ran over the sets of 10 non-averaged classification rates. Thus, a group of 70 rates (7 subjects $\times$ 10 rates) were used in each comparison, concatenating the rates of all subjects. The following p-values were obtained for the null hypothesis that SL-ICA is equal to the other methods against the alternative that SL-ICA outperforms the other methods: 3.59×10^{-6} (SL-ICA *vs*. raw data), 6.05×10^{-4} (SL-ICA *vs*. ICA), and 6.06×10^{-3} (SL-ICA *vs*. SL). These values confirm that SL-ICA outperforms the other methods at a convincing level of significance.

5. DISCUSSION AND CONCLUSION

The power of SL and ICA in EEG research in general and specially in multi-channel EEG classification is widely reported in the literature. A large number of simulations and experimental studies have shown that SL representations reduce the spatial blur of scalp potentials, thus emphasizing localized activity–provided the investigation uses a dense electrode setup (64 electrodes or more). Furthermore, SL has also the advantage of being reference-free and produces a reliable picture of cortical potential distributions generated by radial dipoles or distributed neocortical sources [1]. ICA, in turn, is a powerful statistical tool for solving the Blind Source Separation (BSS) problem given multiple channels, and is based on higher-order statistics or temporal decorrelation [2]. These two methods were combined here to utilize the full potential of both techniques in EEG classification. Statistical evaluations of the outcomes of classification tasks of a mental imagery experiment confirmed our expectations at a convincing level of significance.

References

[1] P.L. Nunez and R. Srinivasan, *Electric Fields of the Brain: The Neurophysics of EEC, 2nd ed.*, Oxford University Press, New York, 2006.

[2] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja, *Independent Component Analysis*, John Wiley & Sons, Inc., 2001.

[3] R. Oostenveld and P. Praamstra, “The five percent electrode system for high-resolution EEG and ERP measurements,” *Clin. Neurophysiol.*, vol. 112, no. 4, pp. 713–719, 2001.

[4] F. Perrin, J. Pernier, O. Bertrand, and JF Echallier, “Spherical splines for scalp potential and current density mapping,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 72, no. 2, pp. 184–187, 1989.

[5] F. Perrin, J. Pernier, O. Bertrand, and JF Echallier, “Corrigendum EEG 02274,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 76, pp. 561, 1990.

[6] G. Wahba, “Spline models for observational data,” in *CBMS-NSF Regional Conference Series in Applied Mathematics, 1987, Philadelphia, SIAM*, 1990.

[7] John David Jackson, *Classical Electrodynamics, 3rd ed.*, Wiley, New York, 1999.

[8] F. Babiloni, C. Babiloni, L. Fattorini, F. Carducci, P. Onorati, and A. Urbano, “Performances of surface Laplacian estimators: A study of simulated and real scalp potential distributions,” *Brain Topography*, vol. 8, no. 1, pp. 35–45, 1995.

[9] F. Perrin, J. Pernier, O. Bertrand, MH Giard, and JF Echallier, “Mapping of scalp potentials by surface spline interpolation,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 66, no. 1, pp. 75–81, 1987.

[10] F. Babiloni, C. Babiloni, F. Carducci, L. Fattorini, P. Onorati, and A. Urbano, “Spline Laplacian estimate of EEG potentials over a realistic magnetic resonance-constructed scalp surface model,” *Electroencephal. Clin. Neurophysiol.*, vol. 98, no. 4, pp. 363–373, 1996.

[11] Marcos Perreau Guimaraes, Dik Kin Wong, E. Timothy Uy, Logan Grosenick, and Patrick Suppes, “Single-Trial Classification of MEG Recordings,” *IEEE Trans. on Biom. Eng.*, vol. 54, no. 3, pp. 436–443, 2007.

[12] Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York, 2001.