Rate-Distortion under Neural Tracking of Speech

A Directed Redundancy Approach

Østergaard, Jan; Jayaprakash, Sangeeth Geetha; Ordoñez, Rodrigo

DOI (link to publication from Publisher):
10.48550/arXiv.2501.16762

Publication date:
2025

Document Version
Early version, also known as pre-print

Rate-Distortion under Neural Tracking of Speech: A Directed Redundancy Approach

Jan Østergaard, Sangeeth Geetha Jayaprakash, and Rodrigo Ordoñez

Department of Electronic Systems
Aalborg University
Fredrik Bajers Vej 7
Aalborg, 9000, Denmark
{jo,sgjay,rop}@es.aau.dk

Abstract

The data acquired at different scalp EEG electrodes when human subjects are exposed to speech stimuli are highly redundant. The redundancy is partly due to volume conduction effects and partly due to localized regions of the brain synchronizing their activity in response to the stimuli. In a competing talker scenario, we use a recent measure of directed redundancy to assess the amount of redundant information that is causally conveyed from the attended stimuli to the left temporal region of the brain. We observe that for the attended stimuli, the transfer entropy as well as the directed redundancy is proportional to the correlation between the speech stimuli and the reconstructed signal from the EEG signals. This demonstrates that both the rate as well as the rate-redundancy are inversely proportional to the distortion in neural speech tracking. Thus, a greater rate indicates a greater redundancy between the electrode signals, and a greater correlation between the reconstructed signal and the attended stimuli. A similar relationship is not observed for the distracting stimuli.

Introduction

Biological cognitive systems are capacity limited in the sense of having finite memory, processing power, and communication capabilities. Since Shannon's foundational work — a mathematical theory of communications [1], there has been numerous works addressing the information rate of the human brain from different perspectives. Factors such as synaptic transmission speed, neural network connectivity, and sensory integration of information across different brain regions all play a role in determining the information rate. For an overview of information theory in cognition, see e.g., [2]. Whilst the body senses and communicates millions of bits per second of data to the brain, only a fraction of this information is processed by the conscious mind. This data reduction is due to the attentional filter being able to compress information and remove irrelevant information [3]. Barlow hypothesized that the brain makes use of efficient coding to compress the information and guarantee that neurons transmit information with minimal redundancy [4]. The information of spiking neurons were considered in e.g., [5–9]. The information capacity of the human visual system has been addressed in [10, 11]. The information capacity and the information loss of the human auditory system was considered in [12] and [13], respectively. A rate-distortion perspective for the human working memory was considered in [14, 15] and analytical information theoretic frameworks for neural processing were given in [16] and [17]. A partial information decomposition [18] approach was taking in [19].

The field of information rates in the brain is vast and we limit here the scope to works related to neural (cortical) speech tracking. Cortical tracking of speech (CTS) refers to the phenomonen that slow moving cortical oscillations in the delta and theta bands are modulated by the envelope of the attended sound [20, 21]. The auditory regions are particularly synchronized to the oscillations of the envelope of the modulating acoustic stimuli, and the strength of the CTS relates to the phase-locking (or neural tracking) capacity of the auditory cortex [22]. CTS is used in e.g., auditory attention detection (AAD) problems, where a subject is exposed to multiple simultaneous acoustic stimuli. It is common to have two competing acoustic stimuli such as speech versus speech [23] or speech versus music [24]. The goal is to detect which of the acoustic stimuli the subject is paying attention to based on the subject's cortical activity [25]. The cortical activity is often represented by non-invasive scalp or in-ear electroencephalogram (EEG) recordings, which are mapped to the envelope-domain of the stimulus using linear or non-linear models [23, 26–29]. The method of regularized least-squares regression utilizes a linear model that is fitted (trained) to the EEG signal [23]. The resulting models are then correlated with the envelopes of the acoustic stimuli, and the stimulus having the greatest correlation is considered to be the one that the subject is paying attention to. Information-theoretic limits on the performance of auditory attention decoders were obtained in [30].

To assess the redundancy in EEG signals, one could potentially use the mutual information measure [31]. The mutual information is only defined for two sets of signals and is therefore not able to quantify the shared information between three or more signals [32]. A lower bound of the redundant information in an arbitrary number of signals were given in [33] and a new notion of directed redundancy was recently presented in [34].

In this paper, subjects were exposed to two competing acoustic stimuli (male and female speech) and asked to focus on one of them (the attended signal) and ignore the other (the distractor). We use linear decoders to reconstruct the stimuli from the signals obtained by scalp electrodes located in the proximity of the left temporal region of the brain. The distortion is quantified by $D = 1 - |\rho|$, where ρ is the Pearson correlation between the reconstructed signal and the attended stimulus. The amount of redundant information causally conveyed from the acoustic stimuli via the left temporal region and to the reconstructed stimuli is measured using the directed redundancy measure of [34]. It is demonstrated that the directed redundancy is inversely proportional to the distortion. Thus, a greater amount of redundant information corresponds to a lower distortion and thereby better neural tracking of the acoustic stimuli. Moreover, an increase in rate from the acoustic stimuli and to the reconstructed stimuli is also inversely proportional to the distortion. Interestingly, we do not observe such relationships when assessing the distracting stimuli.

Notation

Let $X^n = X_1, X_2, \ldots, X_n$, where $X_k \in \mathbb{R}$ are random variables. To simplify the notation, we sometimes write X instead of X^n. We use upper case letters for random variables and lower case letters for their realizations.

Directed Redundancy

Let $\mathbb{TE}$ denote Schreiber's Transfer Entropy, which is a directed information measure defined as [35]:

$$\mathbb{TE}(X^n \to Z^n) \triangleq I(X^{n-1}; Z_n | Z^{n-1}). \tag{1}$$

From (1) it can be noticed that transfer entropy is a variant of conditional mutual information, where the mutual information between the current sample Y_n of the target Y and the past X^{n-1} of the source X is quantified conditioned upon the past Y^{n-1} of the target.

Consider the system depicted in Fig. 1b. Here $\phi = \phi^n$ is a process that drives the redundancy in the system. We are then interested in quantifying the amount of redundant information about ϕ that is causally conveyed via X and Y to Z. This is non-trivial to compute since a positive transfer entropy from X to Z and from Y to Z does not mean that X and Y are necessarily sharing the same information with Z. Moreover, X and Y are not (necessarily) causally coupled to each other, so we cannot use the transfer entropy from X to Y or from Y to X as a measure of redundant information. In this work, we will instead rely on the causal (directed) redundancy measure of [34], which is formally described by the following definition:

Definition 1 *[34]* The transfer-entropy measurable causal redundancy provided to the target Z via the source X and Y, which are driven by the hidden redundancy process ϕ is defined as:

$$\begin{aligned} I^{\text{red}}(X;Y;Z) \triangleq \min\ (&\mathbb{TE}(\phi \to Z), \mathbb{TE}(\phi \to T_{XZ}), \mathbb{TE}(\phi \to T_{YZ}), \\ &\mathbb{TE}(T_{XZ} \to Z), \mathbb{TE}(T_{YZ} \to Z), I(T_{XZ}; T_{YZ})), \end{aligned} \tag{2}$$

where T_{XZ} is a *causal* minimal sufficient statistics for X with respect to Z so that T_{XZ} satisfies the conditional Markov chain

$$X^{n-1}|_{Z^{n-1}} - T_{XZ}^{n-1}|_{Z^{n-1}} - Z_n|_{Z^{n-1}}, \forall n,$$

where

$$\begin{aligned} T_{XZ}^{n-1} = \arg\min_{S_{XZ}^{n-1}} & I(S_{XZ}^{n-1}; Z_n | Z^{n-1}) \\ s.t. \quad & I(S_{XZ}^{n-1}; Z_n | Z^{n-1}) = I(X^{n-1}; Z_n | Z^{n-1}), \forall n. \end{aligned}$$

In Definition 1, the causal sufficient statistics of X and Y with respect to Z are included to make sure that only information that is relevant to Z is taken into account. One can allow additional input to X and Y, e.g., via feedback from Z or other sources directly interacting with X and Y. In such a case, one can further include the terms $I(\phi; T_{XZ})$ and $I(\phi; T_{YZ})$ within the minimization of (2) to make sure only information relevant to ϕ is taken into account.

The causal sufficient statistics, T_{XZ} and T_{YZ} are generally hard to compute and it was suggested in [34] to replace them by their corresponding processes. This would then lead to an upper bound on the directed information.

Lemma 1 *[34]* Consider the dynamical system in Fig. 1b. The amount of redundant transfer entropy communicated via X and Y to Z, and contained in Z, is upper bounded by:

$$\begin{aligned} I^{\text{red}}(X;Y;Z) \leq \min\big(&\mathbb{TE}(\phi \to Z), \mathbb{TE}(\phi \to X), \\ &\mathbb{TE}(\phi \to Y), \mathbb{TE}(X \to Z), \mathbb{TE}(Y \to Z)\big). \end{aligned} \tag{3}$$

In the above example, we considered two processes X and Y. In general, we could have K processes $X(1), \ldots, X(K)$, where $X(i) = X_1(i), X_2(1), \ldots, X_n(1)$.

Auditory Attention Decoding

At any given time, subjects are presented with two competing acoustic signals A and B as illustrated in Fig. 1a. These signals are represented as time series using the convention $A^n = A_1, A_2, \ldots, A_n$, where $A_i \in \mathbb{R}$ for $i = 1, \ldots, n$. A similar notation applies to B. We will only be using the envelopes of the acoustic signals, which we will be denoting $S = S^n$. It will be clear from context whether S is the attended signal or the distractor.

While the subject is listening to the stimuli, the brain response is recorded via EEG electrodes that are spatially distributed on the scalp. We will be focusing on the left-temporal ($\mathcal{LT}$) region, which cover the left auditory cortex and have been shown to be good for cortical tracking of speech [36]. We define $\mathcal{LT}$ as the following set of electrodes in the extended 10-20 international 64 channel EEG electrode layout convention [37]:

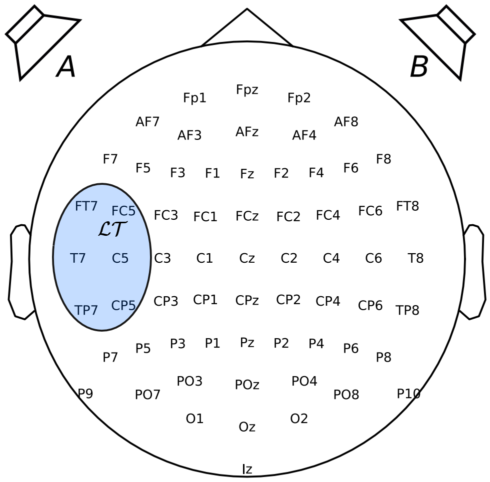
$$\mathcal{LT} = \{\mathrm{FT}_7, \mathrm{T}_7, \mathrm{TP}_7, \mathrm{CP}_5, \mathrm{FC}_5, \mathrm{C}_5\}. \tag{4}$$

Stimulus reconstruction

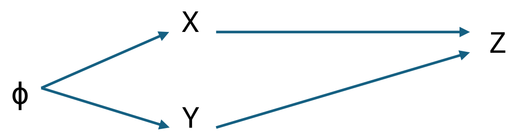
Let s denote the envelope of the stimulus A. To reconstruct the stimulus s from the neural responses, we use a backward stimulus-response mapping that exploits all neural responses in a given time frame [23]. Specifically, let $\hat{s}_t$ be the reconstructed stimuli at time t and let $r(t, c)$ denote the instantaneous neural response sampled at time t at EEG electrode c. Then, if using a linear decoder $g(\tau, c)$, the reconstructed stimuli can be written on the following form [38]:

$$\hat{s}_t = \sum_c \sum_\tau r(t+\tau, c) g(\tau, c), \tag{5}$$

where the outer sum runs over the number of EEG electrodes and the inner sum over time delays τ in a desired range of delays to accommodate the non-instantaneous



(a) A subject is exposed to competing acoustic signals A and B and is focusing on A. Labels indicate EEG electrodes. The shaded area illustrates the electrodes in $\mathcal{LT}$.



(b) The process ϕ drives the redundancy in the system. Information about ϕ is causally conveyed to Z via X and Y. In our work, ϕ models A or B, X and Y represents two of the six electrodes in $\mathcal{LT}$ and Z is the reconstruction of A or B from the electrodes in $\mathcal{LT}$.

Figure 1: The physical setup considered in this work is illustrated in (a) and a model for the physical setup is shown in (b).

effects of the acoustic stimuli. The optimal linear decoder $\mathbf{g}$ that minimizes the mean square error: $\sum_t (s_t - \hat{s}_t)^2$ can be computed by [38]:

$$\mathbf{g} = (R^T R + \lambda I)^{-1} R^T s, \tag{6}$$

where λ is a regularization term, R is a matrix containing the lagged time series of the response $r(t+\tau, c)$, and $\mathbf{g}$ is the vector containing the optimal decoder coefficients [38].

In the sequel, we denote by $\hat{s} = \hat{s}_1, \dots, \dots, \hat{s}_n$ the reconstructed stimuli obtained using the 6 electrodes in $\mathcal{LT}$. We assume that s and $\hat{s}$ are normalized to zero-mean and unit-variance, which means that the stimulus-reconstruction distortion in terms of the mean squared error is an affine function of the correlation between the signals, that is:

$$\frac{1}{2}\mathbb{E}[\|\hat{S} - S\|^2] = 1 - \rho, \tag{7}$$

where $\rho = \mathbb{E}[\hat{S}S]$. To avoid that negative correlations penalizes the distortion, we use the following distortion measure:

$$D \triangleq 1 - |\rho|. \tag{8}$$

Rate Redundancy in EEG Signals

Let $E^n(j) = E_1(j), E_2(j), \dots, E_n(j)$, $j \in \mathcal{LT}$, denote the time-series representation of the j-th EEG electrode signal.

The minimum rate from the electrodes in $\mathcal{LT}$ to the reconstructed stimuli is then found as:

$$R_{E\to\hat{S}} \triangleq \min_{j\in\mathcal{LT}} \mathbb{TE}(E^n(j) \to \hat{S}^n). \tag{9}$$

We note that $\hat{S}^n$ is a deterministic function of $(E^n(1),\dots,E^n(|\mathcal{LT}|))$ according to (5), which implies that $I(E^n(1),\dots,E^n(|\mathcal{LT}|);\hat{S}^n) = 0$ or $I(E^n(1),\dots,E^n(|\mathcal{LT}|);\hat{S}^n) = \infty$. This is not a problem in this work, since $\hat{S}^n$ is a non-deterministic function of the individual electrode signals, which implies that $I(E^n(j);\hat{S}^n)$ is bounded for all $j = 1,\dots,|\mathcal{LT}|$.

The minimum rates from the acoustic stimuli and to the electrodes are given by:

$$R_{S\to E} \triangleq \min_{j\in\mathcal{LT}} \mathbb{TE}(S^n \to E^n(j)). \tag{10}$$

Finally, the rate from the stimulus to the reconstructed stimulus is given by:

$$R_{S\to\hat{S}} \triangleq \mathbb{TE}(S^n \to \hat{S}^n). \tag{11}$$

The directed redundancy for the brain region is then upper bounded by R, which is found using Lemma 1, that is:

$$R \triangleq \min(R_{S\to\hat{S}}, R_{E\to\hat{S}}, R_{S\to E}). \tag{12}$$

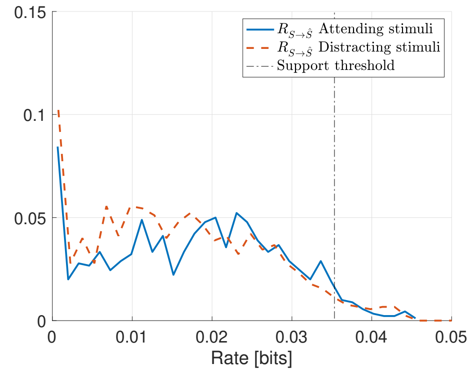
Simulation Study

We use the EEG signals described in detail in [39]. The subjects were exposed to competing acoustic stimuli in the form of male and female talkers. 15 subjects each having 60 repeated trials were used. The temporal envelopes of the target and distractor were acquired and downsampled to 64 Hz to match the sampling frequency of the EEG signals.

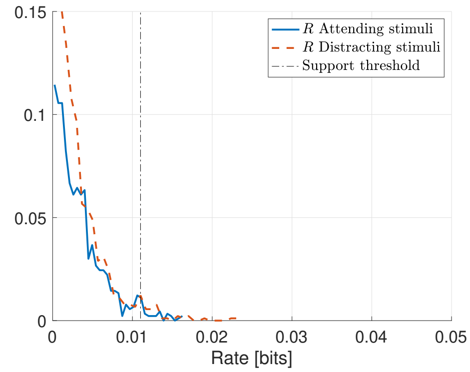
In order to do stimulus reconstruction, we trained subject dependent linear decoders using EEG data only from the electrodes in $\mathcal{LT}$. The decoders for reconstructing the stimuli were obtained using the mTRF-Toolbox [38]. Separate decoders were trained for the attended and distracting stimuli.

The probability density functions of the rates (9), (10), and (11) as well as for the directed rate redundancy (12) are shown in Figs. 2a – 2d. Due to the minimization operation invoked in (9), (10), and (12), the support of the these informational rates are less than for (11). In general, greater rates are increasingly unlikely to appear. In these figures, the vertical dashed-dotted black lines indicate the maximum rate at which the corresponding pdf is above 0.01. In the sequel, we will only be using distortion points, where the associated rates are within the *support* region, i.e., on the left side of this support threshold. We note that rate points outside this support region have very few associated distortion points, and the distortion-rate estimate for these rates would therefore be very poor.

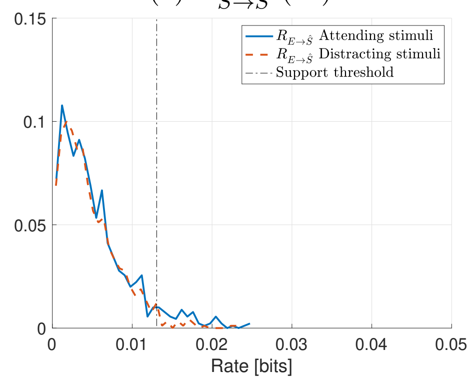
In Figs. 3a and 3b, we have shown the distortion (8) (in dB) as a function of the different types of rates considered. Specifically, we have shown the distortion as a



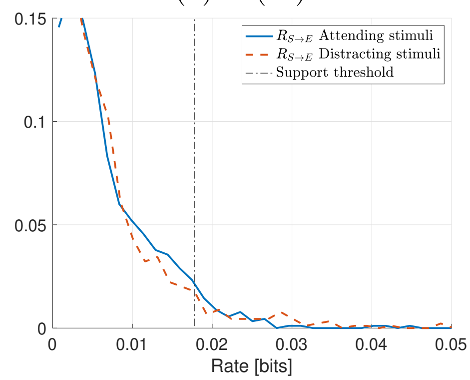
(a) $R_{S\to\hat{S}}$ (11)



(b) R (12)



(c) $R_{R\to\hat{S}}$ (9)

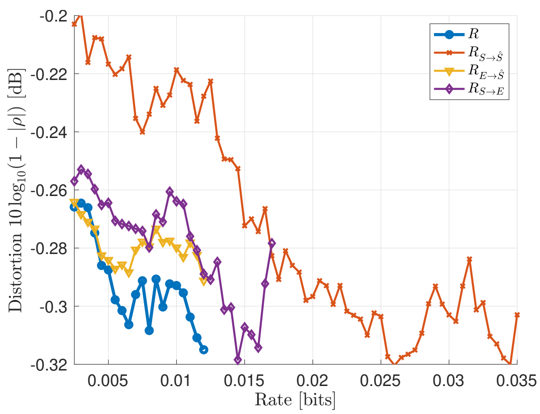


(d) $R_{S\to E}$ (10)

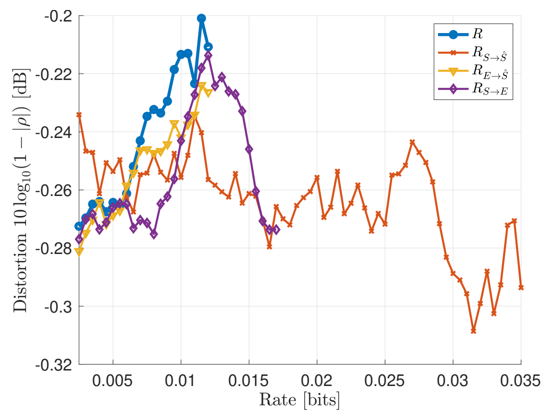
Figure 2: Probability density functions of the rates (9), (10), and (11) as well as for the directed rate redundancy (12). The two cases of attending stimuli (blue) and distracting stimuli (dashed red) are shown.

function of $R_{E\to S}$ (9), $R_{S\to E}$ (10), and $R_{S\to\hat{S}}$ (11) as well as R (12). This is shown for the attended stimuli in Fig. 3a and for the distracting stimuli in Fig. 3b. The distortion-rate curves are obtained by splitting the rate into overlapping intervals of 0.005 bits, and averaging all the distortion points that are associated with rates within the same bin. Only rates within the support regions are considered.

From Fig. 3a, it can be observed that the distortion shows a decreasing trend as a function of the transfer entropy $R_{S\to\hat{S}}$ from the attended stimuli to the reconstructed signal. Such a trend is less clear for the distracting stimuli in Fig. 3b. From the figures, it can also be seen that the distortion shows a decreasing (increasing) trend as a function of the directed redundancy R for attended (distracting) stimuli. Thus, for the attended stimuli, we generally observe that greater correlations (smaller distortions) are approximately proportional to the amount of transfer entropy measured between the acoustic signal and the reconstructed signal. Moreover, there is also an increase in the redundant information R about S, which is conveyed by the electrode signals to $\hat{S}$. Thus, a greater information flow from the attended stimuli to the reconstructed stimuli also leads to greater redundancy in the EEG signals. On the other hand, for the distracting stimuli, the directed redundancy appears to be inversely proportional to the distortion. Thus, an increase in the speech tracking



(a) Attending stimuli



(b) Distracting stimuli

Figure 3: Operational distortion-rate curves as a function of rates (9), (10), and (11) as well as the directed redundancy (12).

performance of the distracting stimuli corresponds to a decrease in the amount of redundant information shared by the EEG electrodes.

To assess the strength of the potentially linear relationship between the distortion (expressed in dB) and the rates and directed redundancy, we have fitted a linear regression model to the corresponding rate-distortion points. We use the fitlm command in Matlab, and test the null hypothesis that the model parameters of the models are zero. The corresponding p-values are reported in Table 1. The boldfaced p-values indicate the cases for which a linear model of the rates is a statistically significant (less than 0.05) predictor for the corresponding distortion. It can be seen that R is a statistical significant linear predictor for the distortion both in case of attended or distracting stimuli. Moreover, in the case the attending stimuli, the slopes are negative, and for the distracting stimuli, the slope is positive, cf. Figs. 3a and 3b.

Table 1: p-values when fitting a linear model to the rate-distortion points. Bold faced values are significant ($p < 0.05$).

	$R_{S\to\hat{S}}$	$R_{S\to E}$	$R_{E\to S}$	R
Attended	**6.1e-06**	0.0930	0.2387	**0.0177**
Distractor	0.1886	**0.0448**	0.0553	**0.0416**

Conclusions

We demonstrated using real-world EEG data that the distortion in an auditory attention decoding study is linearly related to the rate (transfer entropy) from the acoustic stimuli to reconstructed stimuli using the brain response from the left temporal region. Moreover, the distortion is also linearly related to the directed redundancy of the EEG electrode signals in the left temporal region. This suggests that a greater correlation between the envelope of the speech stimuli and the reconstructed stimuli is obtained when a greater amount of redundant information about the stimuli is conveyed in the brain signals.

References

[1] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, 1948.

[2] K. Sayood, "Information theory and cognition: A review," *Entropy*, vol. 20, no. 9, 2018.

[3] D. Broadbent, *Perception and communication*, Pergamon Press, 1958.

[4] H. B. Barlow, *Possible principles underlying the transformations of sensory messages*, In W. Rosenblith (Ed.), Sensory communication, MA: MIT Press, 1961.

[5] C Stevens and A. Zador, "Information through a spiking neuron," in *Advances in neural information processing systems*, 1995.

[6] S. E. Marzen, M. R. DeWeese, and J. P. Crutchfield, "Time resolution dependence of information measures for spiking neurons: scaling and universality," *Frontiers in Computational Neuroscience*, vol. 9, 2015.

[7] M. Veletić, P. A. Floor, Y. Chahibi, and I. Balasingham, "On the upper bound of the information capacity in neuronal synapses," *IEEE Transactions on Communications*, vol. 64, no. 12, pp. 5025–5036, 2016.

[8] K. Aghababaiyan, V. Shah-Mansouri, and B. Maham, "Axonal channel capacity in neuro-spike communication," *IEEE Transactions on NanoBioscience*, vol. 17, no. 1, pp. 78–87, 2018.

[9] K. Aghababaiyan, V. Shah-Mansouri, and B. Maham, "Capacity and error probability analysis of neuro-spike communication exploiting temporal modulation," *IEEE Transactions on Communications*, vol. 68, no. 4, pp. 2078–2089, 2020.

[10] L. Zhaoping, "Theoretical understanding of the early visual processes by data compression and data selection," *Network: Computation in Neural Systems*, vol. 17, pp. 301 – 334, 2006.

[11] C. J. Bates, R. A. Lerch, C. R. Sims, and R. A. Jacobs, "Adaptive allocation of human visual working memory capacity during statistical and categorical learning," *Journal of Vision*, vol. 19, 2019.

[12] H. Jacobson, "The informational capacity of the human ear," *Science (New York, N.Y.)*, vol. 112, 1950.

[13] M. Z. Jahromi, A. Zahedi, J. Jensen, and J. Østergaard, "Information loss in the human auditory system," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 3, pp. 472–481, 2019.

[14] A. M. V. Jakob and S. J. Gershman, "Rate-distortion theory of neural coding and its implications for working memory," *eLife*, vol. 12, pp. e79450, jul 2023.

[15] C. J. Bates and R. A. Jacobs, "Efficient data compression in perception and perceptual memory," *Psychological Review*, vol. 127, pp. 891 – 917, 2020.

[16] X. Yang, K. Wang, and S.A. Shamma, "Auditory representations of acoustic signals," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 824–839, 1992.

[17] T. Berger and W. B. Levy, "A mathematical theory of energy efficient neural computation and communication," *IEEE Transactions on Information Theory*, vol. 56, no. 2, pp. 852–874, 2010.

[18] P. L. Williams and R. D. Beer, "Nonnegative decomposition of multivariate information," *CoRR*, vol. abs/1004.2515, 2010.

[19] A. J. Gutknecht, M. Wibral, and A. Makkeh, "Bits and pieces: Understanding information decomposition from part-whole relationships and formal logic," *Proceedings of the Royal Society A*, vol. 477, 2021.

[20] 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.

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

[22] M. F. Issa, I. Khan, M. Ruzzoli, N. Molinaro, and M. Lizarazu, "On the speech envelope in the cortical tracking of speech," *NeuroImage*, vol. 297, pp. 120675, 2024.

[23] J.A. O'Sullivan, A.J. Power, N. Mesgarani, and et al., "Attentional selection in a cocktail party environment can be decoded from single-trial eeg," *Cereb Cortex*, vol. 25, 2015.

[24] A. Simon, G. Loquet, J. Østergaard, and S. Bech, "Cortical auditory attention decoding during music and speech listening," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, June 2023.

[25] N. Mesgarani and E. F. Chang, "Selective cortical representation of attended speaker in multi-talker speech perception," *Nature*, vol. 485, 2012.

[26] D. D. E. Wong, S. A. Fuglsang, J. Hjortkjær, E. Ceolini, M. Slaney, and A. de Cheveigné, "A comparison of regularization methods in forward and backward models for auditory attention decoding," *Frontiers in Neuroscience*, vol. 12, 2018.

[27] E. Alickovic, T. Lunner, F. Gustafsson, and L. Ljung, "A tutorial on auditory attention identification methods," *Frontiers in Neuroscience*, vol. 13, 2019.

[28] T. de Taillez, B. Kollmeier, and B. T. Meyer, "Machine learning for decoding listeners' attention from electroencephalography evoked by continuous speech," *European Journal of Neuroscience*, vol. 51, no. 5, pp. 1234–1241, 2020.

[29] R. Reetzke, G.N. Gnanateja, and B. Chandrasekaran, "Neural tracking of the speech envelope is differentially modulated by attention and language experience," *Brain Lang.*, vol. 213, 2021.

[30] R. Abeysekara, C. J. Smalt, I. M. Dushyanthi Karunathilake, J. Z. Simon, and B. Babadi, "Information-theoretic limits on the performance of auditory attention decoders," in *2023 57th Asilomar Conference on Signals, Systems, and Computers*, 2023, pp. 1479–1483.

[31] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 2012.

[32] R. Yeung, "A first course in information theory," *Springer*, 2002.

[33] J. Østergaard, "Higher-order common information," *Electronically available at: https://arxiv.org/abs/2406.02001*, 2024.

[34] J. Østergaard, "Directed redundancy in time series," in *IEEE International Symposium on Information Theory (ISIT)*, 2024.

[35] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.*, vol. 85, 2000.

[36] M. F. Issa, I. Khan, M. Ruzzoli, N. Molinaro, and M. Lizarazu, "On the speech envelope in the cortical tracking of speech," *NeuroImage*, vol. 297, 2024.

[37] E. Niedermeyer and F. L. da Silva, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, Lippincott Williams & Wilkins, 2004.

[38] 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, pp. 604, 2016.

[39] S. A. Fuglsang, T. Dau, and J. Hjortkjær, "Noise-robust cortical tracking of attended speech in real-life environments," *NeuroImage*, vol. 156, pp. 435 – 444, 2017.