9th International IEEE EMBS Conference on Neural Engineering
San Francisco, CA, USA, March 20 - 23 , 2019

Removing TMS Artifacts from EEG Recordings Using a Deep Gated Recurrent Unit

Andac Demir[1], Mathew Yarossi[1,2], Damon Hyde[3], Mouhsin Shafi[4], Dana Brooks[1], Deniz Erdoğmuş[1]

Abstract— **The combination of transcranial magnetic stimulation (TMS) and electroencephalography (EEG) provides a direct means of assessing focal and distributed cortical behavior such as excitation/inhibition, intrinsic oscillatory activity and connectivity. However, TMS-EEG poses a number of technical challenges, foremost of which is removal of stimulation-induced artifacts that are several orders of magnitude larger than the neural signals of interest and typically obscure critical early neural responses to the stimulation. Here we describe a non-linear non-causal neural network predictor, built using a Gated Recurrent Unit architecture, and demonstrate its use to remove TMS artifacts from EEG recorded on a phantom, as well as real EEG synthetically contaminated by artifacts from the phantom experiment. Our results indicate that this artifact removal algorithm may decontaminate EEG signals as early as 6ms following stimulation. Given this result we discuss the future development of neural network based predictors for TMS artifact rejection.**

I. Introduction

Simultaneous TMS and EEG (TMS-EEG) provides a direct means of assessing focal and distributed cortical behavior such as excitation/inhibition, intrinsic oscillatory activity and connectivity [1], [2]. However, TMS-EEG poses a number of technical challenges, foremost of which is removal of stimulation–induced artifacts that are several orders of magnitude larger than the neural signals of interest [3], [4]. Many methods have been proposed to deal with these artifacts. One approach is to use a hardware based sample and hold filter to simply blank out the first interval (typically 20 ms) of EEG, preventing contamination of the remaining EEG [3]. While effective, this approach prevents the capture of early responses to the stimulation. It is well known from paired-pulse TMS paradigms that TMS can impact the underlying neural activity as early as 3-10 ms following the TMS pulse. Therefore in order to examine these early responses to TMS an alternative approach is required.

Equivalent circuit models offer an attractive approach to modeling TMS artifacts from first principles [5], [6]. The TMS artifact is predominantly induced by the magnetic flux from the TMS coil. However, a circuit model requires continuous and accurate measurement of a number of parameters of the TMS system (which may not be available for commercial hardware), its interface to the brain including electrode–skin impedance, and the brain itself, including magnetic flux from the TMS coil, contact impedance and the spatial parasitic capacitance. In addition the model is non-linear and switched (since the artifact outlasts the TMS pulse and the system changes upon pulse termination), requiring a complicated circuit. To deal with limited information about the physical characteristics of the system Katayama et. al. [5], [6] performed hand tuning of circuit parameters to fit experimentally collected artifacts. Limited availability of necessary measurements and the need for hand tuning of circuit parameters limit practicality of this approach for wide scale use. Another previously investigated alternative is to use an off-line Kalman filter approach to remove TMS-induced artifacts from EEG recordings. Morbidi et. al. identified two dynamic models describing EEG and TMS signal generation and applied Kalman filtering to the linear system arising from their combination [7]. Like the hardware model, this approach leverages the temporal dynamics of the TMS artifact. However, effectiveness of a linear approach to model a system that is known be highly non-linear is limited, and accurate and repeatable specific EEG and TMS dynamic models are difficult to attain and validate.

Currently, blind source separation approaches such as independent component analysis (ICA) are the predominant approach to TMS artifact rejection. ICA decomposes multi-channel EEG data into multiple components, assuming that artifact and neural signals possess distinguishable spatial-temporal patterns [8], [9]. Typically, temporal dynamics of the artifact are ignored and the problem is treated via spatial unmixing only, in contrast to the two aforementioned approaches. Using ICA, artifact rejection becomes a binary classification problem to distinguish between those components belonging to the artifact and those belonging to the underlying brain activity. Traditionally, classification has been done by hand, involving time-consuming, subjective, and potentially error-prone visual inspection [10], [1]. More recently a number of supervised and unsupervised classification techniques have been introduced to replace the dependence on human selection [11], [8]. Though automated identification of components resolves a major drawback, several issues remain with an ICA approach, foremost of which is the need to still discard the initial ˜10 ms of post stimulus data using a sample and hold filter.

Since the temporal dynamics of the artifact are both

[1]Andac Demir, Mathew Yarossi, Dana Brooks and Deniz Erdoğmuş are with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, USA {ademir, brooks, erdogmus}@ece.neu.edu

[2]Mathew Yarossi is with the Department of Physical Therapy, Movement and Rehabilitation Science, Northeastern University, Boston, MA 02115, USA m.yarossi@northeastern.edu

[4]Damon Hyde is with the Department of Radiology, Boston Children's Hospital, Boston, MA 02115 damon.hyde@childrens.harvard.edu

[3]Mouhsin Shafi is with the Beth Israel Deaconess Medical Center, Boston, MA 02115 mshafi@bidmc.harvard.edu

978-1-5386-7921-0/19/$31.00 ©2019 IEEE

strongly stereotypical and reflect the actual physical cause of the interference (as per the circuit model), here we describe a non-linear non-causal deep neural net dynamic predictor to address the artifact removal problem.

II. Methodology

TMS-EEG was measured on both a phantom and a human volunteer. A melon (Cucumis melo) was used for the phantom as melon skin has been reported to have similar dielectric properties to human skin [12]. The human subject was neurologically healthy and had no family history of epilepsy. Informed consent was obtained in accordance with the Institutional Review Board of Beth Israel Deaconess Medical Center where the experiments were performed.

A. Data Collection and Preprocessing

Data collection procedures were identical for the phantom and human subject. The phantom head (melon) was positioned on a table at a similar distance from the simulators and recording equipment as in human experiment. Blocks of 30 single monophasic pulses of TMS were administered with 5-second inter-pulse intervals beginning at 10% maximum stimulator output (MSO), with successive blocks increasing in intensity at 10% MSO increments to a maximum of 80% MSO. TMS was administered with a Magpro X100 TMS stimulator (Magventure Inc.) using a Cool-B65 figure-of-eight coil with a 75mm outer diameter and external liquid cooling. The stimulator was positioned over the C3 electrode and oriented approximately 45° off midline with the handle facing posteriorly to induce a posterior to anterior current flow. EEG was recorded at 5kHz using a TMS-compatible EEG system (actiCHamp: BrainProducts Inc.) connected to a 64-channel cap with active Ag/AgCl electrodes with impedance conversion at the electrode level (actiCAP slim: BrainProducts Inc.)

a) Preprocessing: Processing for human subject and phantom data were identical. 2000 ms intervals were segmented centered at the TMS pulse. Each channel was baseline corrected with respect to the TMS-free data (1000 to 500 ms prior to the TMS pulse). No additional digital filtering was applied so as to not distort the high frequency artifacts. Since recurrent neural networks (RNNs) are sensitive to the scale of the input, data was scaled linearly to a 0-1 range for every intensity level and channel separately before being fed into the network. To obtain a dataset with known EEG and artifacts components for ground truth evaluation, synthetic data were created by adding non-artifact contaminated EEG (obtained from the rest period prior to TMS stimulation) from the human subject to the melon recordings of the stimulation artifact.

B. Training with Recurrent Cell Architectures

The core of the model implemented in this paper consists of a two-layer deep Gated Recurrent Unit (GRU) neural network. GRU's are an advanced version of RNN's with two gates that control the flow of information: an update gate and a reset gate. The update gate determines how much of the information from previous time steps needs to be passed along to the future, while the reset gate determines how much of that information needs to be forgotten [13].

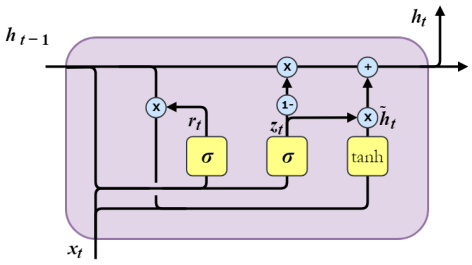

Fig. 1: Gated recurrent unit cell architecture

When the previous hidden state h_{t-1} and an input sample x_t are plugged into the GRU cell, they are multiplied by their own weights W_{hz} and W_{xz} as shown in Fig. 1. We start by calculating the update gate z_t for time step t using the equation:

$$z_t = \sigma\left(W_{hz} \cdot h_{t-1} + W_{xz} \cdot x_t\right) \tag{1}$$

The equation for the reset gate is similar to the one for the update gate:

$$r_t = \sigma\left(W_{hr} \cdot h_{t-1} + W_{xr} \cdot x_t\right) \tag{2}$$

Then we introduce a candidate hidden state which will use the reset gate to store the relevant information from the past:

$$\widetilde{h}_t = \tanh\left(W_{chr} \cdot \left(r_t \cdot h_{t-1}\right) + W_{chx} \cdot x_t\right) \tag{3}$$

The activation h_t of the GRU cell at time t is the linear sum of the previous activation h_{t-1} and the candidate activation $\widetilde{h}_t$:

$$h_t = \left(1 - z_t\right) \cdot h_{t-1} + z_t \cdot \widetilde{h}_t \tag{4}$$

The RNN implemented for this application was a stacked GRU which has 2 hidden layers; each layer is composed of 64 GRU cells. The hidden states of the second layer are fully connected to a dense layer of 64 hidden units that predicts the next sample in the sequence. We segmented the EEG signals from 2 ms before to 20 ms after the TMS. We initialized the network with the first 3 samples in the sequence, predict the next sample, and then slide by one sample in the sequence to predict the next. When phrased as a regression problem, the input variables are x_{t-2}, x_{t-1}, x_t and the output variable is x_{t+1}. The intuition is that the TMS artifact data collected from the phantom head comes from a physical system that can be characterized by an RNN and thus such a model can be learned and applied to human TMS-EEG recordings to estimate, and then remove the TMS artifacts.

Training of the network was performed with the PyTorch framework for 200 epochs using a batch size of 1. In order to avoid overfitting, the model was validated with 5-fold cross validation. For optimization we experimented with Adam, SGD and a Quasi-Newton method: L-BFGS, an unconstrained, non-linear optimization that steers its search

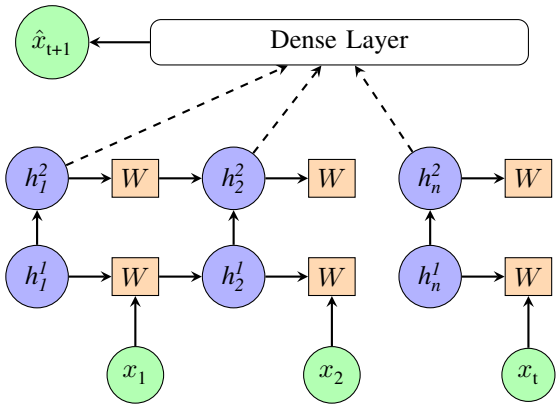

Fig. 2: Graphical model of a deep GRU network. h_i^l denotes the i^{th} hidden state at the l^{th} layer. W denotes the weight matrix that updates the hidden states. $\hat{x}_{t+1}$ denotes the prediction of the next sample in the sequence, computed by matrix multiplication of the weights of the dense layer by the hidden states in the second layer. The network can be fed with the first t samples of the TMS-EEG signal as long as $t \leq n$, the number of hidden states in the first layer.

through variable space by using an estimate to the inverse Hessian matrix [14], [15], evaluating the mean squared error between the actual and the predicted artifacts as the objective function. The lowest training error was obtained from L-BFGS. Although L-BFGS is memory greedy and computationally intensive, since it is a second-order method, it found the local minima better than the variants of stochastic optimization methods, resulting in lower error in predicting the TMS artifacts.

As Fig. 2 shows, the weights, W, of the GRU network are shared across time. This enables the network to be trained with fewer parameters and allows passing input sequences of variable length into the pre-trained network during inference. Network hyper-parameters were determined by performing grid search. Specifically, we minimized the expected L2 loss in the search space trying different number of samples passed into the network: $t \in \{3, 5, 10, 15, 25\}$, different number of hidden units in each layer: $n \in \{16, 32, 64, 128, 256\}$, layers $l \in \{1, 2, 3\}$ and size of the dense layer $n_{dense} \in \{16, 32, 64, 128, 256\}$. Designing a neural architecture that achieved the lowest loss over a reasonable parameter space when tested on independent validation data was the foremost priority. We trained our model on the TMS signals collected from the phantom head per each channel separately for all intensity levels. When tested on human data or synthetic data, the artifact prediction was subtracted from TMS-EEG to estimate the clean EEG signals.

III. Results

Visual inspection of artifact predictions indicate they are well matched to the amplitude, frequency, and phase of the underdamped harmonic oscillation shape characteristic of recorded TMS artifacts (Fig 3). There was no notable decrease in the quality of the artifact fit across channels

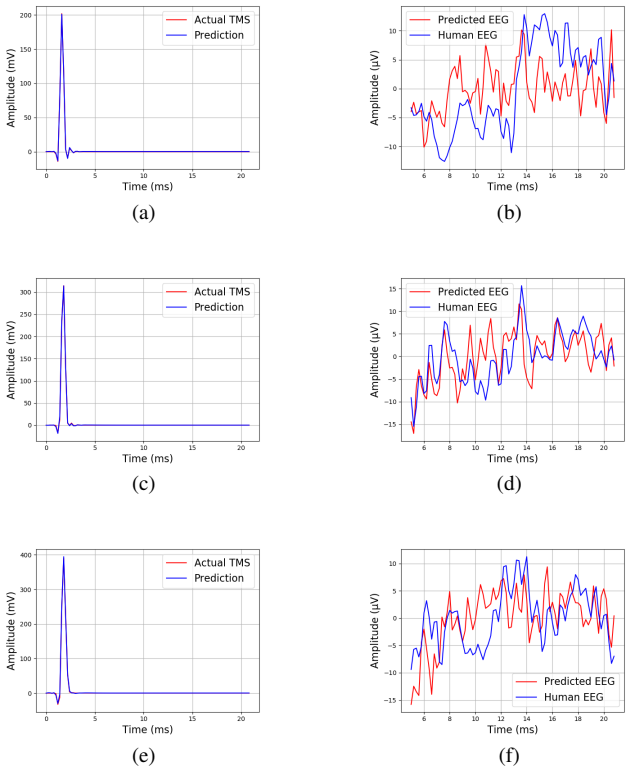

Fig. 3: $(a\text{-}f)$ Actual TMS artifacts and artifacts predicted by the recurrent neural network in the left column. EEG signals and EEG estimated by the model from the synthetic signals in the right column $(a\text{-}b)$ Channel: C3, MSO: 20%. $(c\text{-}d)$ Channel: C3, MSO: 40%. $(e\text{-}f)$ Channel: C3, MSO: 60%.

that were nearer or farther from the site of stimulation. The residuals after subtracting the artifact predictions from the synthetic TMS-EEG data set are the model estimates of clean EEG, as shown in the right column of the figure. The recovered EEG was qualitatively similar to the uncontaminated data. The model was not able to accurately recover the first 6 ms of data. As expected, we observed improved prediction of the underlying EEG as time progressed after stimulus onset. Meaningful signal recovery as early as 6 ms following stimulation is an improvement over currently available techniques that typical require blanking or removal of the first 10-14 ms of data [16].

Fig. 4 reports on performance of the method as a function of spatial position, post-stimulus time, and TMS amplitude. We show results from six channels over four time bins at three MSO levels, for a total of 72 comparisons. Results are quantified by signal-to-interference ratio, calculated as the EEG energy divided by residual artifact energy averaged over trials and time bins.

IV. Discussion

The challenge of TMS artifact removal from EEG recordings persists despite more than a decade of investigation. Results of our proposed solution using a deep GRU network architecture indicate this is a promising (though imperfect) approach. In particular our results were relatively insensi-

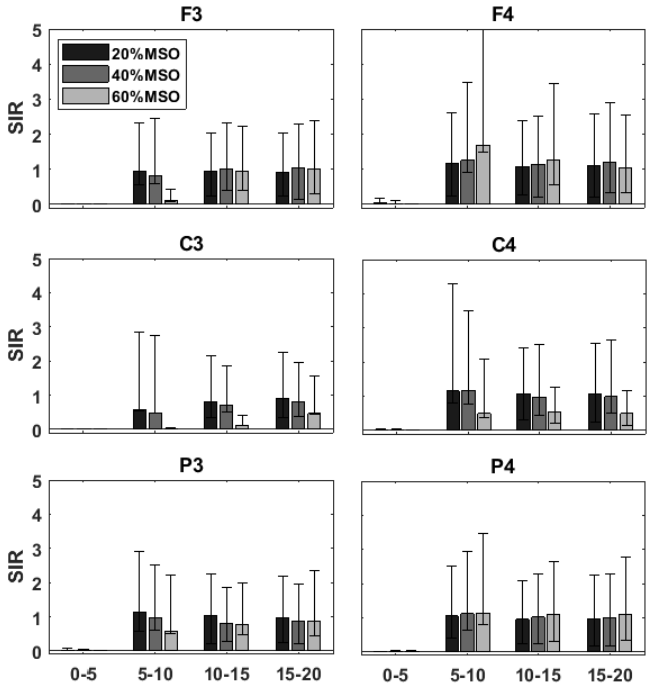

Fig. 4: Mean signal-to-interference ratio (SIR) for estimated EEG calculated as a function of spatial position (six channels), time post-stimulus (four $5ms$ time bins) and TMS amplitude (three levels). Note that the TMS coil was positioned above C3. Bars for each time interval show SIR calculated as EEG energy divided by residual artifact energy. Color shows TMS amplitude: 20% (black), 40% (dark gray), and 60% (light gray). Bar height is mean over trials, error bars show 10 to 90 percentile intervals.

tive to channel location, a key point as one might expect the method to perform worse for larger artifacts. We also achieved significant artifact reduction earlier post-stimulus than most existing methods. Importantly, the method did not require manual intervention to select putative signal components, and was self-calibrating, with no need to learn an equivalent circuit model that might be tuned to a particular TMS hardware system, electrode configuration, or even highly variable electrode-skin contact impedance. We still need to test the system on TMS-EEG from human subjects, where the artifacts are expected to last longer due to increased contact impedance.

TMS artifact still dominated the EEG signal for the first 6-10 ms following stimulation. Future research will seek to improve upon this result by testing modifications in the network architecture and input data. In particular, we speculate that the sigmoid nonlinearities prevent the dynamic model from achieving the rapid rise and fall times needed to approximate the interference; we will explore other nonlinearities that capture locally unstable behavior to follow large spikes that terminate rapidly. Training across channels, requiring a novel spatio-temporal GRU predictor, might capture spatial regularities in the artifact signal. Training across intensities to understand whether artifacts recorded at low intensity (when the cortical response is theoretically negligible or small) can be scaled to approximate artifacts found at greater intensity may prove useful for accommodating artifact variability that occurs due to differences in EEG equipment and human head geometry across individuals.

ACKNOWLEDGMENT

This project was supported by NIH grants P41 GM103545-18 (DB), R01 DC009834(DE), and NSF grants IIS-1149570 (DE), CNS-1544895 (DE), CBET-1804550 (DE, DB).

REFERENCES

[1] Nigel C Rogasch, Caley Sullivan, Richard H Thomson, Nathan S Rose, Neil W Bailey, Paul B Fitzgerald, Faranak Farzan, and Julio C Hernandez-Pavon, "Analysing concurrent transcranial magnetic stimulation and electroencephalographic data: a review and introduction to the open-source tesa software," *Neuroimage*, vol. 147, pp. 934–951, 2017.

[2] Zafiris J Daskalakis, Faranak Farzan, Natasha Radhu, and Paul B Fitzgerald, "Combined transcranial magnetic stimulation and electroencephalography: its past, present and future," *Brain research*, vol. 1463, pp. 93–107, 2012.

[3] Risto J Ilmoniemi and Dubravko Kičić, "Methodology for combined tms and eeg," *Brain topography*, vol. 22, no. 4, pp. 233, 2010.

[4] Risto J Ilmoniemi, Julio C Hernandez-Pavon, Niko N Mäkelä, Johanna Metsomaa, Tuomas P Mutanen, Matti Stenroos, and Jukka Sarvas, "Dealing with artifacts in tms-evoked eeg," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*. IEEE, 2015, pp. 230–233.

[5] Yoshinori Katayama and Keiji Iramina, "Equivalent circuit simulation of the induced artifacts resulted from transcranial magnetic stimulation on human electroencephalography," *IEEE Transactions on Magnetics*, vol. 45, no. 10, pp. 4833–4836, 2009.

[6] Y Katayama and K Iramina, "Fitting and eliminating to the tms induced artifact on the measured eeg by the equivalent circuit simulation improved performance," in *5th Kuala Lumpur International Conference on Biomedical Engineering 2011*. Springer, 2011, pp. 519–522.

[7] Fabio Morbidi, Andrea Garulli, Domenico Prattichizzo, Cristiano Rizzo, and Simone Rossi, "Application of kalman filter to remove tms-induced artifacts from eeg recordings," *IEEE Transactions on Control Systems Technology*, vol. 16, no. 6, pp. 1360–1366, 2008.

[8] Wei Wu, Corey J Keller, Nigel C Rogasch, Parker Longwell, Emmanuel Shpigel, Camarin E Rolle, and Amit Etkin, "Artist: A fully automated artifact rejection algorithm for single-pulse tms-eeg data," *Human brain mapping*, 2018.

[9] Arnaud Delorme, Terrence Sejnowski, and Scott 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] Sravya Atluri, Matthew Frehlich, Ye Mei, Luis Garcia Dominguez, Nigel C Rogasch, Willy Wong, Zafiris J Daskalakis, and Faranak Farzan, "Tmseeg: A matlab-based graphical user interface for processing electrophysiological signals during transcranial magnetic stimulation," *Frontiers in neural circuits*, vol. 10, pp. 78, 2016.

[11] Thea Radüntz, Jon Scouten, Olaf Hochmuth, and Beate Meffert, "Automated eeg artifact elimination by applying machine learning algorithms to ica-based features," *Journal of neural engineering*, vol. 14, no. 4, pp. 046004, 2017.

[12] Nigel C Rogasch, Richard H Thomson, Zafiris J Daskalakis, and Paul B Fitzgerald, "Short-latency artifacts associated with concurrent tms–eeg," *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation*, vol. 6, no. 6, pp. 868–876, 2013.

[13] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

[14] Robert Malouf, "A comparison of algorithms for maximum entropy parameter estimation," in *proceedings of the 6th conference on Natural language learning-Volume 20*. Association for Computational Linguistics, 2002, pp. 1–7.

[15] Galen Andrew and Jianfeng Gao, "Scalable training of l 1-regularized log-linear models," in *Proceedings of the 24th international conference on Machine learning*. ACM, 2007, pp. 33–40.

[16] Wei Wu, Corey J Keller, Nigel C Rogasch, Parker Longwell, Emmanuel Shpigel, Camarin E Rolle, and Amit Etkin, "Artist: A fully automated artifact rejection algorithm for single-pulse tms-eeg data," *Human brain mapping*, vol. 39, no. 4, pp. 1607–1625, 2018.