Information-Theoretic Limits on the Performance of Auditory Attention Decoders

Ruwanthi Abeysekara[1,2], Christopher J. Smalt[4], I. M. Dushyanthi Karunathilake[1,2],
Jonathan Z. Simon[1,2,3], and Behtash Babadi[1,2]
[1]*Department of Electrical & Computer Engineering, University of Maryland, College Park, MD*
[2]*Institute for Systems Research, University of Maryland, College Park, MD*
[3]*Department of Biology, University of Maryland, College Park, MD*
[4]*Human Health & Performance Systems Group, MIT Lincoln Laboratory, Lexington, MA*
E-mails: ruwanthi@umd.edu, christopher.smalt@LL.mit.edu, dushk@umd.edu,
jzsimon@umd.edu, behtash@umd.edu

***Abstract*—Speaker-specific attention decoding from neural recordings to suppress the acoustic background and extract a target speaker in an in-the-wild multi-speaker conversation scenario poses a cornerstone challenge for advanced hearing devices. Despite several recent advances in auditory attention decoding, most existing approaches fail to reach the real-time performance and attention decoding accuracy required by hearing aid devices. In this work, we aim to quantify fundamental limits on the performance of auditory attention decoding by establishing and computing the trade-off between accuracy and decision window length. We demonstrate the utility of our theoretical bounds in benchmarking the performance of existing widely-used attention decoding algorithms using both simulated and experimentally-recorded magnetoencephalography data.**

***Index Terms*—Auditory attention decoding, information theory, channel capacity, error bounds, MEG**

I. Introduction

Auditory Attention Decoding (AAD) from neural recordings has become an active area of research in computational neuroscience due to the emerging smart hearing aid technology [1]. The goal of AAD is to decode the attentional state of a listener to a target speech stream in a multi-speaker environment, and thereby use it as a feedback to the hearing aid device to enhance the attended speech [2], [3]. In research settings, magnetoencephalography (MEG) and electroencephalography (EEG) are among the frequently used non-invasive techniques to record brain activity in auditory experiments, which are then analyzed to decode selective attention [2], [4]–[14].

A practical hearing aid system that utilizes a brain-computer interface (BCI) could potentially decode the listener's attentional state in an environment with multiple talkers and use microphone arrays to enhance the audio from the desired source [15], [16]. To ensure optimal user experience, near real-time operation and high decoding accuracy are critical factors: incorrect switching between audio sources can confuse the user. Meanwhile, a system that fails to respond promptly to the user's desired attentional focus will not be effective in achieving its intended purpose.

Establishing the capabilities and limitations of AAD algorithms is important in practice, especially for manufacturers of hearing devices. However, benchmarking the performance of AAD methods theoretically can be challenging because it requires an in-depth understanding of the underlying neural mechanism. Furthermore, the performance of an AAD may depend on various factors such as the complexity of the acoustic environment, the decision window length, and signal-to-noise ratio (SNR).

In this work, we provide an information-theoretic framework, based on a commonly used forward model used in auditory neuroscience, and establish an algorithm-agnostic lower bound on the AAD error in a two-speaker environment. We model the auditory brain as a communication channel, in which the input is the binary-valued focus of attention and the output is the observed neural response. We then use techniques from information theory to compute the channel capacity and apply Fano's inequality [17]–[19] to obtain a lower bound on the AAD error.

We present simulation studies as well as application to MEG data from an auditory experiment in a double-speaker setting. Our simulation studies confirm the proposed theoretical bounds by revealing a fundamental trade-off between decision window length and decoding accuracy/channel capacity [20]. We finally validate the theoretical error bound by comparing it with the performance achieved by two commonly-used decoders, namely, the Bayes' and correlation-based decoders, using both simulations and application to MEG data.

The outline of this paper is as follows. In Section II we present our information-theoretic formulation of AAD, characterize the capacity and lower error bound, and outline the parameter estimation procedures. In Section III we provide simulation studies as well as analysis of experimentally-recorded MEG data to validate our theoretical framework, followed by our concluding remarks in Section IV.

DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited. This work has been supported in part by the National Science Foundation awards no. SMA1734892 and CCF1552946, and the National Institutes of Health awards no. R01-DC019394, R01-DC014085, and P01-AG055365. This material is based upon work supported by the Department of the Air Force under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Department of the Air Force.

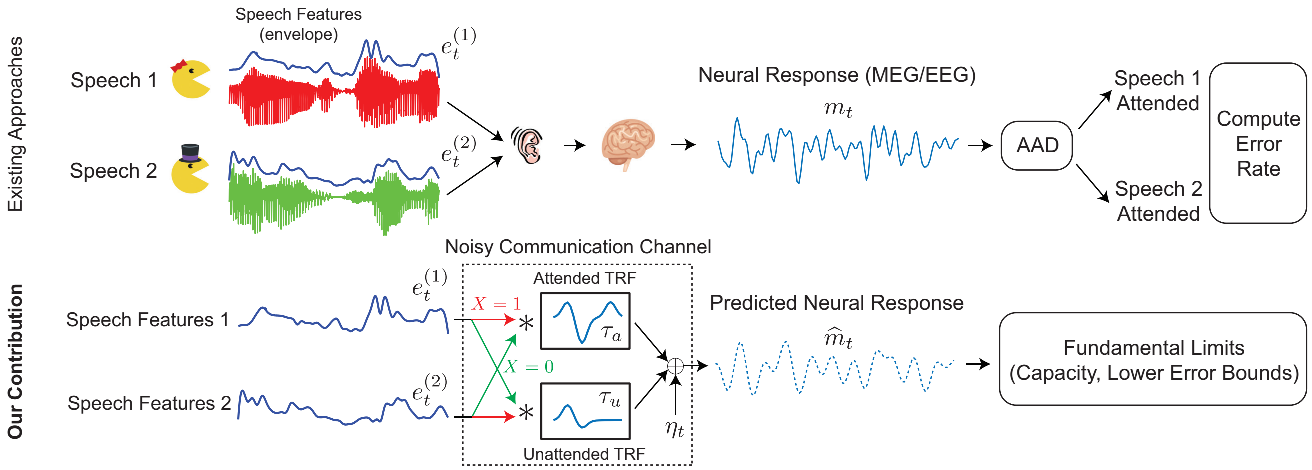

Fig. 1. Schematic depiction of the auditory attention decoding problem. Top: existing approaches use the neural response to decode the attentional state, from which the error rate can be quantified. Bottom: our contribution is to model the AAD problem as a noisy communication channel and use information-theoretic techniques to establish fundamental performance limits such as channel capacity and lower error bound.

II. Problem formulation

We consider a common encoding model of the selective auditory attention, based on the Temporal Response Function (TRF) [21]. In this encoding model, the neural response observed by MEG is a linear convolution of a kernel, known as TRF, with the envelope of the speech. In the presence of two speakers with equal loudness, we consider two TRFs, namely attended and unattended TRFs, denoted by τ_a and τ_u, respectively. The neural response m_t at time t is then modeled as:

$$m_t = \tau_a * e_t^{(a)} + \tau_u * e_t^{(u)} + \eta_t, \tag{1}$$

where $e_t^{(a)}$ and $e_t^{(u)}$ are the speech envelope of the attended and unattended speakers, respectively, and η_t is a noise term representing un-modeled neural processes and measurement noise, which is commonly assumed as a zero-mean Gaussian noise with covariance Σ, i.e., $(\eta_1, \eta_2, \cdots, \eta_T) \sim \mathcal{N}(0, \Sigma)$. For the simplicity of presentation, here we consider a univariate neural response obtained from multi-channel recordings, which is common practice in MEG analysis [22]. Also, the speech feature is taken as the univariate acoustic envelope, which is another common assumption in auditory neuroimaging analysis [4], [21]. As we will explain later, extension of our model to multi-channel recordings and multi-variate speech features is straightforward.

As an example of a conventional AAD algorithm [4], [21], the TRFs are first estimated from pilot/offline data, which then provide two predicted MEG responses as:

$$\begin{aligned} \widehat{m}_t^{(1)} &= \widehat{\tau}_a * e_t^{(1)} + \widehat{\tau}_u * e_t^{(2)}, \quad \text{speaker 1 attended}, \\ \widehat{m}_t^{(2)} &= \widehat{\tau}_a * e_t^{(2)} + \widehat{\tau}_u * e_t^{(1)}, \quad \text{speaker 2 attended}, \end{aligned} \tag{2}$$

where $e_t^{(1)}$, $e_t^{(2)}$ are the speech envelopes of speakers 1 and 2, respectively. Then, the Pearson correlation coefficients between the two predicted responses and the observed signal are computed, and the prediction with the higher correlation is decoded as the attended speaker. Then, the error rate of the AAD can be evaluated using empirical data (Fig. 1, top panel).

In our information theoretic framework, we consider the attentional state as a binary random variable X, taking the value “1” if the listener attended to speaker 1 and taking the value “0” if the listener attended to speaker 2. We also let the MEG observations over a decision window of length T be denoted by $Y := (m_1, m_2, \cdots, m_T)$. Thus, the problem of AAD can be thought of as decoding the transmitted message X from the output of a noisy communication channel given by Y (Fig. 1, bottom panel). In the next subsection, we quantify the capacity and a lower error bound for this noisy communication channel.

A. Computing the Channel Capacity

Recall that the mutual information (MI) between X and Y is defined as:

$$I(X;Y) := h(Y) - h(Y|X), \tag{3}$$

where $h(\cdot)$ denotes the differential entropy. Letting $p := P[X=1]$, the channel capacity is defined as [18]

$$C := \max_{p\in[0,1]} I(X;Y). \tag{4}$$

Letting $f_i(y) := \mathcal{N}(y; \mu_i, \Sigma_i)$ denote the density of y when speaker i is attended, $i = 1, 2$, with $\mu_1 := \tau_a * e^{(1)} + \tau_u * e^{(2)}$ and $\mu_2 := \tau_a * e^{(2)} + \tau_u * e^{(1)}$, the density of y can be expressed as:

$$\begin{aligned} p_Y(y) &= P[X=1]p_Y(y|X=1) + P[X=0]p_Y(y|X=0) \\ &= pf_1(y) + (1-p)f_2(y). \end{aligned} \tag{5}$$

We can thus express $h(Y)$ as:

$$\begin{aligned} h(Y) = &-p\mathbb{E}_{f_1}\left\{\log_2\left(pf_1(y) + (1-p)f_2(y)\right)\right\} \\ &-(1-p)\mathbb{E}_{f_2}\left\{\log_2\left(pf_1(y) + (1-p)f_2(y)\right)\right\}, \end{aligned} \tag{6}$$

where $\mathbb{E}_f\{\cdot\}$ denotes expectation with respect to density f. Also, we have:

$$h(Y|X) = \frac{p}{2}\log_2(2\pi e)^T|\Sigma_1| + \frac{1-p}{2}\log_2(2\pi e)^T|\Sigma_2|. \tag{7}$$

In order to compute the channel capacity, the expression $(h(Y) - h(Y|X))$ needs to be maximized with respect to p. While $h(Y|X)$ has a closed-form dependence on p, $h(Y)$ does not. As such, we use Monte Carlo sampling to approximate the expectations in Eq. (6):

$$h(Y) \approx -\frac{1}{N}\sum_{n=1}^{N}\Big\{p\log_2\Big(pf_1(y_n^{(1)}) + (1-p)f_2(y_n^{(1)})\Big) + (1-p)\log_2\Big(pf_1(y_n^{(2)}) + (1-p)f_2(y_n^{(2)})\Big)\Big\}, \quad (8)$$

where $y_n^{(i)}$ denotes the n^{th} sample from $f_i(y)$, $i = 1, 2, \cdots, N$, and N denotes the number of samples [23].

It is worth noting that drawing samples from high-dimensional distributions and for a large number of iterations is another significant challenge: In addition to the computational challenge of drawing samples from a multi-variate distribution, the number of Monte Carlo samples required for reliable approximations increases exponentially with the dimension, making it computationally expensive when working with high-dimensional data and large sample sizes. To mitigate the first issue, we utilized a reparametrization technique as follows: To draw samples x from the distribution, $\mathcal{N}(\mu, \Sigma)$, we first consider the Cholesky decomposition of $\Sigma = LL^\top$. Then, if z is an i.i.d. normal vector, $x = Lz + \mu$ would be a sample from $\mathcal{N}(\mu, \Sigma)$.

Next, the mutual information as a univariate function of p can be maximized using standard numerical methods, which gives the value of the capacity. Note that extension to multi-variate neural response and multi-variate speech features is straightforward: if there are $K \geq 1$ MEG channels, and $J \geq 1$ speech features, the output Y can be taken as data matrix $Y \in \mathbb{R}^{K\times T}$, and $2J$ TRFs will be used in the encoding model (J per attended/unattended condition). Similarly, extension to multi-speaker environments with $n > 2$ speakers is theoretically possible, but the optimization in Eq. (4) needs to be carried out over a $(n-1)$-simplex, which could be computationally more challenging.

B. Lower Bound on Error

We observe the random variable Y that is related to X by the conditional distribution $p_Y(y|X)$. If $\widehat{X}$ is an estimate of X based on Y, our goal here is to bound the probability of error given by $P[\widehat{X} \neq X]$ [18]. To this end, we use the Fanos inequality [18], [19], [24], [25] as follows: For any two random variables X and Y and for any estimator $\widehat{X}$ such that $X \to Y \to \widehat{X}$, we have:

$$H(P_e) + P_e\log_2(M-1) \geq H(X|\widehat{X}) \geq H(X|Y), \quad (9)$$

where $P_e := P[\widehat{X} \neq X]$, $H(\cdot)$ is the binary entropy function, and M is the cardinality of X. Given that $M = 2$ in our setting, the inequality in Eq. (9) is simplified to

$$P_e \geq H^{-1}\left(H(X|Y)\right), \quad (10)$$

where $H^{-1}(\cdot)$ is the inverse binary entropy function over the domain $[0, 1]$. Note that $H(X|Y) = H(X) - I(X;Y)$, which can be estimated using Monte Carlo sampling as before. Finally, the lower bound in Eq. (10) is computed by considering a uniform prior $p = \frac{1}{2}$.

C. Parameter Estimation

The key parameters to be estimated for the characterization of the model are the TRFs τ_a, τ_u and the covariance matrices Σ_i, $i = 1, 2$. The TRFs are typically be computed using ridge regression [26] or boosting [27]. Here, we use boosting with a 10-fold cross-validation.

Estimating the covariance matrices in real data applications poses a significant challenge, primarily due to the high dimensionality of the estimation problem, given limited data. To address this challenge, we use the shrinkage estimation method of [28]. Let $\{x_i\}_{i=1}^N$ be N vectors drawn from a T-dimensional Gaussian distribution with zero mean and covariance Σ. Despite the fact that the sample covariance $\widehat{S} := \frac{1}{N}\sum_{i=1}^{N} x_i^\top x_i$ is the maximum likelihood estimator and is nearly unbiased for for $N \gg T$, it may not achieve a low Mean Squared Error (MSE) due to its high variance. Moreover, it is ill-posed in the large T regime. A naïve, but well-conditioned estimator for Σ would be:

$$\widehat{F} = \frac{\mathrm{tr}(\widehat{S})}{T} I_T, \quad (11)$$

where I_T denotes the identity matrix in T dimensions and $\mathrm{tr}(\cdot)$ denotes the trace operator. However, this estimator reduces the variance at the expense of a larger bias. To make the bias reasonable, we consider the shrinkage of $\widehat{F}$ towards $\widehat{S}$ by a coefficient $\rho \in (0, 1)$ [28]:

$$\widehat{\Sigma} = (1-\rho)\widehat{S} + \rho\widehat{F}. \quad (12)$$

The so-called oracle estimator $\widehat{\Sigma}_o$ is the solution to:

$$\begin{aligned} \min_{\rho} \quad & \mathbb{E}\left\{\|\widehat{\Sigma}_o - \Sigma\|_F^2\right\} \\ \text{s.t} \quad & \widehat{\Sigma}_o = (1-\rho)\widehat{S} + \rho\widehat{F}. \end{aligned} \quad (13)$$

As shown in [29], for data with unknown sample distribution, the solution to (13) is given by:

$$\rho_o := \frac{\mathbb{E}\left\{\mathrm{tr}((\Sigma - \widehat{S})(\widehat{F} - \widehat{S}))\right\}}{\mathbb{E}\left\{\|\widehat{S} - \widehat{F}\|_F^2\right\}}. \quad (14)$$

Under the Gaussian assumption in our model, Eq. (14) can be expressed as:

$$\rho_o = \frac{(1-2/T)\,\mathrm{tr}(\Sigma^2) + \mathrm{tr}^2(\Sigma)}{(N+1-2/T)\,\mathrm{tr}(\Sigma^2) + (1-N/T)\,\mathrm{tr}^2(\Sigma)}. \quad (15)$$

Given that the true covariance Σ is unknown, in [28] the authors proposed an alternative iterative algorithm to approximate the oracle estimator with provable convergence guarantees. Given an estimate of $\widehat{\Sigma}_\ell$ and ρ_ℓ at iteration ℓ, we update them as:

$$\widehat{\rho}_{\ell+1} = \frac{(1-2/T)\,\mathrm{tr}(\widehat{\Sigma}_\ell\widehat{S}) + \mathrm{tr}^2(\widehat{\Sigma}_\ell)}{(N+1-2/T)\,\mathrm{tr}(\widehat{\Sigma}_\ell\widehat{S}) + (1-N/T)\,\mathrm{tr}^2(\widehat{\Sigma}_\ell)} \quad (16)$$

$$\widehat{\Sigma}_{\ell+1} = (1-\widehat{\rho}_{\ell+1})\widehat{S} + \widehat{\rho}_{\ell+1}\widehat{F}, \quad (17)$$

where the initial estimate $\widehat{\Sigma}_0$ can be taken as either the sample covariance or the Rao-Blackwellized estimate.

III. Results

A. Simulation Studies

In our simulation setting, we considered an experiment in which a subject was tasked with attending to one of two speakers for a period of 60 seconds, and the experiment was repeated 100 times. First, speech envelopes of the two speakers were generated for each of the 100 trials. Attended and unattended TRFs were then generated, preserving the general characteristics of auditory TRFs. Next, the attentional states were generated with equal prior probabilities and taken as the ground truth. Subsequently, the speech envelopes and TRFs were convolved to generate the predicted MEG responses, to which noise was added. For simplicity, we considered equal and diagonal covariance matrices $\Sigma_1 = \Sigma_2$. To simulate realistic levels of noise, the noise variance was computed based on an SNR level of -9 dB, compatible with that observed in actual MEG recordings.

In computing the capacity and the lower bound, the TRFs were estimated using the boosting method described in Section II-C. We considered two AAD approaches: 1) the Bayes' classifier, which computes the posterior probability $P(X|Y)$ and decodes the attentional state based on maximum posterior; 2) We used a correlation-based decoder [4], [21] that calculates the Pearson correlation coefficients between the predicted and observed neural responses and decodes the attentional state based on maximum correlation value.

Since our goal is to analyze how the capacity and the error bound of AAD vary with the available evidence at each decision window, we selected a range of decision window lengths from 1 s to 60 s. For each window length, we performed AAD using the aforementioned methods and then computed the empirical error rate.

Fig. 2A shows the capacity as a function of the decision window T as described in Section II-A. As it can be observed from the figure, at $T > 30$ s, the capacity gets close to 1 bit, implying that there is sufficient information in 30 s of the neural response for reliable recovery of the attentional state. Fig. 2B shows the error rate of the Bayes' and correlation-based decoders, as well as the theoretical lower bound of Section II-B. For large values of $T \sim 60$ s, both the AAD error rates and the lower bound tend to 0, which is expected. However, for small values of T up to 10 s, the AAD methods perform near chance level. Interestingly, there is a significant gap between the lower bound and the AAD error rates for small values of T.This performance gap has two implications: First, for near real-time applications in which $T \sim 1$ s, the theoretical lower bound gives an unimprovable error rate floor of $\sim 27\%$; second, there is theoretically room for improvement by existing AAD algorithms to reach the error rate floor. While the latter provides a benchmark for performance improvement, using more elaborate speech features and access to multi-channel data with higher SNR could mitigate the former.

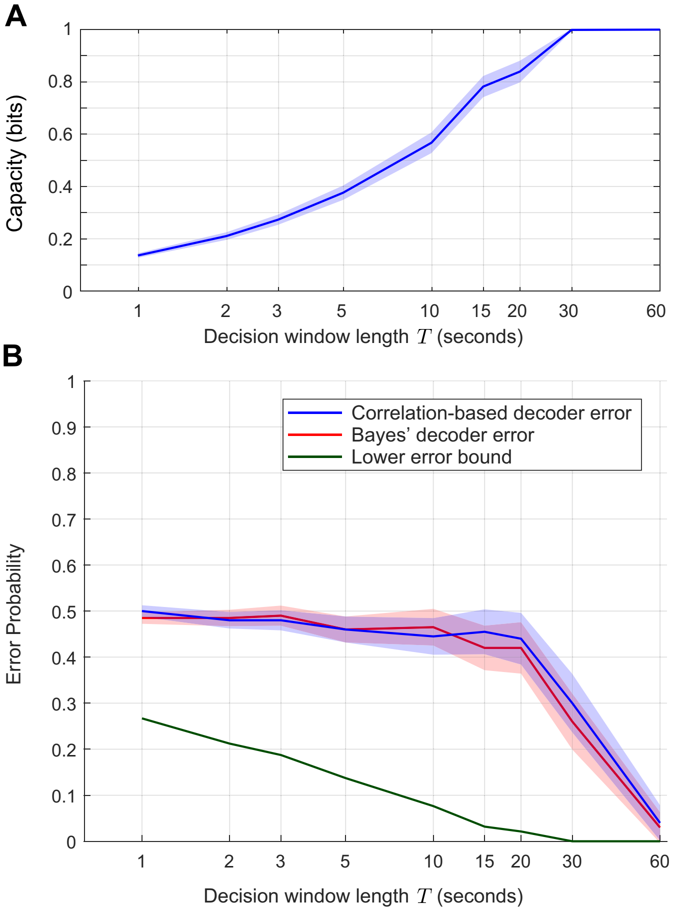

Fig. 2. Results of the simulation study. A) AAD channel capacity as a function of T. B) The error rates of the Bayes' and correlation-based decoders, as well as the theoretical lower bound. The colored hulls show 90% confidence intervals.

B. Analysis of Experimentally-recorded MEG Data

We used an MEG dataset obtained in a cocktail party experiment in which 10 subjects listened to a mixture of two speeches for 60 seconds, over 6 trials [30]. TRFs were computed at a sampling rate of 250 Hz and then everything downsampled to 25 Hz to speed up the computations. We used the Denoising Source Separation (DSS) algorithm [22] to extract the auditory component of the neural response from MEG data and consider the dominant DSS component as the univariate neural response. The length of the TRF kernels was chosen to correspond to 500 ms, consistent with existing literature [21]. We then carried out the same steps as described in Section III-A. Note that the covariance matrices, in this case, are not diagonal and are thus estimated using the shrinkage method of [28] as described in Section II-C.

Fig. 3 shows the error rate of the Bayes' and correlation-based decoders, as well as the theoretical lower bound of Section II-B. Consistent with our simulation study, for large values of $T \sim 60$ s, both the AAD error rates and the lower bound tend to 0. Similarly, for small values of T up to 10 s, the AAD methods perform near chance level. We observe the same gap between the theoretical lower bound and the AAD error rates as in the simulation study, for small values of T.

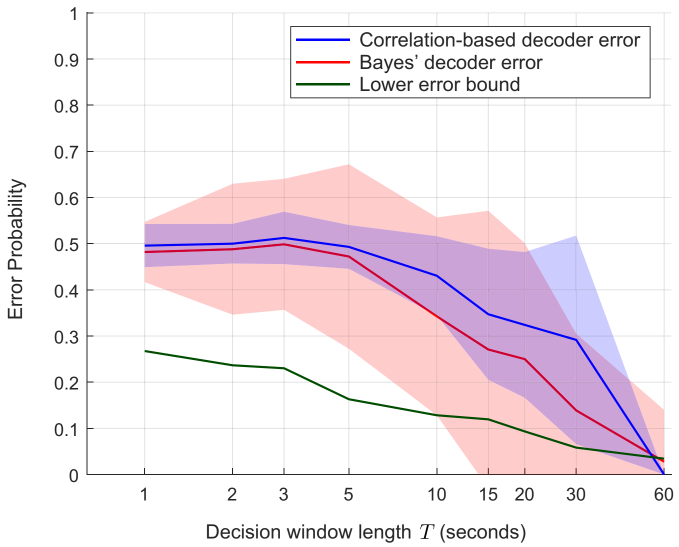

Fig. 3. Results of the real MEG data analysis. The error rates of the Bayes' and correlation-based decoders, as well as the theoretical lower bound. The colored hulls show 90% confidence intervals.

IV. Conclusion

We considered an information-theoretic approach for quantifying the fundamental theoretical limits of AAD. We cast the AAD problem as decoding a message from a noisy communication channel and used the channel capacity and Fano's lower bound to quantify the aforementioned trade-offs. Our approach provides an algorithm-agnostic lower error bound, which can be used to benchmark and assess the performance of existing AAD algorithms. In particular, we demonstrated the impact of limited decision window on the achievable accuracy of AAD decoders in near real-time scenarios. In future work, we will consider an extension of the proposed approach to more complex auditory environments, to more elaborate speech features beyond the acoustic envelope, and to multi-channel EEG/MEG data, and will use the resulting theoretical bounds to benchmark a wider range of existing AAD algorithms.

References

[1] N. Mesgarani and E. F. Chang, "Selective cortical representation of attended speaker in multi-talker speech perception," *Nature*, vol. 485, no. 7397, pp. 233–236, 2012.

[2] S. Geirnaert, S. Vandecappelle, E. Alickovic, A. de Cheveigne, E. Lalor, B. T. Meyer, S. Miran, T. Francart, and A. Bertrand, "Electroencephalography-based auditory attention decoding: Toward neurosteered hearing devices," *IEEE Signal Processing Magazine*, vol. 38, no. 4, pp. 89–102, 2021.

[3] B. J. Borgström, M. S. Brandstein, G. A. Ciccarelli, T. F. Quatieri, and C. J. Smalt, "Speaker separation in realistic noise environments with applications to a cognitively-controlled hearing aid," *Neural Networks*, vol. 140, pp. 136–147, 2021.

[4] J. A. O'sullivan, A. J. Power, N. Mesgarani, S. Rajaram, J. J. Foxe, B. G. Shinn-Cunningham, M. Slaney, S. A. Shamma, and E. C. Lalor, "Attentional selection in a cocktail party environment can be decoded from single-trial eeg," *Cerebral cortex*, vol. 25, no. 7, pp. 1697–1706, 2015.

[5] S. Akram, A. Presacco, J. Z. Simon, S. A. Shamma, and B. Babadi, "Robust decoding of selective auditory attention from meg in a competing-speaker environment via state-space modeling," *NeuroImage*, vol. 124, pp. 906–917, 2016.

[6] J. OSullivan, Z. Chen, J. Herrero, G. M. McKhann, S. A. Sheth, A. D. Mehta, and N. Mesgarani, "Neural decoding of attentional selection in multi-speaker environments without access to clean sources," *Journal of neural engineering*, vol. 14, no. 5, p. 056001, 2017.

[7] 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, p. 262, 2018.

[8] G. Ciccarelli, M. Nolan, J. Perricone, P. T. Calamia, S. Haro, J. Osullivan, N. Mesgarani, T. F. Quatieri, and C. J. Smalt, "Comparison of two-talker attention decoding from eeg with nonlinear neural networks and linear methods," *Scientific reports*, vol. 9, no. 1, p. 11538, 2019.

[9] S. Geirnaert, T. Francart, and A. Bertrand, "An interpretable performance metric for auditory attention decoding algorithms in a context of neuro-steered gain control," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 1, pp. 307–317, 2019.

[10] S. Miran, A. Presacco, J. Z. Simon, M. C. Fu, S. I. Marcus, and B. Babadi, "Dynamic estimation of auditory temporal response functions via state-space models with gaussian mixture process noise," *PLoS computational biology*, vol. 16, no. 8, p. e1008172, 2020.

[11] S. Geirnaert, T. Francart, and A. Bertrand, "Unsupervised self-adaptive auditory attention decoding," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 10, pp. 3955–3966, 2021.

[12] S. Cai, P. Li, E. Su, and L. Xie, "Auditory attention detection via cross-modal attention," *Frontiers in Neuroscience*, vol. 15, p. 652058, 2021.

[13] D. D. Wong, S. A. Fuglsang, J. Hjortkjær, E. Ceolini, M. Slaney, and A. De Cheveigne, "A comparison of regularization methods in forward and backward models for auditory attention decoding," *Frontiers in neuroscience*, vol. 12, p. 531, 2018.

[14] E. Alickovic, T. Lunner, F. Gustafsson, and L. Ljung, "A tutorial on auditory attention identification methods," *Frontiers in neuroscience*, p. 153, 2019.

[15] W. Pu, J. Xiao, T. Zhang, and Z.-Q. Luo, "A joint auditory attention decoding and adaptive binaural beamforming algorithm for hearing devices," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 311–315.

[16] T. Dau and A. de Cheveigné, "Toward a cognitively controlled hearing aid," *The Journal of the Acoustical Society of America*, vol. 141, no. 5, pp. 3893–3893, 2017.

[17] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[18] T. Cover and J. A. Thomas, "Elements of information theory," 2006.

[19] S. Gerchinovitz, P. Ménard, and G. Stoltz, "Fanos inequality for random variables," *Statistical Science*, vol. 35, no. 2, pp. 178–201, 2020.

[20] F. M. Reza, *An introduction to information theory*. Courier Corporation, 1994.

[21] N. Ding and J. 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, 2012.

[22] A. de Cheveigné and J. Z. Simon, "Denoising based on spatial filtering," *Journal of neuroscience methods*, vol. 171, no. 2, pp. 331–339, 2008.

[23] M. F. Huber, T. Bailey, H. Durrant-Whyte, and U. D. Hanebeck, "On entropy approximation for gaussian mixture random vectors," in *2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*. IEEE, 2008, pp. 181–188.

[24] R. G. Gallager, *Information theory and reliable communication*. Springer, 1968, vol. 588.

[25] J. Wolfowitz, *Coding theorems of information theory*. Springer Science & Business Media, 2012, vol. 31.

[26] M. J. Crosse, G. M. Di Liberto, A. Bednar, and E. C. Lalor, "The multivariate temporal response function (mtrf) toolbox: a matlab toolbox for relating neural signals to continuous stimuli," *Frontiers in human neuroscience*, vol. 10, p. 604, 2016.

[27] S. V. David, N. Mesgarani, and S. A. Shamma, "Estimating sparse spectro-temporal receptive fields with natural stimuli," *Network: Computation in neural systems*, vol. 18, no. 3, pp. 191–212, 2007.

[28] Y. Chen, A. Wiesel, Y. C. Eldar, and A. O. Hero, "Shrinkage algorithms for mmse covariance estimation," *IEEE transactions on signal processing*, vol. 58, no. 10, pp. 5016–5029, 2010.

[29] O. Ledoit and M. Wolf, "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection," *Journal of empirical finance*, vol. 10, no. 5, pp. 603–621, 2003.

[30] I. D. Karunathilake, J. L. Dunlap, J. Perera, A. Presacco, L. Decruy, S. Anderson, S. E. Kuchinsky, and J. Z. Simon, "Effects of aging on cortical representations of continuous speech," *bioRxiv*, pp. 2022–08, 2022.