Directed Information Measures for Assessing Perceived Audio Quality using EEG

Ketan Mehta
Klipsch School of Electrical and
Computer Engineering,
New Mexico State University, NM 88003

Jörg Kliewer
Helen and John C. Hartmann Dept.
of Electrical and Computer Engineering,
New Jersey Institute of Technology, NJ 07103

***Abstract*—We exploit directed information to examine the causal relationship between EEG data in response to audio stimulus. Specifically, we conduct experiments wherein the EEG activity of subjects is recorded as they listen to audio with time-varying audio quality between two different levels. Different types of directed information measures are then used to quantify the information flow between EEG sensors, which are grouped into different regions of interest over the cortex. Further, we determine the analytical relationship between these different directional measures and compare how well they are able to distinguish between the perceived audio quality.**

I. Introduction

Brain processing in general is a multistage process that results in the activation and hierarchical interaction of several different regions in the brain. Understanding the dynamics of brain functioning requires the investigation of the information flow over this cortical network and studying the interaction and connectivity between different areas of the brain. Further, sensory and motor information in the brain is represented and manipulated in the form of neural activity. This superposition of electrical activity patterns can be recorded via electrodes on the scalp and is termed as electroencephalography (EEG). In the following, we exploit directed information (DI) measures to examine the causal information flow between EEG sensors on the scalp in response to an external audio stimulus of the human test subject.

The current state-of-the-art approach to subjective audio quality testing, Multi Stimulus with Hidden Anchor (MUSHRA) [1], has human participants assign a single quality-rating score to each test sequence. Such testing suffers from a subject-based bias towards cultural factors in the local testing environment and can be highly variable. Instead of using EEG measurements we can directly capture and analyze the brainwave response patterns that depend only on the perceived variation in signal quality. For example, [2] investigates the use of a time-space-frequency analysis to identify features in EEG brainwave responses corresponding to time-varying audio quality. Also in [3] the authors use linear discriminant analysis classifiers to extract features from EEG for classifying noise detection in audio signals.

Information theory provides a stochastic framework which is well suited to characterize and model EEG responses [4]–[9]. Specifically, there has been a growing interest in using DI measures to analyze EEG data. In [10] the authors use DI to identify the causal relation between EEG signals and later extend this approach to estimate the number [11] and location [12] of the EEG sources. Similarly, [13] applies DI to evaluate emotional elicitation in subjects. Also, [14] reviews and demonstrates the merits of using transfer entropy (TE) among other causal information measures to study effective connectivity between EEG data. To the best of our knowledge however, this is the first time that a DI-based characterization is used in conjunction with EEG measurements to quantify human perception of audio quality. Further, we determine the analytical relationship between multivariate DI measures and TE, and compare how well these measures are able to distinguish between the perceived audio quality.

II. Experiment

In the conducted experiment the EEG response activity of human test subjects was recorded as they listened to a variety of audio test-sequences. The quality of the stimulus test-sequences varied with time between two different levels — 'high' quality and 'distorted' quality.

All audio test sequences were created from three fundamentally different base-sequences sampled at a 'high' quality of 44.1 kHz, with a precision of 16 bits per sample. Two different types of distortions were considered for our analysis, scalar quantization and frequency band truncation. The test sequence for a specific trial was created by selecting one of the two distortion types and applying it to the original base sequence in a time-varying pattern of non-overlapping five second blocks. Here, we employ the same test sequences and distortion quality levels as in [15]. Multiple of such trials were conducted for each subject by choosing all possible combinations of sequences, distortion types, and time-varying patterns.

The trials were conducted in a room specifically built for recording EEGs with a RF shielded testing chamber. We employed an ActiveTwo Biosemi EEG system, which captures data on 128 spatial channels, sampled at 256 Hz. To better manage the large amount of collected data while effectively mapping the activity across different regions of the brain, we group the 128 electrodes of the EEG-system into specific regions of interest (ROI) as shown in Fig. 1.

This work was supported in part by NSF grant CCF-1065603.

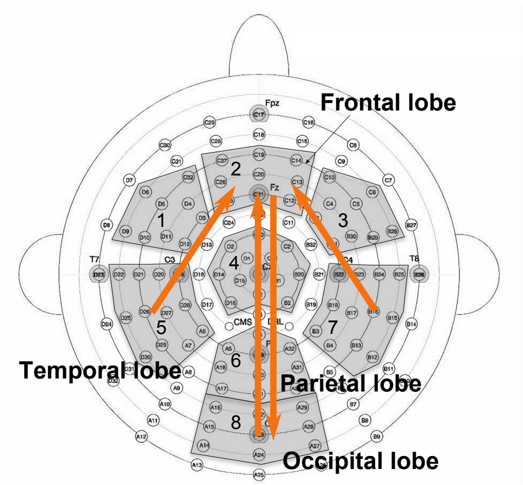

Fig. 1: The 128 electrodes are grouped into eight regions of interest (ROI) to effectively cover the different cortical regions (lobes) of the brain. The arrows illustrate the direction and transmission of information in between the ROIs. For example, shown above is the interaction between the temporal lobes (ROIs 5 and 7) and the pre-frontal cortex (ROI 2).

In our earlier work [16] we demonstrated that the EEG output of an ROI, where every electrode in a given ROI is considered an independent realization of the same random process, converges to a Gaussian distribution with zero mean. The intuition here is that the potential recorded at the EEG electrode at any given time-instant can be considered as the superposition of responses of a large number of neurons. Thus, the distribution of a sufficiently high number of these trials taken at different time instances converges to a Gaussian distribution as a result of the Central Limit Theorem.

III. Directed Information Measures

Given two interacting random processes, DI measures allow us to quantize the causality and the direction of information transmitted between one random process to another. In the following, we will consider three different DI measures presented in the literature.

Notation: $\mathbf{X} = X^n$ is is a vector of n random variables (r.v.) $[X_1, X_2, \ldots, X_n]$, with realizations $x^n = [x_1, x_2, \ldots, x_n]$. Such a random vector in the following represents a sampled EEG signal. Also, $H(X)$ is the entropy of the r.v. X, and $I(X;Y)$ is the mutual information between X and Y.

A. Transfer entropy

TE was first introduced by [17] to analyze information flow over a bi-directional channel, and later formally defined by [18]. It evaluates the deviation of the observed data from a model, assuming the following joint Markov property

$$p(Y_n|Y^{n-1}X^{n-1}) = p(Y_n|Y^{n-1}). \tag{1}$$

The above equation is satisfied when the transition probabilities of the process $\mathbf{Y}$ is independent of $\mathbf{X}$. TE measures the deviation from this condition using the Kullback-Leibler divergence and is defined as

$$\begin{aligned}\mathrm{TE}(\mathbf{X} \to \mathbf{Y}) &\equiv \mathrm{TE}(X^{n-1} \to Y^n) \\ &= E\left[\log \frac{p(Y_n|Y^{n-1}X^{n-1})}{p(Y_n|Y^{n-1})}\right] \\ &= H(Y_n|Y^{n-1}) - H(Y_n|Y^{n-1}X^{n-1}). \end{aligned} \tag{2}$$

Also, we know that the entropy of a k-dimensional multivariate Gaussian distribution with probability density $p(z_1 \ldots z_k)$ is

$$H(Z_1 \ldots Z_k) = \frac{1}{2} \log{(2\pi e)^k} |\mathsf{C}(Z_1 \ldots Z_k)|, \tag{3}$$

where C is the covariance matrix and $|\cdot|$ is the determinant of the matrix. As indicated above, the random vectors X^n and Y^n are jointly Gaussian and therefore the TE can be reduced to

$$\mathrm{TE}(\mathbf{X} \to \mathbf{Y}) = \frac{1}{2} \log \frac{|\mathsf{C}(Y^n)| \cdot |\mathsf{C}(Y^{n-1}X^{n-1})|}{|\mathsf{C}(Y^{n-1})| \cdot |\mathsf{C}(Y^n X^{n-1})|}. \tag{4}$$

B. Massey's directed information

DI as proposed by Massey [19] is an extension of Shannon's mutual information to quantify information flow on a communication channel with feedback. It is defined as the total information transmitted between n past samples of $\mathbf{X}$ and the current sample of $\mathbf{Y}$,

$$\begin{aligned}\mathrm{DI}_1(\mathbf{X} \to \mathbf{Y}) &\equiv \mathrm{DI}_1(X^N \to Y^N) \\ &= H(Y^N) - H(Y^N||X^N) \qquad (5) \\ &= \sum_{n=1}^{N} I(X^n; Y_n|Y^{n-1}), \qquad (6)\end{aligned}$$

where, $H(X^N||Y^N) = \sum_{n=1}^{N} H(X_n|X^{n-1}Y^n)$ is the uncertainty of X^N *causally* conditioned on Y^N. Again, if the random vectors X^n and Y^n are Gaussian then this measure can be conveniently expressed as

$$\mathrm{DI}_1(\mathbf{X} \to \mathbf{Y}) = \frac{1}{2} \sum_{n=1}^{N} \log \frac{|\mathsf{C}(X^n)| \cdot |\mathsf{C}(X^{n-1}Y^n)|}{|\mathsf{C}(X^{n-1})| \cdot |\mathsf{C}(X^n Y^n)|}. \tag{7}$$

C. Kamitake's directed information

Another useful measure of DI is defined by Kamitake in [20] and given as

$$\begin{aligned}\mathrm{DI}_2(\mathbf{X} \to \mathbf{Y}) &= \sum_{n=1}^{N} I(X_n; Y_{n+1}^N | X^{n-1} Y^n) \qquad (8) \\ &= H(Y_{n+1}^N|X^{n-1}Y^n) - H(Y_{n+1}^N|X^n Y^n). \qquad (9)\end{aligned}$$

We notice that this measure is different from Massey's DI in that it measures the influence of the the current sample of $\mathbf{X}$ on the future samples of $\mathbf{Y}$. For the Gaussian case DI_2 can be written as

$$\mathrm{DI}_2(\mathbf{X} \to \mathbf{Y}) = \frac{1}{2} \sum_{n=1}^{N} \log \frac{|\mathsf{C}(X^{n-1}Y^N)| \cdot |\mathsf{C}(X^n Y^n)|}{|\mathsf{C}(X^{n-1}Y^n)| \cdot |\mathsf{C}(X^n Y^N)|}. \tag{10}$$

IV. A Cut-set Bound Approach

A. Information transfer over the brain as a MAC with feedback

Consider a multiterminal communication network where several senders transmit information to only one receiver node via a multiple access channel (MAC). We assume the MAC has a feedback link so the users see the previous outputs of the channel and can use these to choose subsequent channel inputs. This model can be considered analogous to the transmission of information flow over the cortical network, with two or more ROIs (the "senders") being the nodes all communicating to another ROI (the "receiver").

In [21] it is shown that the capacity region for a MAC with feedback can be lower bounded using DI in a form similar to the standard cut-set bound for the MAC [22] for the case without feedback. In order to state the capacity region for the MAC with feedback, we first extend our DI expression to account for the causal influence of additional random processes.

Definition 1. Multivariate Massey DI is defined as the information flowing from $\mathbf{X}$ to $\mathbf{Y}$ causally conditioned on the sequence $\mathbf{Z}$ as

$$\begin{aligned}\mathrm{DI}_1(\mathbf{X}\to\mathbf{Y}||\mathbf{Z}) &= H(Y^N||Z^N) - H(Y^N||X^N Z^N)\\ &= \sum_{n=1}^{N} I(X^n; Y_n|Y^{n-1}Z^{n-1}). \end{aligned}\tag{11}$$

Now consider the two user discrete memoryless MAC with feedback, channel inputs X^N and Y^N, and channel output Z^N as shown in Fig. 2. The cut-set based bound for this channel can be given as follows [21]

$$R_1 \le \frac{1}{N}\mathrm{DI}_1(X^N \to Z^N||Y^N), \tag{12}$$

$$R_2 \le \frac{1}{N}\mathrm{DI}_1(Y^N \to Z^N||X^N), \tag{13}$$

$$R_1 + R_2 \le \frac{1}{N}\mathrm{DI}_1(X^N, Y^N \to Z^N), \tag{14}$$

for all

$$\begin{aligned}p(x_n, y_n|x^{n-1}, y^{n-1}, z^{n-1}) &= p(x_n|x^{n-1}, z^{n-1})\\ &\quad\cdot p(y_n|y^{n-1}, z^{n-1}).\end{aligned}\tag{15}$$

Note that for $N \to \infty$ the equality condition in (12)-(14) represents the capacity region for the MAC with feedback.

We now aim to apply these results to our cortical region model in Fig. 1, where the quantities A^N, B^N, C^N and D^N, resp., represent random vectors describing the sampled output of the EEG signals in each region. However, note that these random vectors are very likely correlated, and most certainly the cortical network is not memoryless. Further, we do not know whether the distribution of these source signals match the capacity achieving distribution of the underlying multiple access communication channel. The rate-bounds in (12)-(14) are therefore not a valid bound on the capacity region of the information transfer between the ROI. However, our interest here is not to calculate the capacity but rather to analyze the nature of causal information flow over the cortical network in response to external stimulus. Therefore, the equations on the right-hand side of (12)-(14) provide us with a useful measure for calculating the multivariate directional information flow between the different interacting ROIs.

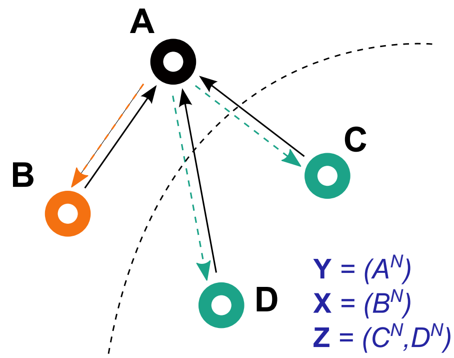

Fig. 2: Cut-set region graph: Information transfer over the ROIs is considered equivalent to a MAC with feedback. In the figure shown above, A is the receiver (output), B is the sender (input), C and D represent the side information (inputs) available at the receiver.

B. Relationship between different measures

Similar to DI_1 TE can be extended for the case of analyzing the information transfer between three or more random processes [14].

Definition 2. Multivariate TE can be defined as follows

$$\mathrm{TE}(\mathbf{X}\to\mathbf{Y}|\mathbf{Z}) = E\left[\log\frac{p(Y_n|Y^{n-1}X^{n-1}Z^{n-1})}{p(Y_n|Y^{n-1})Z^{n-1}}\right] \tag{16}$$

$$= I(X^{n-1}; Y_n|Y^{n-1}Z^{n-1}). \tag{17}$$

Proposition 1. The relation between multivariate DI_1 and multivariate TE is given by

$$\begin{aligned}\mathrm{DI}_1(X^N\to Y^N||Z^N) = \sum_{n=1}^{N}\{&\mathrm{TE}(X^{n-1}\to Y^n|Z^n)\\ &+ I(X_n; Y_n|X^{n-1}Y^{n-1}Z^{n-1})\}\end{aligned}\tag{18}$$

Proof. Using the chain rule of mutual information along with (8) we obtain

$$\begin{aligned}I(X^n; Y_n|Y^{n-1}Z^{n-1}) &= \mathrm{TE}(X^{n-1}\to Y^n|Z^{n-1})\\ &\quad + I(X_n; Y_n|X^{n-1}Y^{n-1}Z^{n-1}).\end{aligned}\tag{19}$$

Substituting the above equation in (11) directly yields the desired result. We observe that the DI is a sum over the TE and an additional term describing the conditional instantaneous information between X_n and Y_n. □

Kamitake's DI expression can also be extended to the multivariate case.

Definition 3. Multivariate Kamitake information is defined as

$$\mathrm{DI}_2(\mathbf{X}\to\mathbf{Y}||\mathbf{Z}) = \sum_{n=1}^{N} I(X_n; Y_{n+1}^N|X^{n-1}, Y^n, Z^{n-1}). \tag{20}$$

We would like to note here that even though the multivariate case for DI_2 has been earlier explored by [10], the authors use an incorrect expansion to calculate the measure, resulting in an expression valid only if the time series is i.i.d. In all other cases their result constitutes only a lower bound. Further, the relation between Kamitake's DI and Massey's DI has previously been discussed in [23], [24]. We extend this to the multivariate case,

Proposition 2. The relation between multivariate DI_1 and multivariate DI_2 is given by

$$\begin{aligned}\mathrm{DI}_2(\mathbf{Y} \to \mathbf{X}||\mathbf{Z}) &+ \mathrm{DI}_1(\mathbf{X} \to \mathbf{Y}||\mathbf{Z}) \\ &= \mathrm{DI}_2(\mathbf{X} \to \mathbf{Y}||\mathbf{Z}) + \mathrm{DI}_1(\mathbf{Y} \to \mathbf{X}||\mathbf{Z}). \end{aligned} \tag{21}$$

Proof. Rewriting (20) in terms of entropy we get

$$\begin{aligned}\mathrm{DI}_2(\mathbf{Y} \to \mathbf{X}||\mathbf{Z}) &= \sum_{n=1}^{N} \big[H(X_{n+1}^N|Y^{n-1}X^nZ^{n-1}) \\ &\qquad - H(X_{n+1}^N|Y^nX^nZ^{n-1})\big] \\ &= \sum_{n=1}^{N} \big[H(X^N|Y^{n-1}Z^{n-1}) - H(X^n|Y^{n-1}Z^{n-1}) \\ &\qquad - H(X^N|Y^nZ^{n-1}) + H(X^n|Y^nZ^{n-1})\big] \\ &= \sum_{n=1}^{N} I(X^N;Y_n|Y^{n-1}Z^{n-1}) - \mathrm{DI}_1(\mathbf{X} \to \mathbf{Y}||\mathbf{Z}).\end{aligned}$$

Rearranging yields

$$\begin{aligned}\mathrm{DI}_2(\mathbf{Y} \to \mathbf{X}||\mathbf{Z}) &+ \mathrm{DI}_1(\mathbf{X} \to \mathbf{Y}||\mathbf{Z}) \\ &= \sum_{n=1}^{N} I(X^N;Y_n|Y^{n-1}Z^{n-1}) \\ &\overset{(a)}{=} \sum_{n=1}^{N}\sum_{i=1}^{N} I(X_i;Y_n|X^{i-1}Y^{n-1}Z^{n-1}) \\ &\overset{(b)}{=} \sum_{n=1}^{N} I(Y^N;X_n|X^{n-1}Z^{n-1}) \\ &= \mathrm{DI}_2(\mathbf{X} \to \mathbf{Y}||\mathbf{Z}) + \mathrm{DI}_1(\mathbf{Y} \to \mathbf{X}||\mathbf{Z}),\end{aligned} \tag{22}$$

where (a) follows by using the chain rule, and (b) from interchanging the order of summation. □

V. Analyzing Real EEG Brain Data

In order to compare and examine the effectiveness of each of these directional measures we apply them to the collected experimental EEG data. We begin by both selecting a source region and a destination region which correspond to the random processes $\mathbf{X}$ and $\mathbf{Y}$, respectively. All the other remaining regions are considered to represent the side information $\mathbf{Z}$. For our analysis here, we separately extract the EEG response sections for each of the two audio quality levels and calculate the DI measures individually for each of them. This allows us to compare the differences in the information flow in between the brain regions for the case where the subject listened to good quality audio as opposed to the case where the subject listened to bad quality audio. Also for the purposes of this analysis we only select a subset of 8 subjects who provide the largest mutual information values on the event related potential (ERP) channel connecting the audio stimulus and the quantized EEG sensor outputs (see [16] for details).

Also, since we know that the interacting random processes are Gaussian, we can directly calculate the information measures via (4), (7), and (10) by estimating the corresponding covariance matrices. We assume stationarity of the observed EEG signal and extract 125 ms long samples ($N = 32$ points) to create the sample covariance matrix. The samples are extracted from all trials across all subjects for different test sequences and distortion types.

Our choice of source and destination ROIs here is closely based on the direction and order of the auditory sensing pathway in the brain. The primary auditory cortex (PAC) located in the left and right temporal lobes is the first region of the cerebral cortex to receive auditory input. The higher executive functions and subjective responses are a result of the information exchange between the PAC and the other cortical regions, predominantly including the prefrontal cortex.

A. Directional information results

Fig. 3 illustrates multivariate Massey DI_1 calculated for four different ROIs for a second of EEG trial data. The plot shows DI calculated separately for both the audio qualities. The results indicate a notable difference between the amount of information flow for good and distorted audio. In particular, there appears to be a higher amount of information flow between the regions when the subject listened to distorted quality audio. This strongly indicates an increase in brain activity, possibly as a result of paying increased attention to identify the drop in audio quality. We also note that in Fig. 3 in one or more cases of ROI pairs a sharp distinction in the behavior of the information flow at around the 400-500 ms mark can be observed. This coincides with the P300 ERP [25] that marks a subject's detection and preliminary reaction to a stimulus.

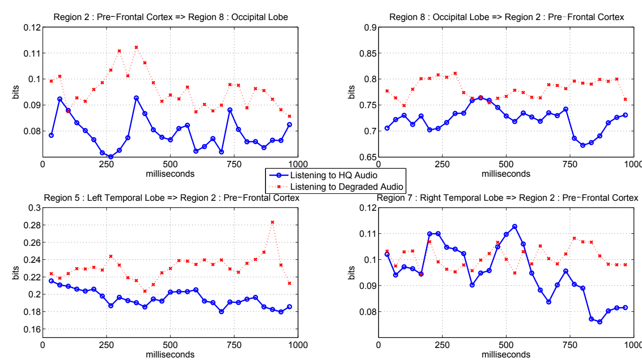

Fig. 3: Multivariate Massey DI

Similarly, Fig. 4 plots multivariate Kamitake DI_2 for the same four regions and a second of trial data. The results show that the causal connectivity between the regions as revealed by DI_2 is very similar to that of DI_1. As before, the amount of information flow in general is higher for distorted audio quality, and the P300 ERP event is observed as well. This similarity between the performances of the two DI measures can can be explained by Proposition 2 which shows that the total flow of information between two random sequences is the same for both measures. Therefore both the directional and temporal behavior of the information flow between the regions is similar for both measures. Finally, Fig. 5 shows the multivariate TE results for the same regions. We notice that, unlike DI, TE performs rather poorly and that there is not a substantial amount of difference in the TE between the two audio qualities. This is due to the fact, as indicated in Proposition 1, that TE is not calculated over all N samples and therefore contains only a fraction ($\approx \frac{1}{N}$) of the information measured by DI, in addition to the absence of the instantaneous mutual information term on the right-hand side of (18).

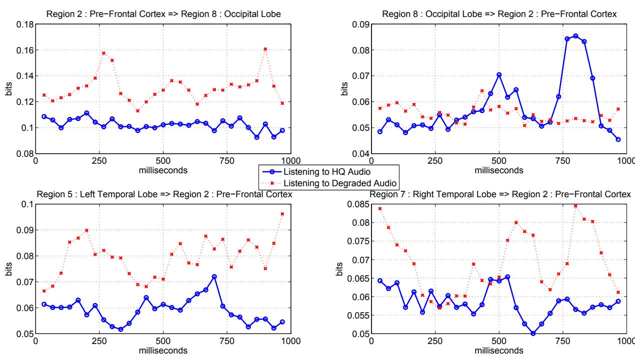

Fig. 4: Multivariate Kamitake DI

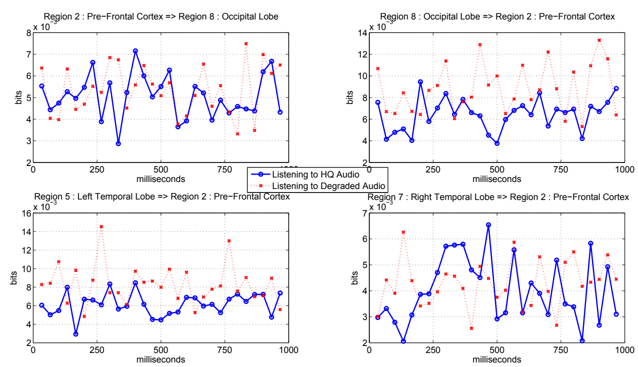

Fig. 5: Multivariate transfer entropy

VI. Conclusion

We presented a novel method to assess changes in perceived audio quality by directly measuring the EEG response of human subjects listening to time varying distorted audio. In particular, we used DI measures to analyze the casual flow of information between EEG sensors grouped into ROIs over the cortex. Further, we considered the information flow between the ROIs analogous to a MAC with feedback where several senders transmit to only one receiver. We also examined the relationship between two commonly used DI measures and TE, and compared their performance as to how well they are able to distinguish between the perceived audio quality. The results demonstrate that typically a change of the information flow between different brain regions occurs as the subject listens to different audio qualities.

References

[1] *Method for subjective assesment of intermediate quality levels of coding systems*, Question ITU-R 220/10 Recommendation ITU-R BS.1534-1, Geneva, Switzerland, 1998-2001.

[2] C. D. Creusere, J. Kroger, S. R. Siddenki, P. Davis, and J. Hardin, "Assessment of subjective audio quality from EEG brain responses using time-space-frequency analysis," in *Proceedings of the 20th European Signal Processing Conference, 2012*, Bucharest, Hungary, 2012, pp. 2704–2708.

[3] A. K. Porbadnigk, J. Antons, B. Blankertz, M. S. Treder, R. Schleicher, S. Moller, and G. Curio, "Using ERPs for assessing the (sub) conscious perception of noise," in *Proceeding of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Buenos Aires, Argentina, 2010, pp. 2690–2693.

[4] A. G. Dimitrov, A. A. Lazar, and J. D. Victor, "Information theory in neuroscience," *Journal of Computational Neuroscience*, vol. 30, no. 1, pp. 1–5, 2011.

[5] A. Borst and F. E. Theunissen, "Information theory and neural coding," *Nature Neuroscience*, vol. 2, no. 11, pp. 947–957, 1999.

[6] R. K. Kushwaha and W. J. Williams, "An information flow technique for category related evoked potentials," *IEEE Transactions on Biomedical Engineering*, vol. 39, no. 2, pp. 165–175, 1992.

[7] D. Ostwald, C. Porcaro, and A. P. Bagshaw, "An information theoretic approach to EEG–fMRI integration of visually evoked responses," *Neuroimage*, vol. 49, no. 1, pp. 498–516, 2010.

[8] J. Xu, Z. Liu, R. Liu, and Q. Yang, "Information transmission in human cerebral cortex," *Physica D: Nonlinear Phenomena*, vol. 106, pp. 363–374, 1997.

[9] J. Jeong, J. C. Gore, and B. S. Peterson, "Mutual information analysis of the EEG in patients with Alzheimer's disease," *Clinical Neurophysiology*, vol. 112, pp. 827–835, 2001.

[10] O. Sakata, T. Shiina, and Y. Saito, "Multidimensional directed information and its application," *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)*, vol. 85, no. 4, pp. 45–55, 2002.

[11] ——, "Estimation of the number of EEG sources in the human brain using multidimensional directed information analysis," *Electrical Engineering in Japan*, vol. 145, no. 4, pp. 38–45, 2003.

[12] O. Sakata and Y. Suzuki, "Visualization of information flows in a brain stemming from alpha waves: Using multidimensional directed information," in *Proceedings of the IEEE Region 10 Conference, TENCON*. IEEE, 2010, pp. 1752–1756.

[13] P. Petrantonakis and L. Hadjileontiadis, "Frontal EEG asymmetry and affective states: A multidimensional directed information approach," in *Mediterranean Conference on Medical and Biological Engineering and Computing*. Chalkidiki, Greece: Springer, 2010, pp. 687–690.

[14] M. Wibral, R. Vicente, and J. T. Lizier, *Directed information measures in neuroscience*. Heidelberg: Springer, 2014.

[15] C. D. Creusere and J. C. Hardin, "Assessing the quality of audio containing temporally varying distortions," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 711–720, 2011.

[16] K. Mehta and J. Kliewer, "Assessing subjective perception of audio quality by measuring the information flow on the brain-response channel," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Florence, Italy, May 2014, pp. 5884–5888.

[17] H. Marko, "The bidirectional communication theory–a generalization of information theory," *Communications, IEEE Transactions on*, vol. 21, no. 12, pp. 1345–1351, 1973.

[18] T. Schreiber, "Measuring information transfer," *Physical review letters*, vol. 85, no. 2, p. 461, 2000.

[19] J. Massey, "Causality, feedback and directed information," in *Proc. Int. Symp. Inf. Theory Applic.(ISITA-90)*. Citeseer, 1990, pp. 303–305.

[20] T. Kamitake, H. Harashima, and H. Miyakawa, "A time-series analysis method based on the directed transinformation," *Electronics and Communications in Japan (Part I: Communications)*, vol. 67, no. 6, pp. 1–9, 1984.

[21] G. Kramer, "Causal conditioning, directed information and the multiple-access channel with feedback," in *Information Theory, 1998. Proceedings. 1998 IEEE International Symposium on*, Aug 1998, p. 189.

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

[23] M. Al-Khassaweneh and S. Aviyente, "The relationship between two directed information measures," *IEEE Signal Processing Letters*, no. 15, pp. 801–804, 2008.

[24] V. Solo, "On causality and mutual information," in *IEEE Conference on Decision and Control*, Dec 2008, pp. 4939–4944.

[25] T. W. Picton, "The P300 wave of the human event-related potential," *Journal of Clinical Neurophysiology*, vol. 9, no. 4, pp. 456–479, 1992.