Real-Time Tracking of Magnetoencephalographic Neuromarkers during a Dynamic Attention-Switching Task

[1]Alessandro Presacco, [2]Sina Miran, [1,2]Behtash Babadi, [1,2,3]Jonathan Z Simon
[1]Institute for Systems Research, [2]Department of Electrical and Computer Engineering [3]Department of Biology
University of Maryland, College Park, MD, USA
Email: {$apresacc$, $smiran$, $behtash$, $jzsimon$}@umd.edu

***Abstract*— In the last few years, a large number of experiments have been focused on exploring the possibility of using non-invasive techniques, such as electroencephalography (EEG) and magnetoencephalography (MEG), to identify auditory-related neuromarkers which are modulated by attention. Results from several studies where participants listen to a story narrated by one speaker, while trying to ignore a different story narrated by a competing speaker, suggest the feasibility of extracting neuromarkers that demonstrate enhanced phase locking to the attended speech stream. These promising findings have the potential to be used in clinical applications, such as EEG-driven hearing aids. One major challenge in achieving this goal is the need to devise an algorithm capable of tracking these neuromarkers in real-time when individuals are given the freedom to repeatedly switch attention among speakers at will. Here we present an algorithm pipeline that is designed to efficiently recognize changes of neural speech tracking during a dynamic-attention switching task and to use them as an input for a near real-time state-space model that translates these neuromarkers into attentional state estimates with a minimal delay. This algorithm pipeline was tested with MEG data collected from participants who had the freedom to change the focus of their attention between two speakers at will. Results suggest the feasibility of using our algorithm pipeline to track changes of attention in near-real time in a dynamic auditory scene.**

I. Introduction

One of the most remarkable features of the brain is its ability to use attention to select a speaker in a cocktail party scenario, that is in a multi-speaker environment. Great effort has been spent in the past decade to understand the mechanisms underlying the ability of the brain to segregate multiple sound sources and to direct its attention to the intended speaker. Several studies where participants were asked to attend to one, while trying to ignore a competing speaker, have suggested that low-frequency oscillations of the brain are more phase-locked to the speech envelope of the attended stimulus [1]–[3]. A number of studies using non-invasive electrophysiological measurements (EEG and MEG) have attempted to devise "attentional decoders", whose goal is to decode the focus of attention when more than one speaker is present in the mix [4]–[7].

A major challenge in using M/EEG neuromarkers to identify the listener's attentional focus is the poor accuracy of attention decoding algorithms in near real-time settings. Current and past attempts to use M/EEG neuromarkers to determine a listener's attentional focus often use tens of seconds before making decisions. This long delay prevents the rapid decisions required in realistic auditory scenes where switching from one auditory source to another is commonplace, or where the audio signal might even be altered in response to changes in the neuromarkers themselves to increase signal saliency and clarity. Overcoming this limitation would be critical in order to adopt these algorithms for clinical applications. Recent efforts have focused on the feasibility of using neuromarkers extracted from EEG signals as feedback to optimize hearing-aid parameters [8], [9].

A state-space model based on Bayesian filtering has been recently proposed by our group as a solution to the near real-time estimation of attentional state [10], [11]. Notably, this model has an aggregate decision delay of only a few seconds, making it a potential candidate to be embedded in future generation hearing-aids processors. Even though results were promising, subjects were not tested in the case of a dynamic task where they were allowed to switch attention on their own volition, which mimics the more realistic auditory user experience. Here we present an experiment where participants had the freedom to switch attention at will. To the best of our knowledge, this is the first attempt to apply attention decoding algorithms to a setting similar to this real-life usage. We also propose the addition of a three state Hidden Markov Model (HMM) in an attempt to better track attentional-related changes in neuromarkers.

Results from this study show the feasibility of our algorithm pipeline to track changes in auditory attention in near real-time. Additionally, the new HMM component enhanced our ability to track attentional-related changes in neuromarkers, thus improving the decision making accuracy of our algorithm.

II. Methods

The experimental protocol and all procedures were reviewed and approved by the Institutional Review Board of the University of Maryland. Participants gave written,

This project was supported by a National Science Foundation grant, NSF SMA1734892, and National Institutes of Health grants, NIH R01-DC14085, NIH P01-AG055365, and DARPA grant N6600118240224. The views, opinions and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government. We would like to thank Anna Namyst for excellent technical support and Joshua Pranjeevan Kulasingham for his help in creating the stimuli and in collecting the data.

978-1-5386-1311-5/19/$31.00 ©2019 IEEE

informed consent, according to principles set forth by the University of Maryland's Institutional Review Board.

Participants. Participants comprised 5 younger adults (22-33 yr) recruited from the University of Maryland. All participants were native English speakers, had no history of neurological disorders and had clinically normal hearing.

Stimuli and Recording. Participants were asked to attend to one of two stories presented diotically while ignoring the other one. The stimuli consisted of two segments from the book, The Legend of Sleepy Hollow by Washington Irving. One of the segments was narrated by a male speaker, while the other one by a female speaker. The speech mixture was presented at ~70 dB sound pressure level and was constructed by digitally mixing two speech segments into a single channel, with a duration of 90 seconds and at a signal-to-noise ratio of 0 dB, as defined by calculating the logarithm of the root-mean-square value. Participants listened to 3 trials of the same speech mixture and were instructed to start attending to the male speaker first and then to switch the focus of their attention at their own will for a minimum of 1 time and a maximum of 3 times. Participants were also given a switching button that they were instructed to press every time they decided to switch attention. Neuromagnetic signals were recorded at a sampling frequency of 2000 Hz using a 157-sensor whole-head MEG system (Kanazawa Institute of Technology, Nonoichi Ishikawa, Japan) in a magnetically shielded room.

Data Analysis. MEG data were analyzed offline using MATLAB. Three reference channels were used to measure and cancel the environmental magnetic field by using time shift-principal component analysis [12]. The 157 raw MEG data channel responses were first filtered between 2 and 8 Hz, then decomposed using *n* spatial filters into *n* signal components (where $n \leq 157$) using the denoising source separation (DSS) algorithm [13], [14]. The first DSS component filter was then used for the analysis. The signal components used for analysis were then re-extracted from the raw data for each trial, spatially filtered using the first DSS filter just constructed and then band-pass filtered between 1 and 8 Hz. A total of 3 time series, one per trial, were obtained and used for the final analysis. The reconstructed envelope was obtained from the unmixed speech of the single speakers used for the task, not from the acoustic stimulus mixture. The envelope was computed as the 1- to 8-Hz band pass-filtered magnitude of the analytic signal. Both speech envelope and neural data were then downsampled to 200 Hz.

Estimation of the Temporal Response Function (TRF). Neuromarkers were extracted from the attention-modulated coefficients of the encoder estimated for the envelope of each speaker. In the context of the encoding model, these coefficients are referred to as the Temporal Response Function (TRF) [1]. The TRF carries two important characteristics: it has a high degree of sparsity and is modulated by attention. TRFs were estimated by using the FASTA software package [15] available online at [16], which allowed us to calculate the coefficients in consecutive non-overlapping 500 ms windows. For more details on the TRF estimation details please refer to [10]. At each time window, the absolute value of the peaks between 75 and 250 ms was extracted, then smoothed using a Savitzky-Golay FIR smoothing filter and the maximum value was denoted as the magnitude of the M100 peak. The M100 peak in TRF has been extensively studied as an attention-modulated neuromarker [1], [5].

Hidden Markov Model (HMM). An HMM was used to estimate the internal state of the dynamics of the M100 peak based on its first derivative. Let $A(t)$ denote the amplitude of the M100 peak at time t. The HMM consisted of the following three states: State 1 indicated no significant change in $\frac{dA(t)}{dt}$, State 2 indicated a significant increase in $\frac{dA(t)}{dt}$, while State 3 indicated a significant decrease in $\frac{dA(t)}{dt}$. The Viterbi algorithm was used to estimate the most likely path of the hidden states. A threshold of $Th = 3 \times 10^{-5}$ was used to classify the derivative in stable (S) $-Th < \frac{dA(t)}{dt} < Th$, positive (P) $\frac{dA(t)}{dt} > Th$ and negative (N) $\frac{dA(t)}{dt} < -Th$ states. The M100 peak time-series was smoothed using weighted linear least squares in order to facilitate the estimation of $\frac{dA(t)}{dt}$. The HMM transition probabilities were chosen as:

$$P_{ij} = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.15 & 0.8 & 0.05 \\ 0.15 & 0.05 & 0.8 \end{bmatrix},$$

and the following observation likelihoods were used:

$$P(S \mid 1) = 0.7, P(S \mid 2) = 0.15, P(S \mid 3) = 0.15$$
$$P(P \mid 1) = 0.25, P(P \mid 2) = 0.7, P(P \mid 3) = 0.05$$
$$P(N \mid 1) = 0.25, P(N \mid 2) = 0.05, P(N \mid 3) = 0.7$$

The initial probabilities were the following: $\Pi_1 = 0.6, \Pi_2 = 0.2, \Pi_3 = 0.2$.

An adjustment variable *adj* was initialized to 0 and updated each time the state of the peak was determined. Specifically, for state 1 no changes were made to *adj*, for state 2 the variable was incremented by 1.3% of the peak amplitude, while for state 3 the variable was decremented by 1.3% of the peak amplitude. The variable was then added to the original value of the maximum peak to create the final neuromarker. By doing so, we "boosted" the peak when in state 2 and "penalized" it when in state 3. The specific value of the percentage increase/decrease of the peak was critical to allow smooth changes in the neuromarker. This procedure was adopted to incorporate the presence of a significant rise or fall in magnitude of the M100 peak, indicating a likely switch of attention.

Bayesian Filtering. The neuromarkers extracted were then fed into a near real-time state-space estimator that translated them to robust and statistically interpretable estimates of the attentional state with a minimal delay [10], [11]. The forward lag was set at ~1.5 seconds, with the backward lag at ~13.5 seconds. The forgetting factor was set at 0.95, while the regularization parameter was set at 0.001. More details of this algorithm can be found in [10], [11].

III. Results

Representative subject: Here we report the results from representative subject R2082. Figure 1 shows the results

of the TRF estimate (Panel A) and the real-time state-space estimates of the attentional state (Panel B) during the second trial. Vertical lines in panel B represent the time when the subject pressed the switching button to signal her attentional switch. As expected, the TRF of both male (top) and female (bottom) speakers exhibit some degree of sparsity, as shown by two major peaks appearing at ~50 ms and ~100 ms throughout the duration of the task, consistent with [1]. M100 magnitude displayed in Panel B (top) show a strong modulation consistent with the switch of attention of the participant. Specifically, during the first ~18 seconds the M100 peak corresponding to the male speaker (black waveform) is larger than the one corresponding to the female speaker (red waveform). Then as the subject starts switching her attention towards the female speaker, as indicated by the vertical line at ~22 seconds, the trend is reversed and the attend-female neuromarker becomes stronger than the attend-male one. The subject then switches attention two more times (~47 and ~66 seconds) during the task and in both cases a change in the strength of the neuromarkers is observed. Attentional state is correctly recognized by the state-space model, that shows a change in probability following the trend of the attend-male and attend-female neuromarkers.

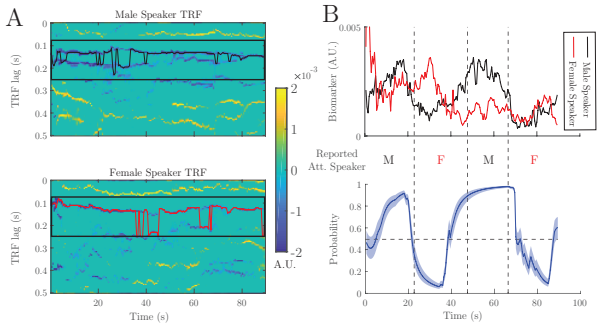

Fig. 1. Panel A: TRF estimated with FASTA of the male (top) and of the female (bottom) speakers.The black rectangular boxes show the time window (75 to 250 ms) used to estimate the M100 peak at each point in time. The narrow black and red line are used to track the peak position used for the analysis for male and female TRF, respectively. The scale of the color bar has been adjusted to better visualize the peaks. Panel B: M100 peak magnitudes (i.e. neuromarkers) (top) extracted at each point in time for male (black) and female (red) speakers and near real-time state-space estimates of the attentional state (bottom). The state-space model output displays the estimated probability of attending to the male speaker. The blacked dashed line indicates the threshold for attentional switch (> 0.5 attending male), while the black vertical lines show the time when the subject pressed the switching button to indicate change of attention. Colored hulls indicate 90% confidence intervals of the state-space estimates. The black and red letters M and F indicate the speaker that the subject reported attending to.

Results from all the 3 trials for subject R2082 are shown in Figure 2. Overall, the neuromarkers reliably follow attention switch of the subject in all the 3 trials. The real-time state-space algorithm also transforms the extracted neuromarkers into a robust and interpretable measure of the attentional state for all the 3 trials.

The HMM performance: Figure 3 shows the results of the estimates of attentional state without (middle row) and with (bottom row) the inclusion of the HMM in our algorithm pipeline. Boosting or penalizing the peaks of the

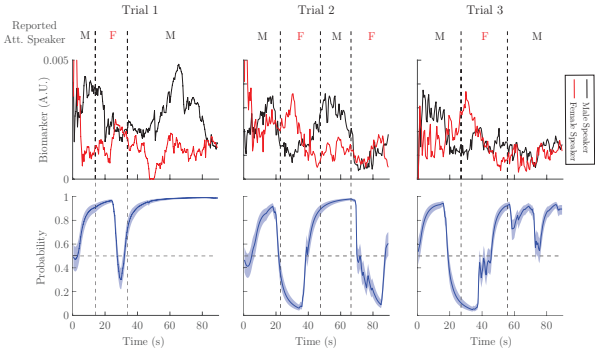

Fig. 2. Results from the 3 trials recorded from subject R2082. The top row shows the M100 peak magnitudes (i.e. neuromarkers) extracted at each point in time for male (black) and female (red) speakers and near real-time state-space estimates of the attentional state (bottom). The blacked dashed line indicates the threshold for attentional switch (> 0.5 attending male), while the black vertical lines show the time when the subject pressed the switching button to indicate change of attention. Colored hulls indicate 90% confidence intervals of the state-space estimates. The black and red letters M and F indicate the speaker that the subject reported attending to. Overall, neuromarkers reliably follow the speaker the participant is trying to focus on.

TRF, depending on their predicted state, resulted in a more reliable estimation of the neuromarkers, as it is particularly evident in the last ~23 seconds of trial 2. At ~67 seconds the subject indicated the intention to switch attention, which is corroborated by a steep decline in the strength of the male neuromarker, as shown in the top panel. However, without penalizing the male neuromarker for this steep decline, the change in attentional state is not detected until ~84 seconds, because of the failure of the female neuromarker to rapidly increase above the level of the male neuromarker.

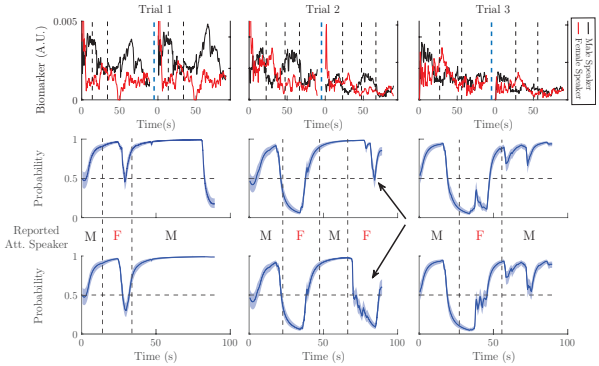

Fig. 3. Results of the real-time state-space estimates of the attentional state with and without HMM. The top row shows the neuromarkers without (left panel) and with (right panel) HMM. The middle and bottom row shows the real-time estimates of the attentional state without (middle row) and with (bottom row) the addition of the HMM to the algorithm pipeline. The blacked dashed line indicates the threshold for attentional switch ($>$ 0.5 attending male), while the black vertical lines show the time when the subject pressed the switching button to indicate attention switch. Colored hulls indicate 90% confidence intervals of the state-space estimates. The black and red letters M and F indicate the speaker that the subject reported attending to. Boosting and penalizing the neuromarkers based on the current state of their slope results in a more reliable estimate of the attentional state.

The penalization of the attend-male neuromarker allows the attend-female neuromarker to raise above the attend-male

neuromarker, resulting in the correct detection of attention switch. The penalization and boosting of the peaks is small enough to preserve the smoothness of the neuromarkers, thus maintaining the general trend of the dynamics.

IV. Discussion

In this paper, we have shown the feasibility of using a near real-time algorithm pipeline to track the attention state in a dual-speaker setting during a dynamic-attention switching task, thus taking an additional and important step towards the possibility of embedding attention decoding algorithms in future generation of hearing-aids processors. Importantly, the task devised for this experiment reproduced very closely a real-life scenario, where individuals switch their attention frequently at their own will. Previous experiments have constrained the subjects to either focus their attention to the same speaker for the whole duration of the task [1], [2] or to switch their attention half way through the task only once [5], [10]. Multiple and free attention switches will also give us the opportunity to investigate the dynamics of the relevant neural processes at the critical moment when individuals willingly shift the focus of their attention.

The addition of a derivative-based three state HMM to our algorithm pipeline also proved to be beneficial in tracking the oscillatory patterns of the neuromarkers. Boosting or penalizing the amplitude of the peaks of the TRF based on their trend could be critical to facilitate detection of attention switches in situations where intrinsic characteristics of the speech envelope (e.g. pitch) may be responsible for weaker or stronger representation of the TRF peaks. This scenario was particularly relevant in the last ~23 seconds of trial 2 (Fig. 2), where our participant indicated her intention to switch attention from the male to the female speaker. The male neuromarker exhibited a steep decrease in amplitude, which was not paired by a rapid and sufficient increase in amplitude of the female neuromarker, resulting in failure to correctly detect a switch of attention. The addition of the HMM model allowed us to capture such oscillatory trends in the male and female neuromarkers. This boosting/penalization approach could be beneficial in real-time applications, as it would allow the algorithm to "speed up" the recognition of changes in the oscillatory patterns initiated by switching of attention.

Two interesting observations can be extrapolated from our results. The first one is the delay between the time that the attentional neuromarkers started changing their pattern and the time that the participant pressed the switching button. We speculate that this trend could be simply explained by a transition time necessary for the subject to switch her attention. It is possible that the focus of her attention started to naturally and gradually shift towards the competing talker, seconds before she decided to press the switching button. Therefore, pressing the switching button may have signaled the exact time where the subject was able to phase-lock to the other speaker rather than the initial stage of her attempt at switching attention. The second observation is related to situations where attention may be equally split or switch back and forth rapidly between the two speakers. This situation is particularly evident in the third trial Fig. 2, in the ~38 seconds to ~45 seconds interval, where the output of the state-space estimator fluctuated around 0.5. Interestingly, this fluctuation precedes the pressing of the switching button, thus suggesting that the subject may have found herself in a situation where her attention was splitting or switching back and forth rapidly between the two speakers. This is indeed a very familiar scenario that we face every day and that will need to be addressed in our future studies.

In conclusion, this paper reports preliminary results suggesting the feasibility of tracking attention in a scenario that closely represents our daily listening experience, thus taking an additional and important step towards the feasibility of using attention decoding algorithms for practical applications.

References

[1] N. Ding and J. Z. Simon, "Emergence of neural encoding of auditory objects while listening to competing speakers," *Proc Natl Acad Sci USA*, vol. 109, no. 29, pp. 11 854–9, 2012.

[2] N. Ding and J. Z. Simon, "Adaptive temporal encoding leads to a background-insensitive cortical representation of speech," *J Neurosci*, vol. 33, no. 13, pp. 5728–35, 2013.

[3] N. Ding, M. Chatterjee, and J. Z. Simon, "Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure," *Neuroimage*, vol. 88c, pp. 41–46, 2014.

[4] B. Mirkovic, S. Debener, M. Jaeger, and M. De Vos, "Decoding the attended speech stream with multi-channel EEG: implications for online, daily-life applications," *J Neural Eng*, vol. 12, no. 4, pp. 12–46 007, 2015.

[5] S. Akram, A. Presacco, J. Z. Simon, S. Shamma, and B. Babadi, "Robust decoding of selective auditory attention from MEG in a competing-speaker environment via state-space modeling," *Neuroimage*, vol. 124, no. Pt A, pp. 906–917, 2016.

[6] S. Akram, J. Z. Simon, and B. Babadi, "Dynamic estimation of the auditory temporal response function from MEG in competing-speaker environments," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 8, pp. 1896–1905, 2017.

[7] J. O'Sullivan, A. Power, N. Mesgarani, S. Rajaram, J. Foxe, B. Shinn-Cunningham, M. Slaney, S. Shamma, and E. Lalor, "Attentional selection in a cocktail party environment can be decoded from single-trial EEG," *Cereb Cortex*, vol. 25, no. 7, pp. 1697–706, 2015.

[8] L. Fiedler, J. Obleser, T. Lunner, and C. Graversen, "Ear-EEG allows extraction of neural responses in challenging listening scenarios a future technology for hearing aids?" *Conf Proc IEEE Eng Med Biol Soc.*, pp. 5697–5700. doi: 10.1109/EMBC.2016.7 592 020, 2016.

[9] C. Bech Christensen, R. Hietkamp, J. Harte, T. Lunner, and P. Kidmose, "Toward EEG-assisted hearing aids: Objective threshold estimation based on ear-EEG in subjects with sensorineural hearing loss," *Trends in Hearing*, vol. 22, p. 2331216518816203, 2018.

[10] S. Miran, S. Akram, A. Sheikhattar, J. Z. Simon, T. Zhang, and B. Babadi, "Real-time tracking of selective auditory attention from M/EEG: A bayesian filtering approach," *Frontiers in Neuroscience*, vol. 12, no. 262, 2018.

[11] S. Miran, S. Akram, A. Sheikhattar, J. Z. Simon, T. Zhang, and B. Babadi, "Real-time decoding of auditory attention from EEG via bayesian filtering," *Conf Proc IEEE Eng Med Biol Soc*, pp. 25–28, 2018.

[12] A. de Cheveigné and J. Z. Simon, "Denoising based on time-shift PCA," *J of Neurosci Methods*, vol. 165, no. 2, pp. 297–305, 2007.

[13] A. de Cheveigné and J. Z. Simon, "Denoising based on spatial filtering," *J of Neurosci Methods*, vol. 171, no. 2, pp. 331–9, 2008.

[14] J. Särelä and H. Valpola, "Denoising source separation," *J. Mach. Learn. Res.*, vol. 6, pp. 233–272, 2005.

[15] T. Goldstein, C. Studer, and R. Baraniuk, "A field guide to forwardbackward splitting with a FASTA implementation," *arXiv:abs/1411.3406*, 2014.

[16] T. Goldstein, C. Studer, and R. Baraniuk, "FASTA: A generalized implementation of forward-backward splitting," Available online at: http://arxiv.org/abs/1501.04979, 2014.