AN ALTERNATIVE APPROACH FOR AUDITORY ATTENTION TRACKING USING SINGLE-TRIAL EEG

Bradley Ekin[⋆]*, Les Atlas*[⋆]*, Majid Mirbagheri*[†]*, and Adrian KC Lee*[†]

[⋆] Department of Electrical Engineering, University of Washington, Seattle, USA
[†] Institute for Learning & Brain Sciences, University of Washington, Seattle, USA

ABSTRACT

Auditory selective attention plays a central role in the human capacity to reliably process complex sounds in multi-source environments. Stimulus reconstruction has been widely used for the investigation of selective auditory attention using multichannel electroencephalography (EEG). In particular, the influence of attention on sound representations in the brain has been modeled by linear time-variant filters and have been used to track the attentional state of individuals in multi-source environments. Detection of auditory attention is of interest and is important in the study of attention-related disorders and has potential application in the hearing aid and advertising industries. In analogy with the rake receiver from wireless communications, we propose a new strategy, adapting principles from minimum variance beamforming, to reconstruct stimuli for decoding the attentional state of listeners in a competing speaker environment. We show through experiments with real electrophysiological data how decoding accuracies can be improved using our proposed scheme.

Index Terms— auditory, attention, BCI, EEG

1. INTRODUCTION

Non-invasive studies of the brain and brain-computer interfaces (BCI) have typically depended upon electrophysiology. Unfortunately, these electrophysiological methods have often suffered from low signal levels and, thus, low signal-to-noise ratios. Repetition of stimuli for averaging of responses, and many other more sophisticated techniques, such as independent component analysis (e.g. [1]) and inverse channel estimation (e.g. [2]), are allowing electrophysiology, such as for electroencephalograms (EEGs), to become more meaningful measures of activity within the brain.

There are two main reasons that modern data communications are reliable, in spite of shockingly low signal-to-noise ratios: 1) Spatial diversity and 2) temporal diversity. Diversity means multiple elements of a signal or its response is spread out over space or time, respectively. Our paper does not claim innovation in spatial diversity in electrophysiology, since spatial diversity processing has already reached a relatively mature stage via processing of channels from multiple electrodes. These spatial approaches have been deployed and successfully used by the research community (e.g. [3]). The results here instead focus on the integration of both spatial and temporal diversity, which we will show has potential for much more utilization for some key tasks in non-invasive brain interfaces when modeled appropriately, such as detecting a person's state of attention.

This work was funded by ARO grant numbers W911NF-12-1-02770 and W911NF-15-1-0450, as well as the Office of Naval Research Young Investigator Program award N00014-15-1-2124.

In particular our initial results show that modeling a persons attended and unattended temporal EEG responses to dichotic auditory stimuli (i.e. a cocktail party scenario) greatly magnifies electrophysiological evidence of the unattended stimuli, providing a potential complementary method to advance BCI. Our primary goal is to validate the benefit of modeling temporal EEG responses via reconstruction of both attended and unattended speech tokens and passages. This joint use of attended and unattended spatio-temporal diversity for electrophysiological signals provides a new view and tool for the brain science community. Note that the details of this temporal diversity are related to the patterned dynamics of evoked brain responses, and that spatial diversity is related to EEG electrode positions on the scalp. We do not argue against the importance of the structure of the spatio-temporal dynamics of evoked responses and the past excellent progress in this area. We instead propose to sidestep these patterns to mine a successful concept from modern data communications, the rake receiver, to improve the relative detection of attention, especially, as shown by experiments in this paper, by increasing sensitivity to the lack of attention.

1.1. Spatio-Temporal Diversity and the Rake Receiver

Modern data communications techniques such as the rake receiver can advantageously use this dispersion to achieve large signal-to-noise gains by appropriately processing temporal diversity, and even larger gains when combined with spatial structure. The rake receiver estimates the channel, and all their multipath reflection delays and gains. Instead of inverting the channel to estimate an impulse response, as is more conventionally done in electrophysiology (e.g. [4]), this rake approach then aligns the multipath components, phase matching those that differ by delays and weighting by positive or negative values. These delay and gain matched signals are then added constructively, allowing the multipath to combine to greatly increase the received signal-to-noise ratio and reduce estimation variance. Ill-posed and ill-conditioned inverse systems are avoided. With some typical channel assumptions and models, signal level gains of 1-3 orders of magnitude are possible when the number of aligned and added reflections is increased from 0 to 6. (For example, see figure 7 in [5].) To illustrate the temporal element of its function, a low complexity rake receiver use in signal shortening (illustration from [6]) is shown in Figure 1.

The main role of the system in Figure 1 is to take the temporally dispersed evoked response from a single small level input, apply appropriate weights, align in time, and then constructively sum it to form a larger output level, resulting in a larger output signal-to-noise ratio. The weights for each receiver are optimized using both channel estimations as well as the spatial structure between receivers to provide minimum variance reconstruction of the transmission.

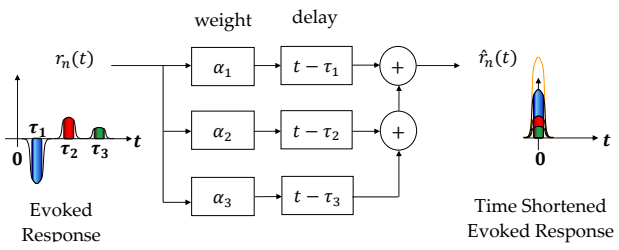

Fig. 1. A simplified picture of a 3-finger rake receiver with its naturally time-spread input and compacted output. The weights and time delays are chosen to maximize the constructive addition so the response is as short duration, with the highest amplitude signal level and lowest estimation variance, as possible.

2. APPLICATION TO SINGLE-TRIAL AUDITORY ATTENTION

Humans possess an outstanding capability to segregate within most classes of multiple simultaneous complex sounds through an act of focusing their attention on a specific source of a sound or spoken words known as selective attention. While the underlying neural processes behind selective attention is not fully known, recent studies suggest that the encoding of auditory objects in the brain while listening to competing speakers is affected by attention. Previous studies show that low-frequency (2 to 8 Hz) cortical activity recorded are linearly related to an 8 Hz low-pass temporal envelope of speech [7]. It has been demonstrated that in dichotic listening (one source to the left ear, another to the right), cortical encoding of an attended speech envelope is substantially stronger than to an unattended envelope [8]. This contrast has been used as a criterion to characterize listener's attention in environments with multiple speakers simultaneously talking using single repetition EEG [9, 10] recordings. For instance, in [9] it is possible to ascertain which of two speech sources is focused on ("attended") based on similarity between a reconstruction of the attended stimulus envelope from neural recordings and the envelope of each of the two speech waveforms over long periods (60 seconds). The parameters of reconstruction filters in this previous method are tuned for the attended signal via a minimum mean squared error (MMSE) criteria. This approach does not require knowledge of forward transformation of attended and unattended waveforms in the brain. These transforms are often modeled via linear temporal response functions (TRF), and can be estimated before a typical experiment, which is to record when the listener is confirmed to be attending to or not attending to the input sound stimuli. By utilizing different optimality criteria, other techniques can be used as an alternative to this standard MMSE method so that we can leverage our knowledge about the TRFs for attended and unattended sounds and the difference between them.

In this paper, we instead use a Capon minimum variance distortionless response (MVDR) beamforming method, similar to the beamformers discussed in [11, 12], to build reconstruction filters for both attended and unattended signals separately. The MVDR beamformer minimizes the total energy of the reconstructed signal while simultaneously keeping the total gain of the channel for the desired signal fixed. Because the gain on the signal is fixed, any reduction in the output energy is obtained by suppressing undesired signal and noise. Similar to [9], we use a similarity measure based on correlation between the reconstructed signals and the true speech envelopes to decode attentional state of the listener.

3. STIMULUS RECONSTRUCTION MODELS

We denote all matrices as bold uppercase letters, for example $\mathbf{A}$, and all vectors as bold lowercase italic letters, for example $\boldsymbol{x}$. The Hermitian transpose of a matrix or vector is denoted by $(\cdot)^H$ and transpose of a matrix or vector is denoted by $(\cdot)^\top$.

A subject is presented with two simultaneous speech waveforms in which one is attended to while the other is not attended to. We denote the 8 Hz low-pass temporal envelopes of these waveforms as $s_a(t)$ and $s_u(t)$ at discrete times $t = 0, ..., T$, respectively. The EEG observed in response to these waveforms is denoted as $r_n(t)$ for a set of $n = 1, ..., N$ electrodes. The underlying assumption for both the forthcoming stimulus reconstruction models assume a linear relationship between the EEG response and stimuli, i.e.:

$$r_n(t) = \sum_\tau [a_n(\tau)s_a(t-\tau) + u_n(\tau)s_u(t-\tau)] + v_n(t) \quad (1)$$

where $a_n(\tau)$ and $u_n(\tau)$ are the attended and unattended temporal response functions, respectively, and $v_n(t)$ is assumed additive noise.

3.1. Minimum Mean Squared Error Reconstruction

The method commonly used in the field for stimulus-reconstruction is centered on finding the filter $\boldsymbol{g}_{\text{MMSE}}$ which minimizes the mean-squared error between the reconstructed stimulus $\hat{s}(t)$ and the true stimulus $s(t)$. This method is known as minimum mean squared error (MMSE) reconstruction, or more commonly, normalized reverse correlation [13, 14]. Details are discussed in [9], where the reconstruction filter is determined by the auto-correlation of EEG data across all time-lags and channels, denoted by $\mathbf{R}$ in the study, and the cross-correlation between $\mathbf{R}$ and the stimulus $\boldsymbol{s}$, i.e. $\boldsymbol{g}_{\text{MMSE}} = (\mathbf{R}\mathbf{R}^\top)^{-1}\mathbf{R}\boldsymbol{s}^\top$.

As each reconstruction filter represents a multivariate impulse response function, filter parameters estimated from numerous training trials can be combined by simply averaging filters together. For training and validation of the model, leave-one-out cross-validation is used. That is, for each test trial, the filter $\boldsymbol{g}_{\text{MMSE}}$ is obtained using the averaged parameters of the filters learned on every other trial. Stimulus reconstruction is performed by evaluating $\hat{\boldsymbol{s}} = \boldsymbol{g}_{\text{MMSE}}^\top\mathbf{R}$. For each trial we develop two reconstruction filters, $\boldsymbol{g}_{\text{a-MMSE}}$ and $\boldsymbol{g}_{\text{u-MMSE}}$, for estimating both the attended and unattended stimulus envelopes, respectively.

3.2. The Rake Reconstruction Filter

One drawback of using the above MMSE reconstruction is that the interference and noise of the system are not directly modeled, but are instead minimized using MMSE criterion on the target stimuli. Our proposed reconstruction method takes advantage of these extra conditions by adapting a spatio-temporal beamforming strategy, recently used in [11, 12] for acoustic signals, which aims to explicitly minimize the noise and interference in the reconstruction, while maintaining unity gain reconstruction of the target stimuli. To achieve this, we estimate parameters of a minimum variance distortionless response (MVDR) rake reconstruction filter. These parameters are estimated using a leave-one-out cross-validation procedure, with the same goal of obtaining two reconstruction filters, $\boldsymbol{g}_{\text{a-MVDR}}$ and $\boldsymbol{g}_{\text{u-MVDR}}$.

3.2.1. Estimating the temporal response function

Temporal response functions (TRFs) for both the attended and unattended states are estimated offline using neural data recorded in a competing-speaker environment with both the attended and unattended streams available. Using vector/matrix notation from (1), the attended TRF, $\boldsymbol{a}_n$, for each EEG channel can be obtained using ridge regression:

$$\boldsymbol{a}_n = (\mathbf{S}_a^\top \mathbf{S}_a + \gamma_n \mathbf{I})^{-1} \mathbf{S}_a^\top \boldsymbol{r}_n \tag{2}$$

where $\mathbf{S}_a$ denotes a Toeplitz convolution matrix of the attended stimulus envelope, and $\mathbf{I}$ denotes identity. A k-fold cross-validation procedure ($k = 5$) was conducted to find the optimum values for the regularization parameters, γ_n, for each EEG channel. The unattended TRFs, $\boldsymbol{u}_n$, are found using this same procedure, where the neural data is fit to the unattended stimulus envelope and its time-lags, $\mathbf{S}_u$.

3.2.2. Minimum variance distortionless response rake filter

The reconstruction filter we propose closely relates to the classic Capon beamformer [15] and the time-domain beamforming methods discussed in [11]. We derive the minimum variance distortionless response (MVDR) rake reconstruction filter assuming the following linear model:

$$\boldsymbol{r}_n = \mathbf{A}_n \boldsymbol{s}_a + \mathbf{U}_n \boldsymbol{s}_u + \boldsymbol{v}_n \tag{3}$$

where

$$\begin{aligned}
\boldsymbol{s}_a &= [s_a(t), s_a(t-1), \ldots, s_a(t-2\tau_{max})]^\top \\
\boldsymbol{s}_u &= [s_u(t), s_u(t-1), \ldots, s_u(t-2\tau_{max})]^\top \\
\boldsymbol{r}_n &= [r_n(t), r_n(t-1), \ldots, r_n(t-\tau_{max}+1)]^\top \\
\boldsymbol{v}_n &= [v_n(t), v_n(t-1), \ldots, v_n(t-\tau_{max}+1)]^\top
\end{aligned} \tag{4}$$

and $\mathbf{A}_n$ and $\mathbf{U}_n$ represent Toeplitz convolution matrices of the attended and unattended TRFs, respectively. Because each trial represents a unique story and neural response, concatenating training set trials together (such that the data in (4) is all-inclusive) yields an improved one-for-all reconstruction filter which is much less arbitrary than averaging filters obtained by each trial individually, as shown with MMSE reconstruction in [16].

By stacking all vectors and matrices of (3), indexed by n, and dropping the index, we obtain its compact model:

$$\boldsymbol{r} = \mathbf{A}\boldsymbol{s}_a + \mathbf{U}\boldsymbol{s}_u + \boldsymbol{v} \tag{5}$$

where

$$\begin{aligned}
\boldsymbol{r} &= [\boldsymbol{r}_1^\top, \boldsymbol{r}_2^\top, \ldots, \boldsymbol{r}_N^\top]^\top \\
\boldsymbol{v} &= [\boldsymbol{v}_1^\top, \boldsymbol{v}_2^\top, \ldots, \boldsymbol{v}_N^\top]^\top \\
\mathbf{A} &= [\mathbf{A}_1^\top, \mathbf{A}_2^\top, \ldots, \mathbf{A}_N^\top]^\top \\
\mathbf{U} &= [\mathbf{U}_1^\top, \mathbf{U}_2^\top, \ldots, \mathbf{U}_N^\top]^\top.
\end{aligned} \tag{6}$$

The optimization problem for the MVDR filter which reconstructs the attended stimulus envelope is expressed as:

$$\text{minimize } E\{|\boldsymbol{g}_a^\top (\mathbf{U}\boldsymbol{s}_u + \boldsymbol{v})|^2\} \text{ subject to } \boldsymbol{g}_a^\top \mathbf{A} = \boldsymbol{\delta}_\tau^\top \tag{7}$$

where $\boldsymbol{\delta}_\tau$ is a vector of zeros, except with the last element equal to one, and $E\{\cdot\}$ denotes expected value. For computational and notation simplicity, the interference term can be absorbed into the noise term such that $\boldsymbol{v}' = \boldsymbol{r} - \mathbf{A}\boldsymbol{s}_a$. The objective can now be developed into:

$$E\{|\boldsymbol{g}_a^\top \boldsymbol{v}'|^2\} = \boldsymbol{g}_a^\top \boldsymbol{\Sigma}_{vv} \boldsymbol{g}_a \tag{8}$$

where $\boldsymbol{\Sigma}_{vv}$ is the auto-covariance matrix of the noise $\boldsymbol{v}'$. The solution to this quadratic program is:

$$\boldsymbol{g}_{\text{a-MVDR}} = \boldsymbol{\Sigma}_{vv}^{-1} \mathbf{A} (\mathbf{A}^\top \boldsymbol{\Sigma}_{vv}^{-1} \mathbf{A})^{-1} \boldsymbol{\delta}_\tau. \tag{9}$$

By redefining $\mathbf{R}$ using the notation of (4) and (6), for $t = 0$, we obtain:

$$\mathbf{R} = [\boldsymbol{r}(\tau), \boldsymbol{r}(\tau+1), \ldots, \boldsymbol{r}(\tau+T)] \tag{10}$$

such that the output of the reconstruction filter can now be evaluated by $\hat{\boldsymbol{s}}_a = \boldsymbol{g}_{\text{a-MVDR}}^\top \mathbf{R}$. The processing above can easily be modified to reconstruct the unattended stimuli analogously.

3.2.3. Inverting the auto-covariance matrix

Although the covariance matrices are estimated using the complete training set, we found the result may still be ill-conditioned. There are numerous ways of inverting ill-conditioned matrices. In this paper we explored principal-component (or eigenspace) covariance inversion [17, Ch. 6.8]. By taking advantage of symmetry properties inherent to covariance matrices, we can express $\boldsymbol{\Sigma}_{vv}$ in terms of its eigenvalues and eigenvectors:

$$\boldsymbol{\Sigma}_{vv} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^H \tag{11}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix of ordered eigenvalues and $\mathbf{Q}$ is a $(N \cdot \tau_{max}) \times (N \cdot \tau_{max})$ matrix of corresponding eigenvectors. The covariance matrix is projected onto a reduced-rank subspace by retaining the first L largest eigenvalues, where L is a parameter tuned on the training set. This subset of eigenvalues can similarly be represented by the matrix $\hat{\boldsymbol{\Lambda}}$, a $L \times L$ ordered diagonal matrix which retains the $L < (N \cdot \tau_{max})$ eigenvalues such that:

$$\hat{\boldsymbol{\Sigma}}_{vv} = \hat{\mathbf{Q}}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{Q}}^H \tag{12}$$

where $\hat{\mathbf{Q}}$ is a $(N \cdot \tau_{max}) \times L$ matrix of the corresponding retained eigenvectors. Inverting the reduced-rank covariance matrix is now better behaved and can be performed by:

$$\hat{\boldsymbol{\Sigma}}_{vv}^{-1} = \hat{\mathbf{Q}}\hat{\boldsymbol{\Lambda}}^{-1}\hat{\mathbf{Q}}^H. \tag{13}$$

4. EXPERIMENTS AND RESULTS

We conducted a pilot study as a proof-of-concept for our proposed method using neural data recorded from a single subject giving informed consent according to the procedures approved by the University of Washington.

4.1. Experiment Setup and Data Acquisition

The experiment consisted of presenting two audio books simultaneously to the subject, one story to the left ear and the other to the right ear. Stories were presented into the same ears throughout the duration of recording. The experiment was segmented into 36 one-minute trials where the subject was asked to attend to one of the two stories throughout the entire trial. To ensure the correct passage was attended to, the subject was required to answer multiple-choice questions on the attended passage after each trial.

EEG data were collected using a 60-electrode EEG cap (Brain Vision products). All EEG data were resampled offline to a sample rate of 64 Hz. The temporal envelopes of the speech stimuli were obtained using the Hilbert transform, and then downsampled to 64 Hz, allowing us to relate their dynamics to those of the EEG. Because envelope frequencies between 2 and 8 Hz are linearly relatable to

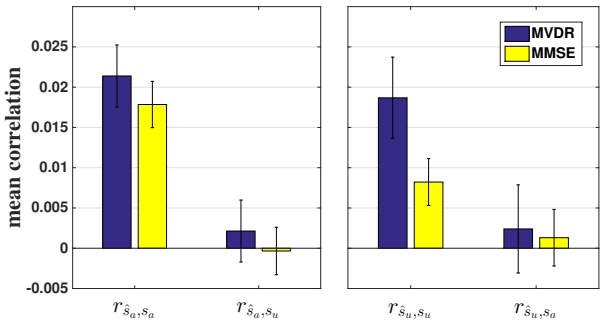

Fig. 2. Average correlation coefficient between the reconstructed stimulus envelopes and actual stimulus envelopes using conventional MMSE reconstruction (yellow) and our proposed MVDR reconstruction (blue). Average correlation is across all 36 test trials.

neural responses in the auditory cortex [18, 19], the EEG data were digitally filtered offline with a band-pass filter between 2 and 8 Hz, and the speech envelopes were low-pass filtered below 8 Hz, similar to preprocessing of [9]. Previous research also indicates EEG activity reflects the dynamics of the speech envelope at latencies up to 250 ms [2], thus, we aimed to develop the reconstruction filters, for both models, over the range of $\tau = 0$ to $\tau_{max} = 250$ ms.

4.2. Evaluation and Results

Following the same evaluation metrics as [9], the reconstruction accuracy was measured based on the correlation coefficient (Pearson's r) between the reconstructed stimuli envelope and the true stimuli envelope over a 60 second test trial. The correlation coefficient $r_{\hat{s}_a,s_a}$ denotes an attended reconstruction correlated against the true attended envelope. The coefficient $r_{\hat{s}_u,s_u}$ is defined similarly for the unattended stimuli. We will refer to these as the "desired" correlation coefficients. To get a relative measure, we also calculated the correlation between the reconstructed envelope and the opposite stimuli, which we will refer to as the "undesired" correlations. We denote these as $r_{\hat{s}_a,s_u}$ and $r_{\hat{s}_u,s_a}$. Measuring all test trials provided 36 correlation coefficients for each of the 4 cases and 2 reconstruction models. Figure 2 depicts the average of these correlations.

Although difficult to make generalized conclusions due to the small sample size of subjects, our initial results show that the proposed MVDR model provides a higher correlated reconstruction with the true stimulus, for both the attended and unattended stimuli, when compared to conventional MMSE reconstruction. It also shows that by modeling the noise (and interference) in the MVDR model, the separation between the desired and undesired correlations are improved, in turn allowing for attentional decoders to perform more efficiently. This improvement in separation between the desired and undesired correlations is shown in Table 1.

The evaluation metric that is most desired in auditory attention detection is the attentional decoder accuracy. Similar to [9], we use the sign of the difference between the desired and undesired correlations for a particular test trial to determine if the attended and unattended passages were detected correctly. That is, for each test trial, where both the MMSE and MVDR models were trained using all other trials, if $r_{\hat{s}_a,s_a} > r_{\hat{s}_a,s_u}$ then the attended passage was detected correctly. Similarly, the unattended passage was detected correctly if $r_{\hat{s}_u,s_u} > r_{\hat{s}_u,s_a}$. It is easy to see the importance between the separation of the desired and undesired correlations and how it relates to attentional state decoding accuracy. The decoding results for all 36 trials is shown in Table 2.

	Conventional MMSE	Proposed MVDR
$E\{r_{\hat{\boldsymbol{s}}_a,\boldsymbol{s}_a} - r_{\hat{\boldsymbol{s}}_a,\boldsymbol{s}_u}\}$	0.0181	**0.0192**
$E\{r_{\hat{\boldsymbol{s}}_u,\boldsymbol{s}_u} - r_{\hat{\boldsymbol{s}}_u,\boldsymbol{s}_a}\}$	0.0062	**0.0165**

Table 1. Difference between desired and undesired correlation coefficients for both attended and unattended stimuli envelope reconstructions. Average correlation is across all 36 test trials.

	Conventional MMSE	Proposed MVDR
Attended Decoder	**86.1%**	**86.1%**
Unattended Decoder	58.3%	**80.6%**

Table 2. Single-trial attentional decoding accuracy. Chance is 50%.

As expected from the correlation separation shown in Table 1, both MMSE and MVDR reconstruction models have similar detection accuracy for the attended audio passage, but our proposed MVDR model greatly outperforms MMSE for unattended audio passage detection. We hypothesize this imbalance in performance is due to the channel response shortening property of the MVDR filter, introducing less estimation variance than results from the deconvolution inherent in the MMSE approach. That increased variance is potentially much more problematic for the weaker unattended decoders.

5. CONCLUSIONS

Based on concepts from communication's ubiquitous rake receiver, we propose an auditory stimulus envelope reconstruction method to be used on single-trial EEG recordings obtained in response to a competing talker environment. Our method was inspired from the classic MVDR beamformer so that not only a desired target is modeled and reconstructed, but any system interference and noise is modeled and directly minimized. Although results are based using only a single subject, we show that both the reconstruction accuracy and auditory attention decoding accuracy of our model performs comparable to the traditional MMSE reconstruction for an attended stimuli, but outperforms MMSE when evaluating the reconstruction of the unattended stimuli.

Future directions for investigation start with evaluation on more subjects. There is also an opportunity to explore the performance of different stimulus envelope representations, for example [16] has shown that by using sub-band envelope extraction with proper power-law compression, as opposed to the standard broadband Hilbert envelope, reconstruction models may yield even better performance in decoding accuracy. Another area of interest is the investigation of other maximal correlation reconstruction models, such as multivariate linear regression and canonical correlation analysis.

6. REFERENCES

[1] Scott Makeig, Anthony J. Bell, Tzyy-Ping Jung, and Terrence J. Sejnowski, "Independent component analysis of electroencephalographic data," in *Proc. of the 1995 Conference on Advances in Neural Information Processing Systems (NIPS)*, Denver, CO, USA, Nov. 1996, vol. 8, pp. 145–151, MIT Press.

[2] Edmund C. Lalor and John J. Foxe, "Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution," *European Journal of Neuroscience*, vol. 31, no. 1, pp. 189–193, Dec. 2010.

[3] Arnaud Delorme and Scott 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, Mar. 2004.

[4] David K. Warland, Pamela Reinagel, and Markus Meister, "Decoding visual information from a population of retinal ganglion cells," *Journal of Neurophysiology*, vol. 78, no. 5, pp. 2336–2350, Nov. 1997.

[5] Kyungwhoon Cheun, "Performance of direct-sequence spread-spectrum rake receivers with random spreading sequences," *Communications, IEEE Transactions on*, vol. 45, no. 9, pp. 1130–1143, Sept. 1997.

[6] Bradley Ekin and Les Atlas, "Modern data communications theory suggests processing for EEG signals," in *Proc. of the IEEE Engineering in Medicine and Biology Society (EMBS) BRAIN Grand Challenges Conference*, Washington, DC, Nov. 2014.

[7] Steven J. Aiken and Terence W. Picton, "Human cortical responses to the speech envelope," *Ear and Hearing*, vol. 29, no. 2, pp. 139–157, Apr. 2008.

[8] Nai Ding and Jonathan Z. Simon, "Neural coding of continuous speech in auditory cortex during monaural and dichotic listening," *Journal of Neurophysiology*, vol. 107, no. 1, pp. 78–89, Jan. 2012.

[9] James A. O'Sullivan, Alan J. Power, Nima Mesgarani, Siddharth Rajaram, John J. Foxe, Barbara G. Shinn-Cunningham, Malcolm Slaney, Shihab A. Shamma, and Edmund C. Lalor, "Attentional selection in a cocktail party environment can be decoded from single-trial EEG," *Cerebral Cortex*, p. bht355, 2014.

[10] Majid Mirbagheri, Bradley Ekin, Les Atlas, and Adrian KC Lee, "Flexible tracking of auditory attention," in *Proc. of the Sixteenth Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Dresden, Germany, Sept. 2015.

[11] Robin Scheibler, Ivan Dokmanic, and Martin Vetterli, "Raking echoes in the time domain," in *Proc. of the 40th IEEE International Conference on Acoustics, Speech and Signal processing (ICASSP)*, Brisbane, Australia, Apr. 2015, IEEE.

[12] Ivan Dokmanic, Robin Scheibler, and Martin Vetterli, "Raking the cocktail party," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 9, no. 5, pp. 825–836, Aug. 2015.

[13] William Bialek, Fred Rieke, Rob R. de Ruyter Van Steveninck, and David Warland, "Reading a neural code," *Science*, vol. 252, no. 5014, pp. 1854–1857, June 1991.

[14] Garrett B. Stanley, Fei F. Li, and Yang Dan, "Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus," *The Journal of Neuroscience*, vol. 19, no. 18, pp. 8036–8042, Sept. 1999.

[15] Jack Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. of the IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

[16] Wouter Biesmans, Jonas Vanthornhout, Jan Wouters, Marc Moonen, Tom Francart, and Alexander Bertrand, "Comparison of speech envelope extraction methods for EEG-based auditory attention detection in a cocktail party scenario," in *Proc. of the 37th Annual Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Milan, Italy, Aug. 2015, pp. 5155–5158, IEEE.

[17] Harry L. Van Trees, *Optimum array processing. Part IV of detection, estimation, and modulation theory*, John Wiley & Sons, Apr. 2004.

[18] Brian N. Pasley, Stephen V. David, Nima Mesgarani, Adeen Flinker, Shihab A. Shamma, Nathan E. Crone, Robert T. Knight, and Edward F. Chang, "Reconstructing speech from human auditory cortex," *PLoS-Biology*, vol. 10, no. 1, pp. 175, Jan. 2012.

[19] Elana M. Zion Golumbic, Nai Ding, Stephan Bickel, Peter Lakatos, Catherine A. Schevon, Guy M. McKhann, Robert R. Goodman, Ronald Emerson, Ashesh D. Mehta, Jonathan Z. Simon, David Poeppel, and Charles E. Schroeder, "Mechanisms underlying selective neuronal tracking of attended speech at a cocktail party," *Neuron*, vol. 77, no. 5, pp. 980–991, Mar. 2013.