33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Robust EEG Preprocessing for Dependence-Based Condition Discrimination

Bilal H. Fadlallah, Sohan Seth, Andreas Keil and José C. Príncipe, Fellow, IEEE

Abstract— **This paper addresses the robustness of the filtering schemes in processing high resolution electroencephalogram (EEG) data in the context of discriminating two stimuli flickering at a given frequency. The raw data consists of recordings from a 128-channel HydroCell GSN where the subject was visually stimulated with two images flickering at 17.5 Hz, representing two distinct conditions, referred to as Face and Mock. These signals were then passed through a band pass filter to only capture the modulation at the flickering frequency, and a connectivity analysis was performed on the filtered signal using generalized measure of association, to observe if the network connectivity changes from one stimulus to the other. In this paper, we investigate the effect of the bandpass filter on the discriminability of the stimuli over different filter orders and quality factors. We observe that the network connectivity is stable over a significant range of parameter values of the filter, thus establishing the desired robustness.**

I. INTRODUCTION

The problem of determining how different parts of the human brain connect functionally, has been the focus of much research in the past decades. Current methods of studying cognitive interactions vary in their spatial-temporal resolution depending on the approach used, such as synaptic, hemodynamic, or nuclear. Recently, fMRI has been the tool of choice to study brain connectivity as it exhibits a high spatial resolution. However, its dependence on the blood flow drastically reduces its temporal resolution ($\sim$ 1 sec). Given that a rough time estimate for neural communication i.e. the time for a path completion from a neuron to another ranges between 1 ms and 100 ms depending on the neurons characteristics [1], fMRI is not suitable for exploring cognitive tasks where time scales of interest do not exceed 100 ms. This actually applies to all the currently available MRI based brain imaging techniques. The high temporal resolution ($\sim$ 1 ms) of EEG or MEG (magnetoencephalogram), on the other hand, makes them particularly useful for studying the dynamics of the brain. Although combining EEG recordings with fMRI imaging has also been suggested [2], such method introduces residual artifacts in the EEG, that is mainly caused by the cardioballistogram (BCG) and the changing fields applied during the fMRI image acquisition [3], thus making the process of analyzing the resulting signals even a harder task.

Manuscript received April 15, 2011. This work was partially funded by NSF grant 0964197 and the CNRS, Lebanon

B. Fadlallah, S. Seth and J. Príncipe are with the Department of Electrical and Computer Engineering at the University of Florida, Gainesville, FL, {bhf, sohan, principe}@cnel.ufl.edu

A. Keil is with the Department of Psychology, University of Florida, Gainesville, FL, (akeil@ufl.edu)

EEG signal, however, suffers from both poor spatial resolution and signal-to-noise ratio. To circumvent the latter, we exploit the steady state visual evoked potential (ssVEP) i.e. the modulation created in the visual cortex by the flickering stimulus in the range of 8 to 20 Hz [4]. The basic idea of using ssVEP is to concentrate the analysis on the power modulations at the flickering frequency to study the connectivity among the visual cortex and the rest of the brain regions, rather than exploring the entire signal. Therefore, following this method, the task of exploring the functional connectivity essentially boils down to two independent subtasks of first, implementing an appropriate bandpass filter to extract the desired modulation, and second, to employ an appropriate dependence measure to analyze the network connectivity. The acquired connectivity can then be exploited to infer the flow of information over time or to discriminate between two different stimuli.

Choosing an appropriate bandpass filter, however, is crucial since depending on the bandwidth of the filter, we can either pass a single frequency, thus destroying the modulation, or the entire signal, thus sending unneccessary information. Both these situations disrupt the connectivity analysis by creating artifically high or low dependence, and thus, making any inference or discrimination difficult. Therefore, it is natural to question the realiability of this method, and to investigate if the method produces meaningful dependence values over a certain range of bandwidths, way from the two extremes. In this paper, we evaluate the robustness of this method to the variation in the design of appropriate bandpass filter, in the context of appropriately discriminating two stimuli, referred to as Face and Mock. To assess the allowable variation in the network connectivity, we employ Kolmogorov-Smirnov (KS) test between the dependence values induced by Face and Mock stimuli, respectively, over each pair of channels.

The rest of the paper is organized as follows. In section II, we describe the experimental setting i.e. we discuss how to clean the raw signal from noisy components via notch filters, and design a scheme to bandpass this signal around the flickering frequency using filters of different orders and quality factors. Here, we also introduce the notion of generalized measure of association, a novel measure of dependence [5], to calculate the dependence between the filtered signals. In section III we apply the Kolmogorov-Smirnov test as a way to assess the variation in the network connectivity, and describe a suitable experiment to evaluate the robustness of the bandpass filters. Finally in section V we conclude with a summary and suggestions for future work.

II. METHODS

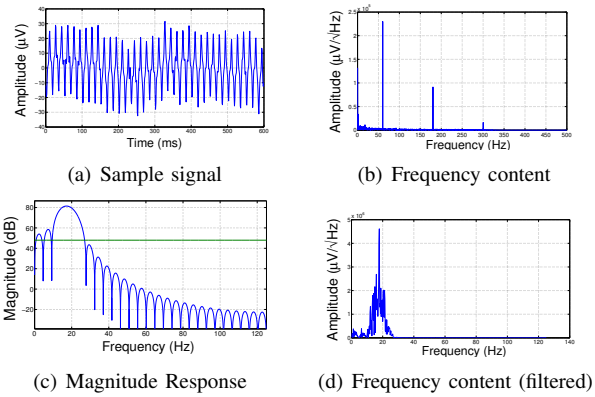
A. Experimental Setting

The dataset consists of electrode recordings from a 128 Hydro-Cell Geodesic Sensor Net (HCGSN) montage, using a geodesic tessellation of the head [6]. The subject's visual cortex is stimulated with two types of images flickering at a frequency of 17.5 ± 0.20 Hz, where the uncertainty arises from the discrepancy of the analog display frequency. The first condition shows a human face (to which we refer as "Face") and the second displays a flickering pattern of stripes (to which we refer as "Mock"). It is expected from cognitive studies that the subjects will react differently to these two types of images since it is well known that the brain has evolved special circuitry to decode faces [7][8][9][10]. The sampling rate is 1000 Hz with 15 trials performed for each condition to ensure consistency. The baseline segment of each channel consists of the first 400 samples, and the remaining 4200 samples correspond to the condition (Face or Mock). The recorded datasets can therefore be represented as spatio-temporal data matrices $\mathbf{X}^{(k)} \in \mathbb{R}^{N_S \times P}$ per trial k for Face and Mock each, where N_S and P denote the number of channels and number of sampled data points, respectively. It is well known that there are several factors that interfere with the process of having a faithful display of brain activity when recording an EEG signal. The effect of some of these artifacts can be attenuated (like the subject's motion during the recording session and to some extent the blinking of the eyes), whereas others are uncontrollable such as the electrical activity of some muscles, electrocardiograms and especially the effect of volume conduction. To reduce the effect of the latter, we follow the procedure described in [11], in which the current source density (CSD) is derived from interpolating the scalp potentials at sites between electrodes.

B. Signal Processing

We proceed by enhancing the frequency band of interest (centered at the flickering frequency f_o) and attenuating all other frequencies. Fig. 1 shows the Fourier transform of the raw recording which contains visible noise components at odd harmonics of the fundamental 60 Hz line noise. We use second order Butterworth filters to notch these frequencies. Note that an implementation with Chebyshev (Type I) filters is also valid. However, we opt for the former since: (i) we're interested in reducing the effect of the phase non-linearity introduced even if working relatively far from the band of interest (phase response is more near-linear) and (ii) for the same order, Butterworth has a better shaping factor than a Chebyshev Type I filter. Filters are 2^{nd} order, centered at the odd harmonics with a quality factor defined as $Q = f_c/\Delta f_c$ equal to 20, where f_c denotes the filter center frequencies and Δf_c the corresponding frequency band at 3 dB.

For bandpassing, we need a linear-phase filter since a constant group delay is desired when measuring the dependence between signals in time domain. We can design such filter simply by allocating N zeros on the unit circle and using pairs of conjugate poles to control the filter bandwidth. This



(a) Sample signal (b) Frequency content

(c) Magnitude Response (d) Frequency content (filtered)

Fig. 1. (a) 600 samples of the original signal corresponding to 600 ms. (b) Frequency content of the signal in (a) with huge spikes at line noise frequencies. (c) Magnitude response of a bandpass filter with $Q = 1$ and $N = 54$. (d) Filtered signal after passing through notch filters and the bandpass filter in (c).

can be illustrated for the case when $N = 40$ by setting the transfer's function numerator to $N(z) = z^N - 1$ and setting the center of the filter on $n = 3\delta$ with a pass band of 2δ, where $\delta = f_d/N = f_s/(dN) = 250/40 = 6.25$ Hz and $d = 4$ being the downsampling factor. In this case, the transfer function would look like:

$$H(z) = \frac{z^N - 1}{\prod_{i \in \{n_1,\ldots,n_\alpha\}} (z^2 - 2cos(\theta_i) + 1)}$$

where $\theta_i = (2\pi)i/N$ and $\{n_1, ..., n_\alpha\}$ refer to angles of adjacent zeros from both sides of the center frequency such that α varies according to whether we want a filter with low quality factor or not. Fig. 1 illustrates an example on a sample channel.

Although this design approach works, it does not prove to be handy in our context since it allows only little flexibility in varying the quality factor for the same order. Therefore, we instead use a least-square linear-phase FIR filter for which the problem can be stated as follows: $\mathbf{h}_{N,Q} = \min_{\mathbf{h}} ||\mathbf{\Gamma h} - \mathbf{D}||^2$, where $\mathbf{h}_{N,Q} = [h(n)]$ denotes the impulse response of a Type II (even symmetry) filter of length $N + 1$ (N even), $\mathbf{D} = [D(w_r)]$ is a length R vector containing the ideal response at a set of frequencies $\{w_r\}$. $\mathbf{D}$ implicitly defines the quality factor Q and the center frequency f_c. Row t of $\mathbf{\Gamma}$ is given by $[1, 2\cos(w_t), \ldots, 2\cos(w_t(N-1))]$ where $t \in [0, \ldots, N-1]$. This expression of $\mathbf{\Gamma}$ can be obtained by expressing the filter as $H(w_r) = \mathbf{h}_{N,Q}(0) + 2\sum_{n=0}^{N/2} \mathbf{h}_{N,Q}(n)\cos(w_r n)$ with $r \in [0, \ldots, N-1]$ as of even symmetry and shifting by $L/2$ samples. The problem reduces to a least-square optimization where $\hat{\mathbf{h}}_{N,Q} = \arg\min_{\mathbf{h}_{N,Q}} ||\mathbf{\Gamma h}_{N,Q} - \mathbf{D}||$, hence:

$$\hat{\mathbf{h}}_{N,Q} = [(\mathbf{\Gamma}^T\mathbf{\Gamma})^{-1}\mathbf{\Gamma}^T]\mathbf{D} \quad (1)$$

This ideal response estimate is used to test several bandpassing schemes with different orders and quality factors.

C. Generalized Measure of Association

We propose to compute the dependence between the filtered signals for both conditions using the generalized

Algorithm 1: Generalized Measure of Association

Input: Bivariate time series $\{x_t, y_t\}_{t=1}^{n}$ assuming values in the joint space $\mathcal{X} \times \mathcal{Y}$
Output: Estimated dependence $d \in [0:1]$
for $i \in \{1...n\}$ **do**
  Find x_{j^*} closest to x_i, i.e. $j^* = \arg\min_{j \neq i} \delta_x(x_i, x_j)$, where δ_x denotes Euclidean distance in $\mathcal{X}$.
  Find the rank r_i of y_{j^*} in terms of δ_y so that $r_i = \text{count}\{j : j \neq i, \delta_y(y_j, y_i) \leq \delta_y(y_{j^*}, y_i)\}$.
Compute C as the empirical CDF of $\{r_1, \ldots, r_n\}$.
d is the area under C normalized by $(n-1)$

measure of association (GMA) [5]. We choose this measure since, unlike correlation, it is capable of capturing nonlinear structure, and unlike mutual information (MI), it can be easily estimated without any free parameters such as kernel size. The description of computing GMA is given in Algorithm 1. When the time series are independent then the ranks r_i are uniformly distributed over $\{1, \ldots, n-1\}$ and therefore GMA is closer to 0.5, whereas when the time series are almost the same, then the ranks are close to 1 and therefore GMA is closer to 1.

III. Testing

It is easy to observe that decreasing the filter's bandwidth, or alternatively increasing its quality factor, increases the dependence between two channels of filtered signals, since both this signal become close to a single frequency sinusoid with a specific frequency i.e. the flickering frequency. On the other hand, filters with low quality factors are more affected by noise because of the larger bandwidth, and as a result, the dependences between the filtered channels drop. Therefore, to ensure robustness it should be established that there exists a sufficiently large set of parameter values where the dependences captured among the filtered signals are meaningful and stable. Moreover since our goal is to maximize the separability between the two conditions, Face and Mock, we consider it to be a criteria to judge the robustness of the effect of filter variation.

Since working with all pairs of channels is computationally expensive, we select a subset S from the initial 128-electrode setting to compute the pairwise dependence values per condition, per trial and per filter, and then apply statistical tests to see how effectively we can differentiate the two conditions for each filter. In this paper, we use the two-sample Kolmogorov-Smirnov (KS) test, which is a non-parametric test to compare two sample vectors. The KS test tries to estimate the distance between the empirical distribution functions of the two sets of samples. The null hypothesis is that both samples are drawn from the same distribution. Assuming $\gamma_1(x)$ and $\gamma_2(x)$ to be the sample vectors, the KS statistic can be calculated as:

$$KS_{\gamma_1,\gamma_2} = \max_x |F_{\gamma_2}(x) - F_{\gamma_1}(x)|$$

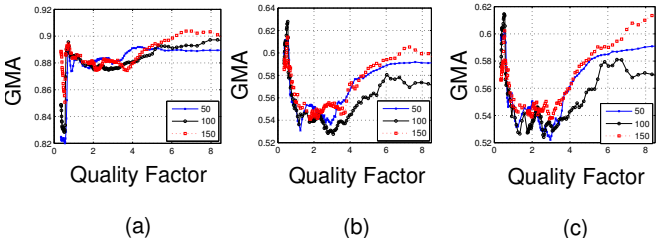


Fig. 2. *GMA vs. quality factor plot for 3 different filter orders. (a) Two channels expected to have high dependence* $(81, 88)$*. (b) and (c) Two channels expected to have low dependence, respectively* $(81, 27)$ *and* $(88, 27)$*.*

where $F_{\gamma_1}(x)$ and $F_{\gamma_2}(x)$ denote the empirical cumulative distribution functions for the n iid observations: $F_{\{X_1,\ldots,X_n\}}(x) = \frac{1}{n}\sum_{i=1}^{n} I_{X_i \leq x}$ where I_k denotes the indicator function. The null hypothesis is rejected at the α-level if $\sqrt{(n_1 n_2)/(n_1 + n_2)} KS_{\gamma_1,\gamma_2} > K_\alpha$, where n_1 and n_2 denote the number of samples from each observation vector and K refers to the Kolmogorov distribution [12]. In our case, we are interested in applying the KS test on the Face and Mock data so $n_1 = n_2$. α is set to 0.1. The complete procedure is detailed in Algorithm 2.

Before proceeding, we check the dependence values for two pairs of channels that we expect to be highly and less dependent, respectively. Fig. 2 shows the GMA values between a pair of channels in the occipital, and a pair having one channel in the occipital and the other in the frontal. Results were averaged over 15 trials to ensure consistency. As expected, we observe that GMA increases with the quality factor for the channels likely to be dependent since the effect of noise fades with the reduction of the passband. For the channels less likely to be dependent, the reduction of the noise level is masked by the differences between the two signals in the passband.

IV. Results and Discussion

Following the procedure described in Algorithm 2 with $N_S = 32$, we obtain the plot shown in Fig. 3. We present the results for only 3 different filter orders out of 20 between 10 and 300, since the effect of the filter order on the shape of the curve is minimal. GMA was evaluated for 250 quality factors in the range $\sim$0.1 and 175. Again, results were averaged over trials. The region corresponding to $Q < 18$ is of special interest because a filter's bandwidth drops below 1 Hz afterwards. Since the readings of any two pair of channels is affected by the delay in signal propagation, we embed the signal in $\tau = 8$ dimension before computing the GMA values. $\tau = 8$ corresponds to 32 ms.

The curve starts with relatively low values corresponding to quality factors in the range $(Q < 0.4)$, then stays almost stable in the range $Q \in [0.6, 1.5]$, and then decreases again. This is inline with our expectations since for a low quality factor, the dependence level is reduced due to the presence of noise and the method looses discriminability, whereas on the other hand, for a high quality factor the dependence values are usually high, and therefore the method again

Algorithm 2: Testing Procedure

Input: S is a subset of size N_S of the original electrode set, f_o the flickering frequency, $\mathbf{O}$ is a vector of M filter orders, $\mathbf{f}$ is a vector of N filter bandwidths, τ is the embedding dimension

Output: An $M \times N$ matrix $\mathbf{R}$ of pairwise KS test decisions per filter order per quality factor

for $m \in \{1, \ldots, M\}$ **do**
  for $n \in \{1, \ldots, N\}$ **do**
    Compute $h_{\mathbf{O}(m),\mathbf{f}(n)}$ as in (1).
    for $i \in \{1 \ldots N_S\}$ **do**
      for $j \in \{1 \ldots N_S\}$ **do**
        for cond $\in \{\text{Mock}, \text{Face}\}$ **do**
          for $k \in \{1 \ldots T\}$ **do**
            Compute $\mathbf{Y}^{(k)(\text{cond})}_{m,n,i,(\cdot)} = \mathbf{h}_{\mathbf{O}(m),\mathbf{f}(n)} \star \mathbf{X}^{(k)(\text{cond})}_{i,(\cdot)}$ and
            $\mathbf{Y}^{(k)(\text{cond})}_{m,n,j,(\cdot)} = \mathbf{h}_{\mathbf{O}(m),\mathbf{f}(n)} \star \mathbf{X}^{(k)(\text{cond})}_{j,(\cdot)}$, where $\star$ denotes convolution.
            Embed filtered series $\mathbf{Y}^{(k)(\text{cond})}_{m,n,i,(\cdot)}$ and $\mathbf{Y}^{(k)(\text{cond})}_{m,n,j,(\cdot)}$ in τ dimension to get $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$
            Set $\mathbf{D}^{(k)(\text{cond})}_{m,n,i,j} = \text{GMA}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$
        Compute $\mathbf{R}_{m,n}$ as $\sum_i \sum_j \text{KS}\left(\mathbf{D}^{(\cdot),\text{face}}_{m,n,i,j}, \mathbf{D}^{(\cdot),\text{mock}}_{m,n,i,j}\right)$.

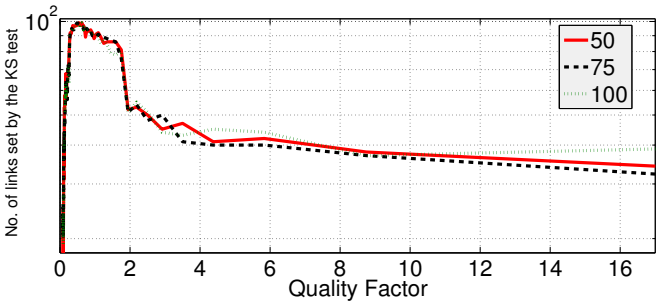


Fig. 3. *Results obtained from Algorithm 2. The values of* $\mathbf{R}(i, j)$ *increase with the quality factor* j*, reaches a plateau and then decreases as the quality factor goes up, clearly showing that there is indeed a region where the estimated dependence values are robust.*

perform worse. The stable performance of the method in the range $Q \in [0.6, 1.5]$ can be justified since for this range of quality factors the bandwidth range becomes $[\sim 11, \sim 29]$. It is possible that this particular bandwidth range covers the modulation of the flickering frequency well, thus extracting the essential information for discriminating the stimuli.

V. CONCLUSION AND FURTHER RESEARCH

In this paper, we have discussed the issue of robustness in discriminability of stimuli for variations in the filter design in the context of processing ssVEP signal. We have evaluated several filter parameters in the context of maximizing separation between the two conditions of interest, which was obtained by applying the KS test on the dependence values induced by these two stimuli over pairs of channels. We observe that the method is indeed robust in terms of the variation in the filter design, producing stable discriminability over a wide range of parameter values.

REFERENCES

[1] J. Anderson, *Cognitive Psychology and Its Implications*, 6th ed., Worth, ISBN 0716701103, NY: pp. 17-18 (2004).

[2] F. Huang-Hellinger, H. Breiter, G. McCormack, M. Cohen, K. Kwong, J. Sutton, R. Savoy, R. Weisskoff, T. Davis, J. Baker, J. Belliveau and B. Rosen, "Simultaneous functional magnetic resonance imaging and electrophysiological recording", *in Human Brain Mapping*, vol. 3, pp. 13-23, (1995).

[3] P. Allen, O. Josephs and R. Turner, "A Method for Removing Imaging Artifact from Continuous EEG Recorded during Functional MRI", *in NeuroImage*, vol. 12, pp. 230-239 (2000).

[4] S. Moratti, B. A. Clementz, Y. Gao, T. Ortiz and A. Keil, "Neural Mechanisms of Evoked Oscillations: Stability and Interaction With Transient Events", *in Human Brain Mapping*, vol. 28, pp. 1318-1333 (2007).

[5] S. Seth, A. Brockmeier, J. Choi, M. Semework, J. Francis and J. Principe, "Evaluating dependence in spike train metric spaces", *in IJCNN 2011* (accepted).

[6] Electric Geodesics, Inc., "Geodesic Sensor Net Technical Manual", pp. 29-30, Online: http://www.egi.com (2007).

[7] P. Lang, M. Bradley and B. Cuthbert, "International affective picture system: Technical manual and affective ratings". Gainesville, FL: NIMH Center for the Study of Emotion and Attention (2005).

[8] A. Keil, T. Gruber, M. Muller, S. Moratti, M. Stolarova, M. Bradley and P. Lang: "Early modulation of visual perception by emotional arousal: Evidence from steady-state visual evoked brain potentials", *in Cognitive, Affective, & Behavioral Neuroscience*, vol. 3, pp. 195-206 (2003).

[9] A. Keil, D. Sabatinelli, M. Ding, P. Lang, N. Ihssen and S. Heim, "Re-entrant Projections Modulate Visual Cortex in Affective Perception: Evidence From Granger Causality Analysis", *in Human Brain Mapping*, vol. 30, pp. 532-540 (2009).

[10] B. Clementz, J. Wang, and A. Keil, "Normal electrocortical facilitation but abnormal target identification during visual sustained attention in schizophrenia," *in Journal of Neuroscience*, vol. 28, no. 50, pp. 13411-13418 (2008).

[11] M. Junghofer, T. Elbert, P. Leiderer, P. Berg and B. Rockstroh, "Mapping EEG-potentials on the surface of the brain: a strategy for uncovering cortical sources", *in Brain Topography*, vol. 9, pp. 203-217 (1997).

[12] G. Marsaglia, W. Tsang and J. Wang, "Evaluating Kolmogorov's Distribution", *in Journal of Statistical Software*, vol. 8, no. 18, pp. 1-4, (2003).