A novel algorithm for removing artifacts from EEG data

Yongcheng Li[1], Po T Wang[2], Mukta P Vaidya[3,4,5], Charles Y. Liu[6], Marc W Slutzky[3,4,5] and An H Do[1]

***Abstract*— In recent years, many studies examined if EEG signals from traumatic brain injury (TBI) patients can be used for new rehabilitation technologies, such as BCI systems. However, extraction of the high-gamma band related to movement remains challenging due to the presence of surface electromyogram (sEMG) caused by unconscious facial and head movement of patients. In this paper, we proposed a modified independent component analysis (ICA) model for EMG artifact removal in the EEG data from TBI patients with a hemicraniectomy. Here, simulated EMG was generated and added to the raw EEG data as the extra channels for independent components calculation. After running ICA, the independent components (ICs) related to artifacts were identified and rejected automatically through several criteria. EEG data underlying hand movement from one healthy subject and one TBI patient with a hemicraniectomy were conducted to verify the efficacy of this algorithm. Results showed that the proposed algorithm removed sEMG artifacts from the EEG data by up to 86.72% while preserving the associated brain features. In particular, the high-gamma band (80 to 160 Hz) was found to arise principally from the hemicraniectomy area after this technique was applied. Meanwhile, we found that the magnitude of gamma power during movement improved after removal of sEMG artifacts.**

***Index Terms*— artifact removal; ICA; traumatic brain injury; neural network; rehabilitation technology**

I. Introduction

Prior research demonstrated that the γ bands (usually from 40 to 160 Hz) from brain signals were strongly related to the movement [1], [2]. In particular, the γ band from the electrocorticogram (ECoG) signal is directly related to some kinematic [3], [4] and kinetic [5] parameters. This information may be employed to develop the rehabilitation technologies for paralyzed patients, such as brain-computer interfaces. However, it is difficult to detect the real γ frequency in the electroencephalogram (EEG) signal due to the filtering effects of the skull. Some research recently demonstrated that EEG γ band can be recorded from the hemicraniectomy site in traumatic brain injury (TBI) patients [6]. Even so, the extraction of the EEG γ band features during movement remains challenging due to the large overlap in the γ band between the EEG signal and surface electromyogram (sEMG), which is primarily caused by unconscious facial and head movement.

For EEG signal processing, there are many sEMG artifact removal algorithms which use independent component analysis (ICA), and other high-order statistical methods in current research work [7], [8]. Although ICA is now considered an important technique for artifact rejection [9], this approach has several problems in removing sEMG noise. For example, since the recording channels are limited, the independent components also are limited, thereby making it difficult to "force" all of the noise into a minimal set of independent components. Another issue is the potential for over-correction or under-correction as users attempt to distinguish between neurogenic and myogenic components. Bias is also introduced as experimenters manually reject components. Hence, it is necessary to develop a technique that can more effectively remove sEMG noise so that it does not confound the neurogenic γ band while not affecting any of the underlying signal features.

In order to efficiently remove the sEMG artifacts, we developed a modified ICA model by adding sEMG noise into the EEG data. Either real or simulated sEMG noise was added to the EEG data, which was then subjected to ICA. Several criteria were established to reject noise ICs automatically. In order to verify this approach, the EEG from a healthy subject and a TBI patient with a left-sided hemicraniectomy were used to test whether the sEMG artifacts can be removed efficiently while preserving the brain features underlying motor behaviors. In particular, the high-γ band (80 to 160 Hz) in the hemicraniectomy area was investigated in this work.

II. Method

A. Experiments

This study was approved by the Institutional Review Board of the University of California, Irvine and the Rancho Los Amigos National Rehabilitation Center. Two experiments were carried out to verify this approach. A healthy subject (Subject 1) was asked to wear a 64-channel EEG cap while signals were acquired 2× Nexus-32 EEG amplifiers (Mind Media, Herten, The Netherlands). The subject performed repetitive right fist pumping in response to commands on a computer screen. The subject alternated between right fist pumping and idling for two seconds for a total of 20 trials.

Real sEMG signals from the frontalis, temporalis, masseter, posterior head muscles were recorded with a sampling rate of 2000 Hz for healthy subject. This EMG data was later

This study was funded by the National Institues of Health, R01NS094748
We thank Angelica Nguyen for her assistance in subjects recruitment

[1]Yongcheng Li and An H Do with the Department of Neurology, University of California, Irvine, CA 92697 USA. Correspondence: and@uci.edu; yongchel@uci.edu

[2]Po T Wang with the Department of Biomedical Engineering, University of California, Irvine, CA 92697, USA

[3]Mukta P Vaidya and Marc W Slutzky with the Department of Neurology, Northwestern University, Chicago, Illinois. [4]Department of Physiology, Northwestern University, Chicago, Illinois. and [5]Department of Physical Medicine and Rehabilitation, Northwestern University, Chicago, Illinois

[6]Charles Y. Liu with the Department of Neurosurgery, University of Southern California. Rancho Los Amigos National Rehabilitation Center, California

978-1-5386-3646-6/18/$31.00 ©2018 IEEE

used in the noise removal process (described in subsection C and D below).

Next, a TBI patient with a left-sided frontal hemicraniectomy (Subject 2) was fitted with a 128-channels EEG cap (ActiCap, Brain Products, Gilching, Germany) and asked to perform a pincer gripping task on the right side while the EEG signals were acquired by Blackrock Neuroport system (Blackrock, Salt Lake City, USA). Gripping events were detected by a force sensor. The subject was asked to perform a pincer grasp with varying levels of force, as guided by a cursor on a computer screen. This was repeated for a total of 20 times over a 120s-long session. All of the data were recorded with a sampling rate of 2000 Hz.

B. ICA model based on adding EMG

To efficiently remove the sEMG noise from the EEG data, we added EMG noise into the EEG datasets and used a modified ICA model as follows:

$$\begin{pmatrix} \hat{X}_t \\ n_\tau^* \end{pmatrix} = A_{t+\tau} \times S_{t+\tau}$$

$$\begin{pmatrix} X_t + b_t \times N_t \\ n_\tau^* \end{pmatrix} = A_{t+\tau} \times \begin{pmatrix} s_t + m_t \\ m_\tau^* \end{pmatrix}$$

where $\hat{X}_t = X_t + b_t \times N_t$, and X_t is the real EEG data, N_t is the real noise, b_t is the linear coefficients, n_τ^* is the added EMG noise, t is the number of the EEG electrodes, τ is the number of the added EMG noise, $A_{t+\tau}$ is the mixing matrix, $S_{t+\tau}$ is the independent component sources, in which s_t is the sources representing the real EEG, m_t are the sources representing the real EMG noise, and m_τ^* are the added EMG sources.

In this model, when the added EMG noise are independent, if n_τ^* is linearly related to N_t, then $m_t = 0$. We presume that $m_t \neq 0$, that is, $N_t = a1_t \times m_t + a2_\tau \times m_\tau^*$ and $n_\tau^* = a3_t \times m_t + a4_\tau \times m_\tau^*$ ($a1$, $a2$, $a3$ and $a4$ are the corresponding coefficients of the mixing matrix). Since n_τ^* is linearly related to N_t, that means m_t is linearly related to m_τ^*. This is completely in violation of the ICA principles, that is, the ICA components are independent. Therefore, m_t must equal 0.

Two key assumptions are made in our model. One is the independence among the added EMG noise and the other was the correlation between the real noise and the added EMG noise.

C. Generating the simulated EMG

Since collecting sEMG from TBI patients can often times be difficult, we generated simulated EMG as a substitute. The following approach was used:

1) generating the extracellular current with Hodgkin-Huxley model
2) generating the single fiber action potential (SFAP) based on the volume conduction model
3) generating the motor unit action potential (MUAP)
4) generating the sEMG by defining the firing rate of MUAP
5) customizing the sEMG to mimic the characteristics of different muscles

Step 1: The Hodgkin-Huxley model was chosen since it is a widely accepted model for a typical excitatory cell, such as skeleton muscles.

Step 2: A volume conduction model was generated as in [10]:

$$\begin{aligned} V_E(z,y) =& K[\int_{S_1} \frac{\partial e(z)}{\partial z} \cdot \frac{1}{r} dS + \int_S dS \int_{-\infty}^{+\infty} \frac{\partial^2 e(z)}{\partial z^2} \cdot \frac{1}{r} dz \\ & - \int_{S_2} \frac{\partial e(z)}{\partial z} \cdot \frac{1}{r} dS] \end{aligned} \tag{1}$$

where V_E is the SFAP, $e(z)$ is the extracellular current, z and y are the axial and radial directions, respectively, S_1 and S_2 are the fiber sections at the fiber ends, and r is the distance between the surface element, and dS is the observation point.

The SFAP was generated by discretization of this formula and using known parameter values from the literature, including fiber length, endplate position, observation position, etc. [11], [12].

Step 3: Endplate positions were considered as a Gaussian distribution with 0 mean and standard deviation (SD) = 2.5 mm [11]. A Gaussian distribution was used for the voltage propagation velocities, with an average of 4 m/s and SD = 0.125. A total of 100 SFAPs were first generated and their average served as the MUAP.

Step 4: The firing rate of the MUAPs was modeled as the Poisson process [11]. The sEMG firing rate and amplitude were assumed to co-vary with the hand/finger movements. Hence, sEMG firing rate and amplitudes were increased during these events.

Step 5: Simulated EMG noise was generated for 4 different muscles, including the frontalis, temporalis, masseter, posterior head muscles, for each session. Since each head/facial muscle has different frequency domain characteristics, each muscle's simulated EMG was filtered based on the frequency characteristics found in the literature [12].

Before running ICA, these simulated EMG noises were added as extra virtual channels in the original EEG data. These virtual channels located at different positions on the edge of brain topographic map (details shown in Fig.1-3). The coordinates of these positions corresponded to the actual muscle locations on the head. Note that for the Subject 1, the real EMG (from subsection A) was added and processed in a similar manner to the simulated EMG data.

D. Rejection criteria

The following two criteria were established and an automatic rejection method was developed to reject the noise independent components after running ICA. First, the root mean square (RMS) values of coefficients in each mixing matrix rows corresponding to sEMG noise were calculated and used to establish a threshold. Subsequently, the coefficients which exceeded the threshold in the corresponding mixing matrix rows were identified and defined as high noise values. The components represented by the high noise

values' columns were considered as components related to artifacts and rejected. Second, components which primarily involved channels in the hatband EEG areas were also rejected. Note that this second criteria is used as the rejection criterion for conventional ICA in this paper.

E. Data processing

After adding the EMG noise in the EEG data, the combined data was subjected to a 3-200 Hz 3rd order band pass filter. Each trial, consisting of 1-s idle time followed by 2-s movement, was identified and segmented. The same ICA algorithm was employed in these three conditions (with real EMG noise, with simulated EMG noise and without any EMG noise). Here, the FastICA algorithm from the EEGLAB toolbox [13] was adopted on the EEG data from both subjects. The components related to artifacts were then rejected as above. The time-frequency decomposition underlying short-time Fourier transform was employed to analyze the time-frequency properties on the EEG before and after running ICA. For the data after running ICA, the time-frequency properties under three conditions (with simulated EMG, real EMG and without EMG) were calculated and compared in different frequency bands (α band: 8 to 12 Hz, high-γ band: 80 to 160 Hz), respectively. Both the data before and after the time frequency decomposition were separately normalized to the statistics of the EEG during the idling epochs using a Z-score.

III. RESULTS

Subject 1 completed the fist pumping task successfully, whereas Subject 2 only performed 19 out of the 20 expected pincer grasp trials.

In Subject 1, the α band desynchronization was localized to the left hemispheric areas around the C3 channel in the conditions with real EMG and simulated EMG Fig. 1). In the conventional ICA condition, only a modest α band desynchronization was seen. In Fig. 2, after running ICA, the power of the high-γ band decreased by 10.25% for conventional ICA, 60.08% for real EMG condition and 86.72% for simulated condition, respectively. Specifically, the power of the high-γ band during movement was lower in the real and simulated EMG conditions compared to the conventional ICA condition (reduced by 20.41% for conventional ICA, 65.6% for real EMG condition and 77.27% for simulated condition, respectively).

For Subject 2, the α desynchronization during movement was well-preserved after running ICA in both simulated EMG and conventional ICA conditions and it was localized to the hemicraniectomy areas around C3 channel (Fig. 3 A-C). In the high-γ band, the increased power during movement was observed primarily only in the hemicraniectomy areas just in the simulated EMG condition (Fig. 3 D-F). In non-hemicraniectomy area, the high-γ band decreased by 28.76% for conventional ICA and 65.28% for simulated EMG condition. Fig. 4B showed that, after running ICA, the high-γ power underlying movement in the simulated EMG condition within the hemicraniectomy area was significantly larger

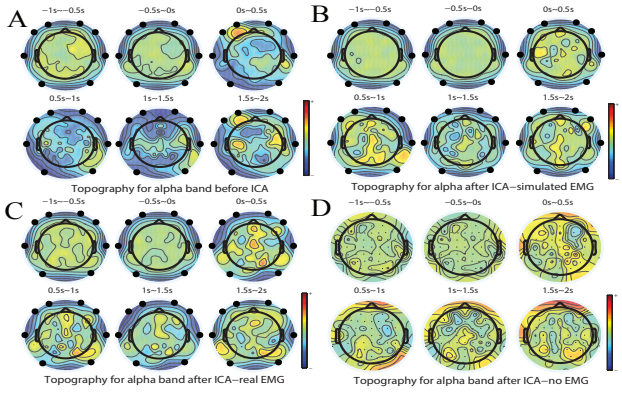

Fig. 1. Brain topography map for displaying the power of the α band (8 to 12 Hz) on the Subject 1. Time bin here was 0.5 second. 0s denoted the starting point of the movement. Negative denoted the time before movement (first two subfigures). The black dots outlined the position of the added virtual channels. The value for the color bar was from -1 to 1.

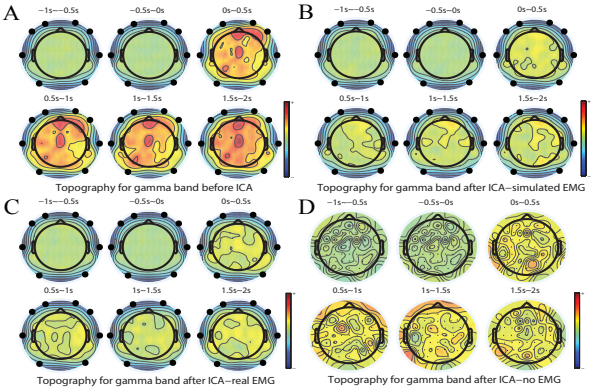

Fig. 2. Brain topography map for displaying the power of the high-γ band (80 to 160 Hz) on the Subject 1. Time bin here was 0.5 second. 0s denoted the starting point of the movement. Negative denoted the time before movement (first two subfigures). The black dots outlined the position of the added virtual channels. The value for the color bar was from -1 to 1.

than that in the conventional ICA condition (Wilcoxon rank sum test, P<0.001). Moreover, only in the simulated EMG condition was the high-γ power underlying movement in the hemicraniectomy significantly larger than that in the non-hemicraniectomy area (Wilcoxon rank sum test, P<0.001). The averaged time series of high-γ power in C3 (which was in the hemicraniectomy area) and amplitude of the squeezing force was shown in Fig. 4A. Fig. 4A showed that the high-γ power during movement in C3 channel was obviously increased after running ICA in the condition with simulated EMG while modest improvement with conventional ICA. However, there is a latency of approximately 250 ms between the onset of movement and the high-γ power increase event.

IV. DISCUSSION

We believe that the new technique removes EMG noise because:

- it successfully removed EMG from the EEG data from a healthy subject, where no high-γ signal (80 to 160 Hz) was expected to originate from the brain (Fig. 2).
- both α desynchronization and high-γ power increase during movement in the TBI patient were observed in anatomically expected areas, making this physiologically plausible (Fig. 3).

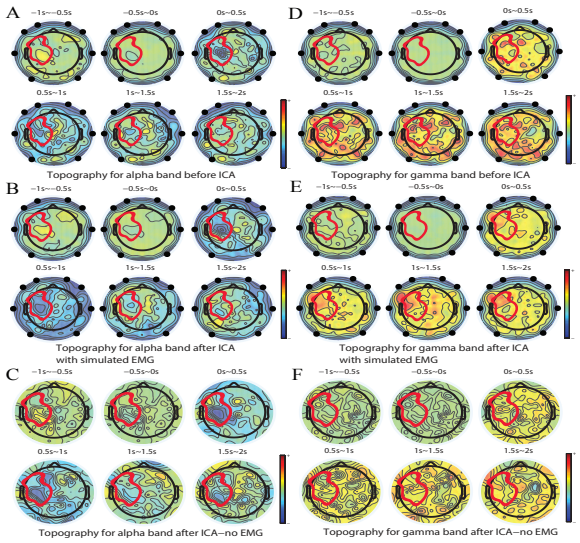

Fig. 3. Brain topography map for displaying the power of α band (8 to 12 Hz, A-C) and γ-band (80 to 160 Hz, D-F) in different conditions on the Subject 2. Time bin here was 0.5 second. 0s denoted the starting point of the movement. Negative denoted the time before movement (first two subfigures). The value for the color bar was from -1 to 1. The black dots outlined the position of the added virtual channels. The red outline in each subfigure defined the hemicraniectomy area.

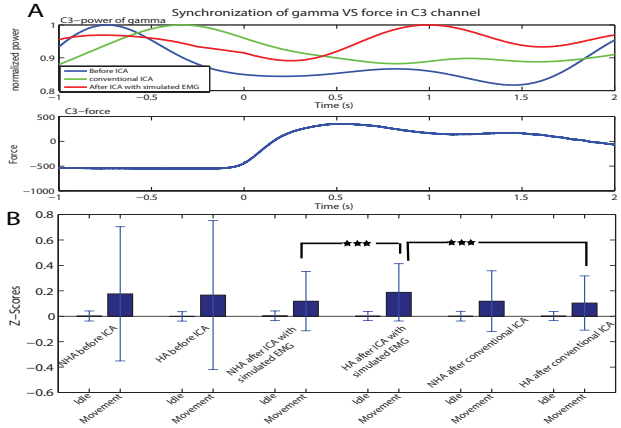

Fig. 4. A. Averaged power of the high-γ band (80 to 160 Hz) and amplitude of squeezing force in the TBI patient experiment. Values here were averaged over 19 trials. The power was normalized to the maximum of absolute value. Time point 0 meant the start of the movement. Negative denoted the time before moving. B. Power of high-γ band (80 to 160 Hz) under different conditions in the TBI patient experiment. Values here were from 19 trials. Wilcoxon rank sum test, asterisks indicate the significant differences between two datasets, and the significance level=***p<0.001. In the figure, NHA was the abbreviation for Non-hemicraniectomy area, HA was Hemicraniectomy area.

We also found the new technique does not remove high-γ signal from the brain - instead, it helped to improve the quality of the high-γ power underlying movement (Fig. 4).

Although the quality of the high-γ power after running ICA with simulated EMG condition has been improved, simulated sEMG noise still could not precisely mimic the real sEMG noise due to the time-varying of real sEMG. Furthermore, the increase of recovered high-γ power during movement lags the onset of movement, suggesting that this event may represent some form of sensory or feedback processing. Future work will involve rigorous testing, which includes verification with synthetic EEG and EMG data and with EEG data from a large group of patients, as well as developing a method that overcomes the time-varying nature of sEMG. Another future work is to perform direct comparisons with other popular EMG removal algorithms. In addition, the technique will need to be implemented in such a way that it can potentially be used in real time for neurorehabilitation applications, such as in BCI systems.

V. CONCLUSION

Here, we proposed a modified ICA model which involved adding sEMG noise into the EEG data to remove the sEMG artifacts automatically. According to the results from one healthy subject and one TBI patient, this approach can potentially remove the confounding overlap between EMG and gamma signals, and preserve the expected brain features underlying motor behavior. Therefore, it may allow researchers to confidently use the resulting high-γ signals for subsequent analysis or interpretation.

REFERENCES

[1] G. Pfurtscheller, B. Graimann, J. E. Huggins, S. P. Levine, and L. A. Schuh, "Spatiotemporal patterns of beta desynchronization and gamma synchronization in corticographic data during self-paced movement," *Clinical neurophysiology*, vol. 114, no. 7, pp. 1226–1236, 2003.

[2] N. E. Crone, D. L. Miglioretti, B. Gordon, and R. P. Lesser, "Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. ii. event-related synchronization in the gamma band." *Brain: a journal of neurology*, vol. 121, no. 12, pp. 2301–2315, 1998.

[3] P. T. Wang, C. M. McCrimmon, C. E. King, S. J. Shaw, D. E. Millett, H. Gong, L. A. Chui, C. Y. Liu, Z. Nenadic, and A. H. Do, "Characterization of electrocorticogram high-gamma signal in response to varying upper extremity movement velocity," *Brain Structure and Function*, vol. 222, no. 8, pp. 3705–3748, 2017.

[4] C. M. McCrimmon, P. T. Wang, P. Heydari, A. Nguyen, S. J. Shaw, H. Gong, L. A. Chui, C. Y. Liu, Z. Nenadic, and A. H. Do, "Electrocorticographic encoding of human gait in the leg primary motor cortex," *Cerebral Cortex*, pp. 1–11, 2017.

[5] T. Pistohl, A. Schulze-Bonhage, A. Aertsen, C. Mehring, and T. Ball, "Decoding natural grasp types from human ecog," *Neuroimage*, vol. 59, no. 1, pp. 248–260, 2012.

[6] B. Voytek, L. Secundo, A. Bidet-Caulet, D. Scabini, S. I. Stiver, A. D. Gean, G. T. Manley, and R. T. Knight, "Hemicraniectomy: a new model for human electrophysiology with high spatio-temporal resolution," *Journal of cognitive neuroscience*, vol. 22, no. 11, pp. 2491–2502, 2010.

[7] G. Barbati, C. Porcaro, F. Zappasodi, P. M. Rossini, and F. Tecchio, "Optimization of an independent component analysis approach for artifact identification and removal in magnetoencephalographic signals," *Clinical Neurophysiology*, vol. 115, no. 5, pp. 1220–1232, 2004.

[8] A. Delorme, S. Makeig, and T. Sejnowski, "Automatic artifact rejection for eeg data using high-order statistics and independent component analysis," in *Proceedings of the third international ICA conference*, 2001, pp. 9–12.

[9] A. Delorme, T. Sejnowski, and S. Makeig, "Enhanced detection of artifacts in eeg data using higher-order statistics and independent component analysis," *Neuroimage*, vol. 34, no. 4, pp. 1443–1449, 2007.

[10] J. Duchene and J.-Y. Hogrel, "A model of emg generation," *IEEE transactions on biomedical engineering*, vol. 47, no. 2, pp. 192–201, 2000.

[11] D. F. Stegeman and W. H. Linssen, "Muscle fiber action potential changes and surface emg: A simulation study," *Journal of Electromyography and Kinesiology*, vol. 2, no. 3, pp. 130–140, 1992.

[12] S. Muthukumaraswamy, "High-frequency brain activity and muscle artifacts in meg/eeg: a review and recommendations," *Frontiers in human neuroscience*, vol. 7, p. 138, 2013.

[13] 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, no. 1, pp. 9–21, 2004.