35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

EyeCatch: Data-mining over Half a Million EEG Independent Components to Construct a Fully-Automated Eye-Component Detector*

Nima Bigdely-Shamlo, Ken Kreutz-Delgado, *Fellow, IEEE,* Christian Kothe and Scott Makeig

Abstract— **Independent component analysis (ICA) can find distinct sources of electroencephalographic (EEG) activity, both brain-based and artifactual, and has become a common pre-preprocessing step in analysis of EEG data. Distinction between brain and non-brain independent components (ICs) accounting for, e.g., eye or muscle activities is an important step in the analysis. Here we present a fully automated method to identify eye-movement related EEG components by analyzing the spatial distribution of their scalp projections (scalp maps). The EyeCatch method compares each input scalp map to a database of eye-related IC scalp maps obtained by data-mining over half a million IC scalp maps obtained from 80,006 EEG datasets associated with a diverse set of EEG studies and paradigms. To our knowledge this is the largest sample of IC scalp maps that has ever been analyzed. Our result show comparable performance to a previous state-of-art semi-automated method, CORRMAP, while eliminating the need for human intervention.**

I. INTRODUCTION

Finding EEG sources through the application of ICA data decomposition has become a popular EEG analysis method [1-6]. An important step in analyzing EEG using ICA is separating brain source processes from the contributions to the scalp data from muscle and eye-movement related processes [7]. There are several algorithms proposed for this task: ADJUST [8] is a fully automatic algorithm that uses a combination of spatial and temporal features of independent components (ICs) to classify blinks, eye movements, and generic discontinuities. The method is based on a handful of spatial features (e.g., variance differences across groups of channels) manually constructed in a trial and error manner. When temporal information is not available, or when the EEG epochs are too short to obtain reliable statistics on temporal features, the performance of the ADJUST algorithm is not established. CORRMAPP [9] is a semi-automated method that classifies eye-related ICs solely based on the correlation of their spatial projections (scalp maps) with one or few templates. Each template is initially specified by the user and later refined by iterative clustering and averaging of detected eye components.

Here we present EyeCatch, a method that uses a large database of exemplar eye scalp maps instead of the single user-initiated template in CORRMAP. The exemplar database is generated by analysis of a very large set of IC scalp maps from multiple studies to capture relevant eye component topographies while being robust to normal variations in subject anatomy, electrode locations, ICA decomposition quality, etc.

II. METHODS

A. Scalp maps Database Preprocessing

We first gathered 106,749 single-subject EEG data sets from file servers of the UC San Diego Swartz Center for Computational Neuroscience (data collected during the period 2002-2012) and selected those with an ICA decomposition (nearly all by Extended Infomax [4] or AMICA [10, 11]) and unique dipolar IC source models computed using EEGLAB [6, 12]. From the selected 80,006 data sets we extracted 638,512 distinct IC scalp maps interpolated on a 67×67 2-D scalp grid using *topoplot()* in EEGLAB.

B. Eye-related template scalp map dataset

The eye-related scalp map template dataset was created in two stages. First we selected a single eye-movement related template scalp map from an RSVP study we knew well [13] and calculated its correlations with the 265 scalp maps from three other laboratory studies. The ten IC scalp maps most highly correlated with the template were visually judged to be eye-activity related and added to the eye-related IC scalp map template

---

*Research supported by Research was sponsored by the Army Research Laboratory under Cooperative Agreement Number W911NF-10-2-0022 and NIH grant 1R01MH084819-03.

N.B-S. and K.K-D. are with the Electrical and Computer Engineering Department and Swartz Center for Computational Neuroscience, Institute for Computational Neuroscience, University of California San Diego, CA 92093 USA and (e-mail: nima@sccn.ucsd.edu).

C.K. and S.M. are with Swartz Center for Computational Neuroscience, Institute for Computational Neuroscience, University of California San Diego, CA 92093 USA.

database. Next, we sorted 499 IC scalp maps from an Attention-Shift study [14] by their maximum correlation to any of the scalp maps in the template database and visually selected 25 eye-activity related component scalp maps to add to the template database.

Next we calculated the highest absolute correlation between all 638,512 distinct IC scalp maps (section A) and any of the eye-related scalp maps in the template database. After sorting by this value and visual inspection, the scalp maps most highly correlated with any template map (max(|r|)>0.994) were clustered into 24 clusters using Affinity Propagation [15]. Sixteen of these clusters mostly contained scalp maps associated with a single type of eye-related activity (e.g., vertical or horizontal eye movements, or eye blinks). The rest were considered to be brain source ICs whose maps had some similarity to eye-activity related maps. We then visually inspected each of the sixteen eye-related scalp map clusters, and retained only scalp maps that were more similar than a visually appropriate correlation threshold to the cluster exemplar (cluster thresholds: 0.8<|r|<0.97; median 0.94). After final visual adjustment (eliminating 13 ICs) we obtained a template database of 3,452 eye-activity related IC scalp maps.

The EyeCatch algorithm then simply calculates the maximum absolute correlation between an input scalp map and all 3,452 eye-activity related template scalp maps in its database. Cross validation results showed that this typically was more reliable than more complex nearest-neighbor distance weighted averaging methods.

III. RESULTS

Fig. 1 shows a sample 96 IC scalp maps in the EyeCatch template database. Many of these represent variations on a single type of template (e.g., accounting for EEG artifact produced by horizontal eye movements or eye blinks) arising from differences in subject anatomy, electrode locations, etc. Including this variability provides an advantage when using a simple similarity-based classification method and can be achieved only by processing data from a large sample of subjects and recording conditions.

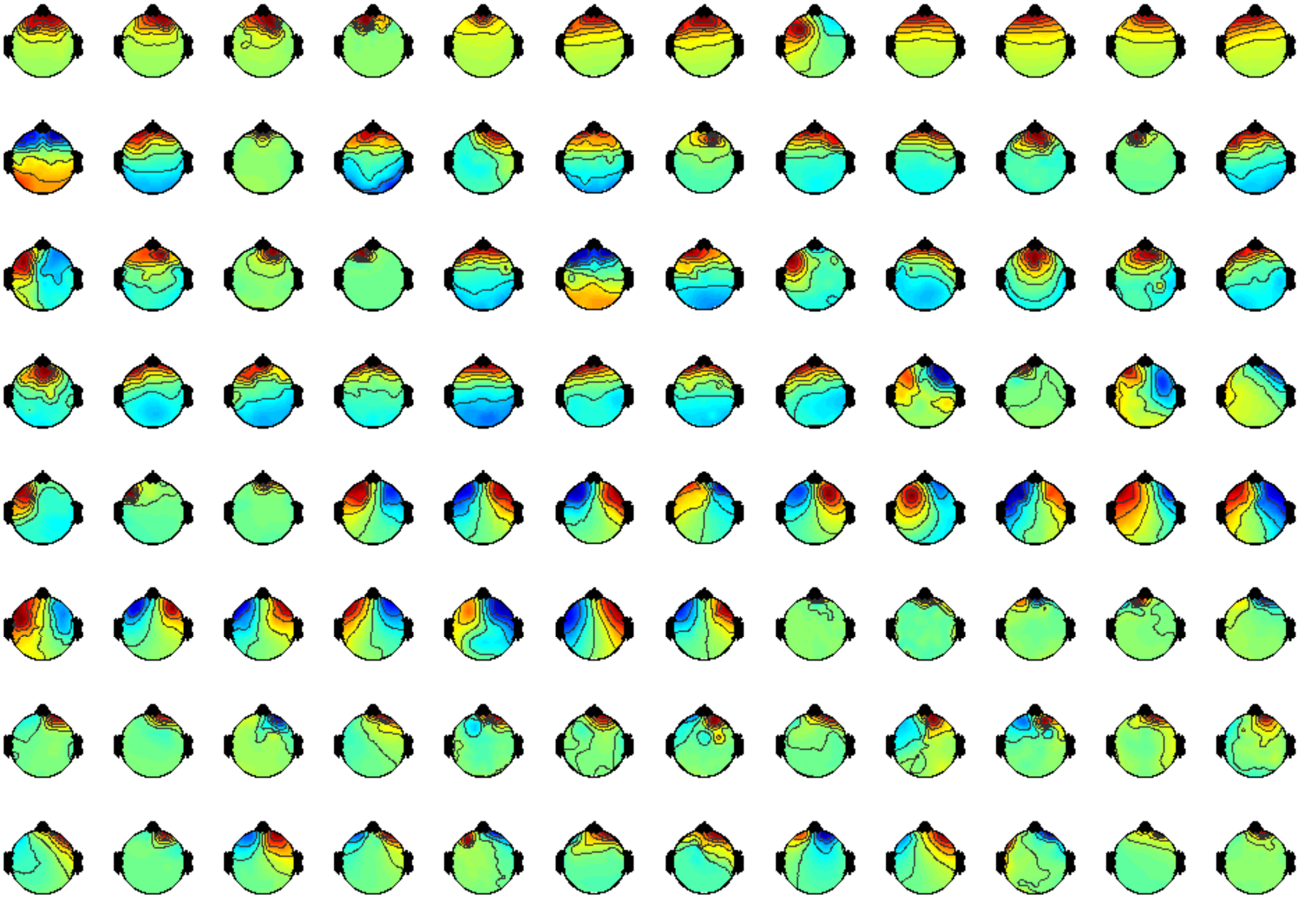

Fig. 1. A sample 96 template component scalp maps (of 3,452) in the EyeCatch eye-related component template database.

We compared the performance of EyeCatch with the reported results of the semi-automatic CORMAP algorithm. The 4,256 IC scalp maps used in the CORRMAP paper [9] plus ratings of these maps by eleven experts were kindly provided to us by the authors of [9]. We applied EyeCatch to these scalp maps using a range of decision correlation thresholds (between 0.95 and 0.99) and compared the results to the average of the [0|1] votes from the 11 experts who judged each given IC scalp map as either accounting for eye-movement activity (e.g., blinks or lateral eye movements) or not. Using Matlab (Mathworks, Inc.) 7.85 s were required to obtain maximum correlation values for the 4,256 input maps (1.8 ms per map). Figure 2 shows the correlations between the EyeCatch output (length 4,256 vector of binary [0|1] values] and the expert vote averages (vector of range [0,1] values) for a range of EyeCatch maximum-correlation decision thresholds.

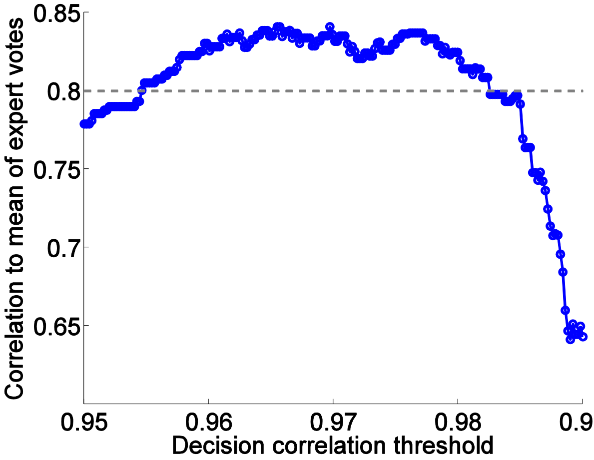

Fig. 2. Correlations between eye-activity related component scalp map judgments by EyeCatch and the average votes (whether each component is eye activity related or not) from eleven experts as a function of the EyeCatch maximum-correlation decision threshold.

We also calculated the Receiver Operator Characteristic (ROC) curve [16] using the majority vote of the 11 experts as binary ground truth (thereby identifying 125 lateral eye movement or blink-related scalp maps) and the maximum absolute correlation similarity between each test scalp map and the 125 scalp maps in the EyeCatch template database as the detection variable. Fig. 3 displays this ROC curve. The area under the ROC curve is 0.993, demonstrating that EyeCatch has both high sensitivity and specificity.

IV. CONCLUSIONS

As seen in Fig. 2, for a range of decision correlation thresholds (from 95.5% to 98.3%) the ROC area is above 0.8. This is highly comparable to the reported performance of CORRMAP, for which mean correlations with expert judgments for each study were 0.85-0.91 for lateral eye movements and 0.83-0.99 for blinks. However, EyeCatch results did not involve the user interaction required by CORRMAP.

Our results show that high-performance eye-related IC classification can be achieved by using a large volume of data and relatively simple measures (here, scalp map correlation thresholding). This suggests that solving other problems in EEG analysis, from muscle-related component detection to robust Brain Computer Interface design, may also benefit from exploiting large databases spanning many EEG studies.

However, still better performance for detecting both eye-activity and other non-brain ('artifact') IC types might be obtained by jointly considering IC scalps and time courses. For example, saccade and blink ICs have strong, fairly predictable time domain features; ICs accounting for scalp muscle (electromyographic, EMG) activity have characteristic spectral profiles, etc.

A freely available, open-source implementation of the EyeCatch algorithm running on Matlab is available in the Measure Projection Toolbox (MPT), an EEGLAB plug-in [17]. Documentation and stand-alone downloads are available at http://sccn.ucsd.edu/wiki/EyeCatch.

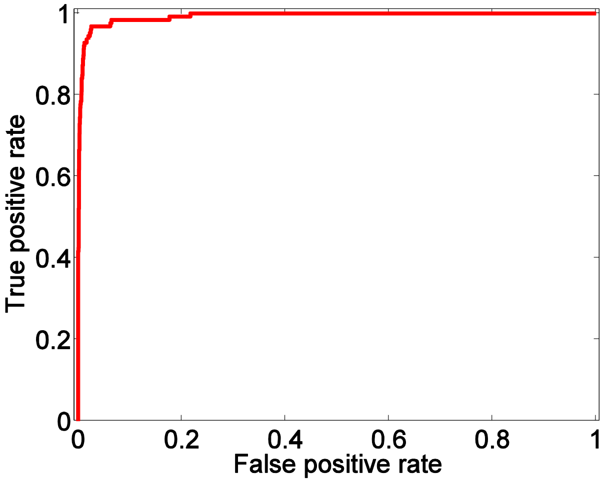

Fig. 3. Receiver Operator Characteristic (ROC) curve for EyeCatch scalp map classification and expert majority voting on the CORRMAP paper component scalp map collection (area under the curve = 0.993).

ACKNOWLEDGMENT

This research was sponsored by the Army Research Laboratory under Cooperative Agreement Number W911NF-10-2-0022 and by NIMH grant 1R01-MH084819-03. 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 Army Research Laboratory or the U.S Government. The U.S Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

REFERENCES

[1] A. J. Bell and T. J. Sejnowski, "An Information Maximization Approach to Blind Separation and Blind Deconvolution," *Neural Computation,* vol. 7, pp. 1129-1159, Nov 1995.

[2] S. Makeig, A. J. Bell, T. P. Jung, and T. Sejnowski, "Independent component analysis of electroencephalographic data," *Advances in Neural Information Processing Systems,* vol. 8, pp. 145-151 1996.

[3] S. Makeig, T. P. Jung, A. J. Bell, D. Ghahremani, and T. J. Sejnowski, "Blind separation of auditory event-related brain responses into independent components," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 94, pp. 10979-10984, Sep 30 1997.

[4] T. W. Lee, M. Girolami, and T. J. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources," *Neural Computation,* vol. 11, pp. 417-441, Feb 15 1999.

[5] T. P. Jung, S. Makeig, M. J. McKeown, A. J. Bell, T. W. Lee, and T. J. Sejnowski, "Imaging brain dynamics using independent component analysis," *Proceedings of the Ieee,* vol. 89, pp. 1107-1122, Jul 2001.

[6] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods,* vol. 134, pp. 9-21, Mar 15 2004.

[7] T. P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T. J. Sejnowski, "Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects," *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology,* vol. 111, pp. 1745-58, Oct 2000.

[8] A. Mognon, J. Jovicich, L. Bruzzone, and M. Buiatti, "ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features," *Psychophysiology,* vol. 48, pp. 229-240, Feb 2011.

[9] F. C. Viola, J. Thorne, B. Edmonds, T. Schneider, T. Eichele, and S. Debener, "Semi-automatic identification of independent components representing EEG artifact," *Clinical Neurophysiology,* vol. 120, pp. 868-877, May 2009.

[10] J. Palmer, K. Kreutz-Delgado, and S. Makeig, "Super-Gaussian mixture source model for ICA," *Independent Component Analysis and Blind Signal Separation,* pp. 854-861, 2006.

[11] J. Palmer, S. Makeig, K. Delgado, and B. Rao, "Newton method for the ICA mixture model," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, 2008, pp. 1805-1808.

[12] A. Delorme, T. Mullen, C. Kothe, Z. Akalin Acar, N. Bigdely-Shamlo, A. Vankov, and S. Makeig, "EEGLAB, SIFT, NFT, BCILAB, and ERICA: new tools for advanced EEG processing," *Computational Intelligence and Neuroscience,* vol. 2011, 2011.

[13] N. Bigdely-Shamlo, A. Vankov, R. R. Ramirez, and S. Makeig, "Brain Activity-Based Image Classification From Rapid Serial Visual Presentation," *Ieee Transactions on Neural Systems and Rehabilitation Engineering,* vol. 16, pp. 432-441, Oct 2008.

[14] R. Čeponienė, M. Westerfield, M. Torki, and J. Townsend, "Modality-specificity of sensory aging in vision and audition: Evidence from event-related potentials," *Brain research,* vol. 1215, pp. 53-68, 2008.

[15] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science,* vol. 315, pp. 972-976, Feb 16 2007.

[16] J. P. Egan, *Signal detection theory and ROC-analysis*. New York: Academic Press, 1975.

[17] N. Bigdely-Shamlo et al, *Measure Projection Toolbox*. WWW publication: http://www.sccn.ucsd.edu/wiki/MPT, 2012.