

Single-trial detection of event-related fields in MEG from the presentation of happy faces : Results of the Biomag 2016 data challenge

H. Cecotti¹, A. Barachant², J.R. King³, J. Sanchez Bornot¹, G. Prasad¹

Abstract—The recognition of brain evoked responses at the single-trial level is a challenging task. Typical non-invasive brain-computer interfaces based on event-related brain responses use electroencephalography. In this study, we consider brain signals recorded with magnetoencephalography (MEG), and we expect to take advantage of the high spatial and temporal resolution for the detection of targets in a series of images. This study was used for the data analysis competition held in the 20th International Conference on Biomagnetism (Biomag) 2016, wherein the goal was to provide a method for single-trial detection of event-related fields corresponding to the presentation of happy faces during the rapid presentation of images of faces with six different facial expressions (anger, disgust, fear, neutrality, sadness, and happiness). The datasets correspond to 204 gradiometers signals obtained from four participants. The best method is based on the combination of several approaches, and mainly based on Riemannian geometry, and it provided an area under the ROC curve of 0.956 ± 0.043 . The results show that a high recognition rate of facial expressions can be obtained at the signal-trial level using advanced signal processing and machine learning methodologies.

I. INTRODUCTION

Brain-Computer Interfaces (BCIs) are mostly based on electroencephalography (EEG) recordings. While there exists the possibility to obtain a high spatial resolution with high density EEG (e.g., 128 channels), it requires a long time for the preparation of a user and can become an obstacle when dealing with BCI for day-today usage. Moreover, where the performance related to the detection of a command is the most critical aspect, other types of brain imaging techniques may be used, such as magnetoencephalography (MEG) that has several advantages over EEG signals [1]. MEG is an efficient brain imaging technique that can enhance pre-surgical planning by localizing relevant brain regions in a non-invasive way. Compared to high density EEG, the time to prepare a subject is significantly reduced because there is no need to optimize the contact between the scalp and the sensors. MEG is currently the best available non-invasive brain imaging technique for research in human brain dynamics [2]. MEG is however mainly used for clinical studies and neuroscience research, and it recently got some attention for neural engineering applications [3], [4].

This paper deals with the detection of facial expression during a serial visual presentation task at the single-trial level, where different images of faces with particular facial

expressions are presented to the user. The main challenge is to extract relevant features to differentiate between happy faces from other faces. Different methods have been proposed in the literature for single-trial detection [5]. This type of task has been used with EEG recordings for the detection of faces [6], and threat detection [7]. In addition to the feat of brain decoding, the detection of particular facial expressions can provide key insights about patients with traumatic brain injury (TBI). Patients with TBI are indeed found to be significantly impaired on expression labelling and matching [8]. The single-trial detection of faces with a particular expression may therefore provide a neuromarker related to TBI. The remainder of this paper is organized as follows: First, the experimental protocol is presented in Section II, then the participants of the competition and the results are given in Section III. Finally, the different methods are detailed in Section IV.

II. EXPERIMENTAL PROTOCOL

Four healthy volunteers participated in the study (age= 31.5 ± 5.6 , 2 females). All participants provided written informed consent, reported normal or corrected-to-normal vision, and no history of neurological problems. The experimental protocol was reviewed by the Faculty Ethics Filter Committee of Ulster University, and was in accordance with the Helsinki Declaration of 1975, as revised in 2000.

Visual stimuli consisted of grayscale images (335×419 pixel) from the FACES database (2052 images), which was developed at the Max Planck Institute for Human Development (MPIB), Berlin, Germany between 2005 and 2007 [9]. The images were centered on the screen (visual angle $\approx 20^\circ$). Participants were seated comfortably 100 cm from the screen in a darkened electromagnetically shielded chamber. They were asked to avoid moving during the experiments to avoid muscular artifacts. Participants had to pay attention to a stream of images (presentation rate=1 Hz (stimulus onset asynchrony=1000 ms, stimulus duration=333 ms)) (40 blocks of 12 images, each block contains faces from the same person, but with different facial expressions corresponding to 6 different classes: anger, disgust, fear, neutrality, sadness, and happiness). The goal of the task was to detect the presence of images with a happy facial expression, by pressing a button. The goal of the data analysis is to detect the presence of a face with happiness by using only the MEG signal and the stimulus onsets.

The data was processed and available directly in the Matlab format as a structure containing: the signal corresponding to the 204 planar gradiometers, the stimulus onsets from the

¹ Northern Ireland Functional Brain Mapping Facility, Ulster University, Magee Campus, Derry ~ Londonderry, UK h.cecotti@ulster.ac.uk

² Burke Medical Research institute, Cornell University, New York NY, USA.

³ Department of Psychology, New York University, New York NY, USA

images and the behavioral responses (triggers). The stimulus onsets are given for each facial expression: t1: Anger (non-target), t2: Disgust (non-target), t3: Fear (non-target), t4: Happiness (target), t5: Neutrality (non-target), t6: Sadness (non-target). The data analysis competition provided the class labels of the stimuli for the first half of the recording session, for each subject. The task of the competition was to predict in which trials of the second half of the session, the participant was presented a happy face, just from MEG data. The evaluation was to be performed with the area under the ROC curve (AUC). For training, there were 40 triggers for each type of face, and for the test, there were 240 unlabelled triggers (6×40).

A. Signal acquisition

The MEG signal was recorded with an Elekta Neuromag 306-channel MEG system in the Northern Ireland Functional Brain Mapping (NIFBM) Facility of the Intelligent Systems Research Centre (ISRC), Ulster University, Derry/Londonderry, UK. The signal was recorded with a sampling rate of 1 kHz using 204 planar gradiometers and 102 magnetometers. Five head position indicator (HPI) coils were placed on the head of the participants to determine how close the head is to the sensors that are collecting the signal. The Neuromag software Maxfilter 2.2 that implements Signal-Space Separation (SSS) was used to preprocess the signal [10]. The method separates magnetic signals coming from within the brain from those coming from outside. Finally, the signal was bandpass-filtered between 0.1 Hz and 41.66 Hz, and then downsampled to 125 Hz.

III. RESULTS

Six teams participated in this data analysis competition. Each team provided a vector containing the classifier outputs for the 240 unlabelled epochs of each participant. The composition of the teams are as followed: **Team 1**: Alexandre Barachant and Jean-Remi King, from France. **Team 2**: Emanuele Olivetti and Paolo Avesani, NeuroInformatics Laboratory (NILab), Bruno Kessler Foundation, Center for Mind and Brain Sciences (CIMeC), University of Trento, Italy. **Team 3**: Cristian Grozea, Fraunhofer Institute FOKUS, Berlin, Germany. **Team 4**: Zafer Ican, Centre for Cognition and Decision Making, National Research University, Higher School of Economics, Russian Federation. **Team 5**: Andrea Vitale, Institute for advanced study (IUSS), Pavia, Italy and Christian Salvatore, National Research Council (CNR) at the Institute of Molecular Bioimaging and Physiology (IBFM), Milano, Italy. **Team 6**: Mohammed Abdulaal, School of EEE The University of Manchester, UK.

The results are presented in Table I for each team. The best method, from Team 1, provides an AUC of 0.956 ± 0.043 .

IV. METHODS

Here, we discuss and highlight the superiority of the single-trial detection approach adopted by the winning team. The best approach consisted in ensembling three classification pipelines built to extract both evoked and induced

TABLE I
AUC FOR EACH TEAM.

	s1	s2	s3	s4	Mean	SD	Position
Team1	0.982	0.893	0.962	0.986	0.956	0.043	Gold
Team2	0.940	0.778	0.819	0.930	0.867	0.081	Silver
Team3	0.808	0.655	0.734	0.890	0.772	0.101	Bronze
Team4	0.756	0.559	0.577	0.745	0.659	0.106	4th
Team5	0.511	0.520	0.618	0.505	0.538	0.054	5th
Team6	0.524	0.498	0.545	0.466	0.508	0.034	6th

responses. This work relies on functions already available in the toolboxes ¹. The complete method is available online [11]. For a comparative evaluation, the methods used other teams are also briefly described.

A. Classification Pipelines

The three pipelines rely on Riemannian geometry classifiers fitted on the covariance matrices of each trial. These pipelines made use of the tangent space mapping followed by a logistic regression [12], [13]. The main difference among these pipelines resides in the definition of the features space and in the estimation of the covariance matrices.

a) Covariance estimation: Let $\mathbf{X}_i \in \mathfrak{R}^{C \times N}$ denote an epoch (trial) of index i with C the number of channels and N the number of time samples. Let be y_i be the class (target (happy) or non-target (not happy)) of \mathbf{X}_i . The spatial covariance matrix $\Sigma_i \in \mathfrak{R}^{C \times C}$ of \mathbf{X}_i is estimated using the sample covariance matrix estimator (SCM):

$$\Sigma_i = \frac{1}{N} (\mathbf{X}_i - \bar{\mathbf{x}}_i) (\mathbf{X}_i - \bar{\mathbf{x}}_i)^T \quad (1)$$

where $\bar{\mathbf{x}}_i \in \mathfrak{R}^C$ is the average of the trial across time. For each pipeline, the covariance estimation only varies in the way \mathbf{X}_i is built. To ensure that the matrices are symmetric and positive definite (a requirement to use Riemannian geometry), regularization can be applied on the estimated covariances matrices. To this end, the Ledoit-wolf Shrinkage [14] or the Oracle Approximation Shrinkage [15] were used.

b) Tangent Space mapping: Riemannian geometry provides a natural way to manipulate and measure difference between Symmetric Positive Definite matrices. Tangent space mapping were used as a way to take into account the manifold structure while having a vector-representation of the matrices. For any covariance matrix Σ_i , we define its “tangent vector” $\mathbf{s}_i \in \mathfrak{R}^{C(C+1)/2}$ as:

$$\mathbf{s}_i = \text{upper} \left(\log \left(\Sigma_\mu^{-1/2} \Sigma_i \Sigma_\mu^{-1/2} \right) \right) \quad (2)$$

where \log is the matrix logarithm, upper is the operator consisting in taking the upper triangular part of the matrix, applying a coefficient $\sqrt{2}$ on its off-diagonal elements, and vectorizing the results. Finally, Σ_μ is also a covariance matrix, defining the reference point of the tangent space. This reference point was chosen to be the log-Euclidean mean of

¹MNE v.18, pyRiemann, Scikit-Learn v.17

all training data:

$$\Sigma_\mu = \exp\left(\frac{1}{N} \sum_i^N \log(\Sigma_i)\right) \quad (3)$$

1) *ERPCov*: This first pipeline is made of: 1) a decomposition of the spatial filters with a Principal Component Analysis (PCA) for dimensionality reduction (70 components), 2) an estimation of the ERF Covariance to capture both evoked and induce responses, 3) a Common Spatial Pattern (CSP) to reduce the dimensionality of the covariance matrices (30 components), 4) a mapping to the tangent space (with log-euclidean mean as reference point) to meaningfully vectorize the data, and 5) a penalized logistic regression for the final classification stage. The ERPCov pipeline relies on a special form of covariance matrix estimation that embeds the temporal information about the evoked response. This estimation consists in the concatenation, along the channel axis, of the averaged ERF of each class before estimating the spatial covariance matrix:

$$\tilde{\mathbf{X}}_i = \begin{bmatrix} \mathbf{P}_t \\ \mathbf{P}_{nt} \\ \mathbf{X}_i \end{bmatrix} \quad (4)$$

where \mathbf{P}_t and \mathbf{P}_{nt} are the averaged (across trial) evoked response for the target and non-target class, respectively. The total dimension of the matrices is $\tilde{C} = 3 \times C = 3 \times 70 = 210$. The resulting covariance matrix is thus composed by the cross-covariance of the trial with the prototypical evoked response of each class and by the standard covariance of the trial. Consequently, if the trial \mathbf{X}_i contains a target evoked response synchronized to the prototypical target response, then it will produce a specific cross-covariance structure with the target prototype. If the trial contains a task-related induced activity, the presence of the sample covariance matrix of \mathbf{X}_i will allow its detection. This type of covariance estimation makes use of both evoked and induced responses within the same pipeline.

2) *XdawnCov*: This pipeline is composed by 1) PCA for dimensionality reduction (50 components), 2) XdawnERP Covariance estimation, 3) Tangent space mapping (with log-euclidean mean as reference point), and 4) logistic Regression for the final classification stage. The XdawnCov pipeline is similar to ERPCov, except that a Xdawn spatial filtering [16], [17] is applied before the covariance estimation in order to increase the SNR of the ERP response. For each class, a set of Xdawn spatial filters are estimated and applied on the data. Let \mathbf{V}_t and $\mathbf{V}_{nt} \in \mathfrak{R}^{C \times K}$ be the spatial filters corresponding to the class target and non target, respectively, then we get:

$$\tilde{\mathbf{X}}_i = \begin{bmatrix} \mathbf{V}_t^T \mathbf{P}_t \\ \mathbf{V}_{nt}^T \mathbf{P}_{nt} \\ \mathbf{V}_t^T \mathbf{X}_i \\ \mathbf{V}_{nt}^T \mathbf{X}_i \end{bmatrix} \quad (5)$$

In this pipeline, $K = 12$ Xdawn filters for each class are used; the total dimension of the covariance matrices is therefore $\tilde{C} = 4 \times K = 48$.

3) *HankelCov*: This pipeline is made of 1) PCA for dimensionality reduction (70 components), 2) Hankel Covariance estimation (delays 1, 8, 12 and 64 time samples), 3) CSP for reduction of covariance matrices (15 components), 4) Tangent space mapping (with log-euclidean mean as reference point), and 5) Logistic Regression for the final classification stage. This estimation is obtained by concatenating multiple time-lagged versions of each trial. This estimation detects the auto-correlation and the cross-correlation between the channels, and it is similar to the Common Spatio-Spectral Pattern algorithm.

$$\tilde{\mathbf{X}}_i = \begin{bmatrix} \delta(n_1, \mathbf{X}_i) \\ \vdots \\ \delta(n_j, \mathbf{X}_i) \\ \vdots \\ \delta(n_J, \mathbf{X}_i) \end{bmatrix} \quad (6)$$

where $\delta(n_j, \mathbf{X}_i)$ is an operator that adds a lag of n_j to the data \mathbf{X}_i . In this case, we chose a logarithmically spaced set of five lags: 0, 1, 8, 12, 64. The dimension of the covariance matrices is therefore $\tilde{C} = 5 * C = 350$.

B. Evaluation

Since the experimental design was not trully randomized, but consisted in shuffled sequences of 12 faces (with two target in each group), a post-processing step on the predictions was applied to ensure that the sum of probabilities of each group of 12 trials was 1. The fusion of the three pipelines was achieved by the summation of their individual predictions. To further improve the robustness of the classification, and to reduce overfitting, results for different set of epoching offset : 10, 20, 30, 40 and 50 time samples with a constant window of 150 time samples were also combined [18]. With a first evaluation on the training (10-fold cross-validation), the best pipeline was ERPCov with an average AUC of 0.972. The averaged scores of the 3 classifiers lead to an AUC score of 0.979.

C. Other teams

1) *Team 2*: The method comprises three stages. First, each epoch is represented as the covariance matrix of the signals measured by 204 MEG gradiometers. Second, each covariance matrix is represented in the tangent space of the associated Riemannian manifold [12], [13]. Third, the output is obtained with a logistic regression classifier. The time window started at 120ms (15 time points) from the stimulus onset, and lasted until 1320 ms (165 time points) from stimulus onset. Such values were defined in order to take into account high-level visual processing and premotor/motor activation during the experiment. With a 10-fold cross-validated ROC AUC over the training set of each subject, the average of the AUC across subjects was 0.93.

2) *Team 3*: For the preprocessing steps, the first ten principal components were extracted, the signal was filtered with a low-pass Butterworth filter of order 6 and cut-off frequency 4 Hz. In each epoch, the chosen time window was

1800 ms after stimulus onset. The classifier was a random forest [19].

3) *Team 4*: Noisy channels were removed by analyzing the raw data and power spectra of the MEG channels after filtering the signal with band-stop (45-55 Hz) and band-pass (1-49 Hz) filters. In the preprocessed dataset, after detrending the data, the signal was filtered with a 2nd degree Butterworth band-pass filter (0.58 Hz). For the classification, the window length was chosen as 80 samples (640 ms) for the responses. Then, the T-Weight method was used after using Principal Component Analysis (PCA) to remove correlations in the data [20].

4) *Team 5*: For each epoch, the chosen time window was 800 ms after stimulus onset, corresponding to 100 time points. The number of trials equalized across conditions: 40 for target and non-target. Feature selection was achieved with a 10 fold cross validation procedure. For each fold, 35 features were extracted with PCA, followed by Fisher Discriminant Ratio (9 features). The classification step was based on Support Vector Machine fitted to all the possible combination of the selected features, with testing on the remaining 1/K of the training data. The most accurate classifier's parameters are applied to predict the unlabelled epochs.

5) *Team 6*: This hierarchical approach is decomposed into two stages. First, the classification is performed by classifying (Sadness, Anger and Fear) versus (Disgust, Neutrality and Happiness). In the second stage, the classification is performed with (Disgust, Neutrality) versus and Happiness. In both stages, the signal from each electrode was subtracted from its following electrode to form a signal with 102 channels. Then, the common spatial patterns method was used, and only the best 6 spatial filters were used to extract features. The classification was achieved with a fitted linear discriminant analysis function.

V. CONCLUSION

Brain decoding at the single-trial level represents a fundamental challenge for machine learning and neuroscience, with significant impact in neural engineering. The best method was based on a multiple-classifier system using a hierarchy of pre-processing steps, with features based on Riemannian geometry, and taking into account the underlying manifold of the data. The extent to which the motor response had a contribution in the facial expressions detection has to be investigated. Further work will be carried out with a larger number of the subjects to determine neuromarkers related to face detection with particular emotions.

ACKNOWLEDGMENT

The authors would like to thank all the participants for their effort and participation in this competition. H.C., J.B., and G.P. are supported by the Northern Ireland Functional Brain Mapping Facility (1303/101154803), funded by InvestNI and Ulster University.

REFERENCES

- [1] D. Cohen and B. Cuffin, "Demonstration of useful differences between the magnetoencephalogram and electroencephalogram," *Electroencephalogr Clin Neurophysiol*, vol. 56, pp. 38–51, 1983.
- [2] M. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa, "Magnetoencephalography: theory, instrumentation, and applications to noninvasive studies of the working human brain," *Rev. Mod. Phys.*, vol. 65, no. 2, pp. 413–497, Apr. 1993.
- [3] H. Cecotti, "Single-trial detection with magnetoencephalography during a dual rapid serial visual presentation task," *IEEE Trans. on Biomed. Eng.*, vol. 63, no. 1, pp. 220–7, 2016.
- [4] J. P. Kauppi, M. Kandemir, V. M. Saarinen, L. Hirvenkari, L. Parkkonen, A. Klami, R. Hari, and S. Kaski, "Towards brain-activity-controlled information retrieval: Decoding image relevance from MEG signals," *Neuroimage*, vol. 112, pp. 288–98, May 2015.
- [5] H. Cecotti and A. J. Ries, "Best practice for single-trial detection of event-related potentials: Application to brain-computer interfaces," *Int. J. of Psychophysiology*, pp. 1–15, 2016.
- [6] J. Touryan, L. Gibson, H. J. Horne, and P. Weber, "Real-time measurement of face recognition in rapid serial visual representation," *Frontiers in Psychology*, vol. 2, no. 42, pp. 1–8, 2011.
- [7] A. Marathe, A. Ries, V. Lawhern, B. Lance, J. Touryan, K. McDowell, and H. Cecotti, "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, 2015.
- [8] R. E. Green, G. R. Turner, and W. F. Thompson, "Deficits in facial emotion perception in adults with recent traumatic brain injury," *Neuropsychologia*, vol. 42, no. 2, pp. 133–41, 2004.
- [9] N. C. Ebner, M. Riediger, and U. Lindenberger, "Faces - a database of facial expressions in young, middle-aged, and older women and men: Development and validation," *Behavior Research Methods*, vol. 42, pp. 351–362, 2010.
- [10] S. Taulu and J. Simola, "Spatiotemporal signal space separation method for rejecting nearby interference in meg measurements," *Phys. Med. Biol.*, vol. 51, pp. 1–10, 2006.
- [11] J.-R. King and A. Barachant, "Team-bk/biomag2016 1.0," Nov. 2016. [Online]. Available: <https://doi.org/10.5281/zenodo.165999>
- [12] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain-computer interface classification by riemannian geometry," *IEEE Trans. on Biomed. Eng.*, vol. 59, no. 4, pp. 920–928, Apr. 2012.
- [13] —, "Classification of covariance matrices using a riemannian-based kernel for BCI applications," *Neurocomputing*, vol. 112, pp. 172–178, July 2013.
- [14] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, vol. 88, no. 2, pp. 365–411, 2004.
- [15] Y. Chen, A. Wiesel, Y. C. Eldar, and A. O. Hero, "Shrinkage algorithms for MMSE covariance estimation," *IEEE Trans. on Signal Processing*, vol. 58, no. 10, pp. 5016–5029, Oct. 2010.
- [16] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert, "xDawn algorithm to enhance evoked potentials: application to brain-computer interface," *IEEE Trans Biomed. Eng.*, vol. 56, no. 8, pp. 2035–43, 2009.
- [17] B. Rivet and A. Souloumiac, "Optimal linear spatial filters for event-related potentials based on a spatio-temporal model: Asymptotical performance analysis," *Signal Processing*, vol. 93, no. 2, pp. 387–398, 2013.
- [18] H. Cecotti, A. Marathe, and A. J. Ries, "Optimization of single-trial detection of event-related potentials through artificial trials," *IEEE Trans. on Biomed. Eng.*, vol. 62, no. 9, pp. 2170–6, 2015.
- [19] T. K. Ho, "Random decision forests," in *Proc. of the 3rd Int. Conf. on Document Analysis and Recognition*, 1995, pp. 278–282.
- [20] Z. Iscan, "MLSP competition, 2010: Description of second place method," in *Proc. of the IEEE Int. Workshop on Machine Learning for Signal Processing*, 2010, pp. 114–115.