

# Directional and Causal Information Flow in EEG for Assessing Perceived Audio Quality

Ketan Mehta<sup>1</sup> and Jörg Kliewer, *Senior Member, IEEE*

**Abstract**—In this paper, electroencephalography (EEG) measurements are used to infer change in cortical functional connectivity in response to change in audio stimulus. Experiments are conducted wherein the EEG activity of human subjects is recorded as they listen to audio sequences whose quality varies with time. A causal information theoretic framework is then proposed to measure the information flow between EEG sensors appropriately grouped into different regions of interest (ROI) over the cortex. A new causal bidirectional information (CBI) measure is defined as an improvement over standard directed information measures for the purposes of identifying connectivity between ROIs in a generalized cortical network setting. CBI can be intuitively interpreted as a causal bidirectional modification of directed information, and inherently calculates the divergence of the observed data from a multiple access channel with feedback. Further, we determine the analytical relationship between the different causal measures and compare how well they are able to distinguish between the perceived audio quality. The connectivity results inferred indicate a significant change in the rate of information flow between ROIs as the subjects listen to different audio qualities, with CBI being the best in discriminating between the perceived audio quality, compared to using standard directed information measures.

**Index Terms**—Electroencephalography (EEG), directed information, causal conditioning, functional connectivity, audio quality.

## I. INTRODUCTION

**D**ETECTION and response to stimuli is in general a multistage process that results in the hierarchical activation and interaction of several different regions in the brain. To understand the dynamics of brain functioning it therefore essential to investigate the information flow and connectivity (interactions) between different regions in the brain. Further, in each of these hierarchies, sensory and motor information in the brain is represented and manipulated in the form of neural activity patterns. The superposition of this electrophysiological

Manuscript received December 27, 2016; revised July 17, 2017 and January 26, 2018; accepted February 1, 2018. Date of publication February 15, 2018; date of current version June 5, 2018. This work was supported by NSF under Grant CCF-1065603. This paper was presented in part at the 49th Asilomar Conference on Signals, Systems and Computers Pacific Grove, CA, USA, November 2015 [1], and at the IEEE International Conference on Communications Paris, France, May 2017 [2]. The associate editor coordinating the review of this paper and approving it for publication was C. Rozell. (*Corresponding author: Ketan Mehta.*)

K. Mehta is with the Klipsch School of Electrical and Computer Engineering, New Mexico State University, Las Cruces, NM 88003 USA (e-mail: ketan@nmsu.edu).

J. Kliewer is with the Helen and John C. Hartmann Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07103 USA.

Digital Object Identifier 10.1109/TMBMC.2018.2806454

activity can be recorded via electrodes on the scalp and is termed as electroencephalography (EEG).

Functional connectivity refers to the statistical dependencies between neural data recorded from spatially distinct regions in the brain [3], [4]. Information theory provides a stochastic framework which is fundamentally well suited for the task of assessing functional connectivity [5], [6] between neural responses. For example, [7], [8] presented mutual information (MI) [9] estimates to assess the correlation between the spike timing of an ensemble of neurons. Likewise, [10]–[12] investigated the effectiveness of calculating pairwise maximum entropy to model the activity of a larger population of neurons. MI has also been successfully employed in the past for determining the functional connectivity in EEG sensors for feature extraction and classification purposes [13], [14]. Similarly, other studies have used MI to analyze EEG data to investigate corticocortical information transmission for pathological conditions such as Alzheimer’s disease [15] and schizophrenia [16], or for odor stimulation [17], [18].

One limitation of MI and entropy when applied in the traditional (Shannon) sense is their inability to distinguish between the direction of information flow, as pointed out by Marko [19]. In the same work Marko also proposed to calculate the information flow in each direction of a bidirectional channel using conditional probabilities based on Markovian dependencies. Massey [20] extended the initial work by Marko and formally defined directed information as the information flow from the input to the output of a channel with feedback. Other measures have similarly been defined for calculating the directional information transfer rate between random processes, most notably, Kamitake’s directed information [21] and transfer entropy by Schreiber [22]. Further, feedback and directionality are also closely related to the notion of *causality* in information measures<sup>1</sup> [20], [23], [24]. Massey’s directed information and transfer entropy are in general referred to as *causal*, since they measure statistical dependencies between the *past* and current values of a process. We adopt this definition in this paper and therefore, causality here takes on the usual meaning of a cause occurring prior to its effect, or a stimulus occurring before response, i.e., how the past states of a system influences its present and future states [25].

Our interest here is in using EEG for assessing human perception of time-varying audio quality. We are inspired by our

<sup>1</sup>Causality and directionality are formally defined in Def. 1 and Def. 2, resp., in Section III.

recent results in [26] which uses MI to quantify the information flow over the end-to-end perceptual processing chain from audio stimulus to EEG output. One characteristic common in subjective audio testing protocols including the current state-of-the-art approach, Multi Stimulus with Hidden Anchor (MUSHRA) [27], is that they require human participants to assign a single quality-rating score to each test sequence. Such conventional testing suffers from a subject-based bias towards cultural factors in the local testing environment and can tend to be highly variable. Instead, neurophysiological measurements such as EEG directly capture and analyze the brainwave response patterns that depend only on the perceived variation in signal quality [28], [29]. As a result, EEG is inherently well suited to assess human perception of audio [30], [31] and visual [31]–[33] quality. For example, Porbadnigk *et al.* [31] and Scholler *et al.* [32] used linear discriminant analysis classifiers to extract features from EEG for classifying noise detection in audio signals and to assess changes in perceptual video quality, respectively. Similarly, [30] identified features in EEG brainwave responses corresponding to time-varying audio quality using a time-space-frequency analysis, while [33] employed a wavelet-based approach for an EEG classification of commonly occurring artifacts in compressed video, using a single-trial EEG.

To the best of our knowledge, however, the work presented here is the first time that functional connectivity has been applied in conjunction with EEG measurements for the purposes of assessing audio quality perception. By using causal information measures to detect a change in functional connectivity we directly identify those cortical regions which are most actively involved in perceiving a change in audio quality. Further, we establish the analytical relationship between the different presented information measures and compare how well each of them is able to distinguish between the perceived audio qualities. Towards this end, we consider two distinct scenarios for estimating the connectivity between EEG sensors by appropriately grouping them into regions of interest (ROIs) over the cortex. In the first scenario, we employ Massey’s and Kamitake’s directed information and transfer entropy, respectively, to calculate the pairwise directional information flow between ROIs while using causal conditioning to account for the influence from all other regions. In the second scenario we propose a novel information measure which can be considered as a causal bidirectional modification of directed information applied to a generalized cortical network setting. In particular, we show that the proposed causal bidirectional information (CBI) measure assesses the direct connectivity between any two given nodes of a multiterminal cortical network by inherently calculating the divergence of the induced conditional distributions from those associated with a multiple access channel (MAC) with feedback. Each presented measure is validated by applying it to analyze real EEG data recorded for human subjects as they listen to audio sequences whose quality changes over time. For the sake of simplicity and analytical tractability we restrict ourselves to only two levels of audio quality (high quality and degraded quality). We determine and compare the instantaneous information transfer rates as inferred by each of these measures for the case where the

TABLE I  
DIFFERENT QUALITY LEVELS PRESENTED TO THE SUBJECT DURING THE COURSE OF AN AUDIO TRIAL. TO GENERATE THE DISTORTION, EACH OF THESE BASE-SEQUENCES WERE PASSED THROUGH A 2048-POINT MODIFIED DISCRETE COSINE TRANSFORM AND EITHER FREQUENCY TRUNCATION OR SCALAR QUANTIZATION WAS APPLIED TO THE COEFFICIENTS PRIOR TO RECONSTRUCTION [34]

Quality Level	Freq. Truncation	Scalar Quantization
	Low-Pass Filter	No. of Significant Bits Retained
Q1	4.4 KHz	4
Q2	2.2 KHz	3
Q3	1.1 KHz	2

subject listens to high quality audio as opposed to the case when the subject listens to degraded quality audio. Finally, note that we are not able to make any assumptions about the actual structure of the underlying cortical channels (e.g., linear vs. non-linear) as our analysis is solely based on the observed empirical distributions of the data at the input and output of these channels.

The rest of the paper is organized as follows. Section II provides an overview of EEG, the experiment, and the stimulus audio sequences. In Section III we review some directional information measures widely used in the literature for estimating connectivity. We assess the information flow between cortical regions using directional information measures in Section IV, along with determining the analytical relationship between these measures. In Section V we introduce CBI and discuss its properties. The results of our analysis on EEG data are presented in Section VI. We finally conclude with a summary of our study and future directions in Section VII.

## II. BACKGROUND

In the conducted study, the EEG response activity of human test subjects is recorded as they listen to a variety of audio test-sequences. The quality of these stimulus test-sequences is varied with time between different “quality levels”. All audio test-sequences were created from three fundamentally different base-sequences sampled at a reference base quality of 44.1 kHz, with a precision of 16 bits per sample. Here, we employ the same test-sequences and distortion quality levels as in [34]. Two different types of distortions are considered for our analysis, scalar quantization and frequency band truncation, where the specific parameters are listed in Table I. The test-sequence for a specific trial is created by selecting one of the two distortion types and then applying it to the original base-sequence in a time-varying pattern of non-overlapping five second blocks as shown in Fig. 1. Multiple of such trials are conducted for each subject by choosing all possible combinations of sequences, distortion types, and time-varying patterns. Note that despite the subjects being presented with all different quality levels in our listening tests, here we focus exemplarily only on the “high” base-quality and the “Q3 degraded” quality audio. This addresses the worst-case quality change and keeps the problem analytically and numerically tractable. A detailed exposition of the experimental setup, test-sequences, and distortion quality levels is provided in [26].

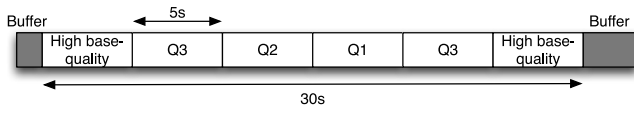


Fig. 1. A 30 second test sequence where the audio quality changes in a time-varying pattern over the whole duration of the sequence. Different possible combinations of quality changes and two distortion types are presented to each subject in a randomized fashion (adopted from [26]).

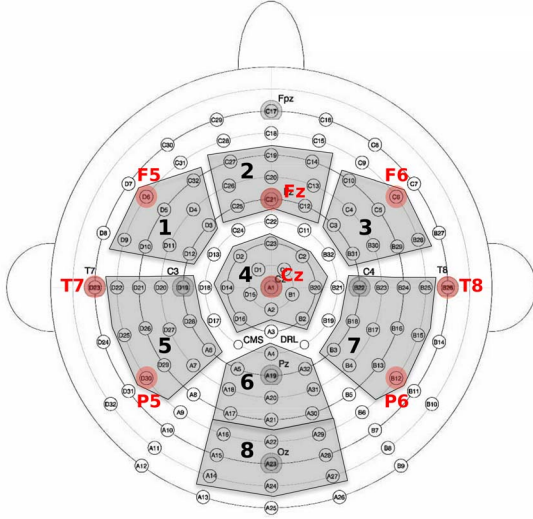


Fig. 2. The 128 electrodes are grouped into eight regions of interest (ROI) to effectively cover the different cortical regions (lobes) of the brain (adopted from [26]). The red electrode locations are used for generating synthetic EEG data in Section VI-C, see Table III for the naming convention.

The EEG data is captured on a total of 128 spatial channels using an ActiveTwo Biosemi system with a sampling rate of 256 Hz. To better manage the large amount of collected data while also effectively covering the activity over different regions of the cortex, we group the 128 electrodes into specific regions of interest (ROI) as shown in Fig. 2. While a large number of potential grouping schemes are possible, this scheme is favored for our purposes as it efficiently covers all the cortical regions (lobes) of the brain with a relatively low number of ROIs. Also, the number of electrodes in any given ROI varies between a minimum of 9 to a maximum of 12. For example, in our region partitioning scheme ROI 2 (9 electrodes) covers the prefrontal cortex, ROI 6 (10 electrodes) the parietal lobe, ROI 8 (9 electrodes) the occipital lobe, and ROI 5 and ROI 7 (12 electrodes each) cover the left and right temporal lobes, respectively. In essence, our goal here then is to investigate the causal connectivity between the different ROI in response to different audio quality levels.

### III. DIRECTIONALITY, CAUSALITY AND FEEDBACK IN INFORMATION MEASURES

Let  $X^n$  denote a random vector of  $n$  constituent discrete valued random variables  $[X_1, X_2, \dots, X_n]$ . Also, let  $x^n = [x_1, x_2, \dots, x_n]$  be the corresponding realizations drawn from the joint probability distribution denoted by  $p(x_1, x_2, \dots, x_n)$ , respectively. Similarly,  $X_{n+1}^N = [X_n, X_{n+1}, \dots, X_N]$  represents a length  $N - n + 1$  random vector. Denoting the expected value

of a random variable by  $\mathbb{E}[\cdot]$ , the entropy of the  $n$ -tuple  $X^n$  can be written as  $H(X^n) = -\mathbb{E}[\log p(X^n)]$ .

The mutual information (MI) between two length  $N$  interacting random processes  $X^N$  and  $Y^N$  is defined as

$$I(X^N; Y^N) = \sum_{n=1}^N I(X^n; Y_n | Y^{n-1}) \quad (1)$$

$$= H(Y^N) - H(Y^N | X^N), \quad (2)$$

with the conditional entropy  $H(Y^n | X^n) = -\mathbb{E}[\log p(Y^n | X^n)]$  and

$$H(Y^N | X^N) = \sum_{n=1}^N H(Y_n | Y^{n-1} X^N). \quad (3)$$

The MI measures the reduction in the uncertainty of  $Y^N$  due to the knowledge of  $X^N$ , and is zero if and only if the two processes are statistically independent. The MI between the two random processes is symmetric, i.e.,  $I(X^N; Y^N) = I(Y^N; X^N)$ , and can therefore not distinguish between the direction of the information flow. Alternatively, a directional information measure introduces the notion of direction in the exchange of information between sources. In this paper we define a directional information measure as follows.

*Definition 1:* A directional information measure from  $X^N$  to  $Y^N$ , represented with an arrow  $X^N \rightarrow Y^N$ , quantifies the information exchange rate in the direction from the input process  $X^N$  towards the output process  $Y^N$ .

As implied by the definition, the information flow measured by a directional information measure is not symmetric, and in general the flow from  $X^N \rightarrow Y^N$  is not equal to  $Y^N \rightarrow X^N$ . In the following, we will examine three different directional information measures presented in the literature.

#### A. Massey's Directed Information

Directed information as proposed by Massey [20] is an extension of the preliminary work by Marko [19] to characterize the information flow on a communication channel with feedback. Given input  $X^N$  and output  $Y^N$ , the channel is said to be used without feedback if

$$p(x_n | y^{n-1} x^{n-1}) = p(x_n | x^{n-1}), \quad \forall n \leq N, \quad (4)$$

i.e., the current channel input value does not depend on the past output samples. If the channel is used with feedback then [20] shows that the chain rule of conditional probability can be reduced to

$$p(y^N | x^N) = \prod_{n=1}^N p(y_n | y^{n-1} x^n). \quad (5)$$

In closely related work [35] introduced the concept of causal conditioning based on (5). The entropy of  $Y^N$  *causally* conditioned on  $X^N$  is defined as

$$H(Y^N || X^N) = \sum_{n=1}^N H(Y_n | Y^{n-1} X^n). \quad (6)$$

*Definition 2:* A measure between two random processes  $X^N$  and  $Y^N$  is said to be causal if the information transfer rate at time  $n$  relies only on the dependencies between their past and

current sample values  $X^n$  and  $Y^n$ , and is not a function of the future sample values  $X_{n+1}^N$  and  $Y_{n+1}^N$ .

Therefore, the notion of causality as used in this work is based on inferring the statistical dependencies between the past states of a system on its present and future states [20], [23]–[25]. This is in contrast to the stronger interventional interpretation about casual inferences such as in [36] which draws conclusions about *causation*, e.g., process  $X^N$  causes  $Y^N$ .

Massey's directed information is a causal measure between sequence  $X^N$  and  $Y^N$  defined as

$$DI_1(X^N \rightarrow Y^N) \triangleq H(Y^N) - H(Y^N | X^N) \quad (7)$$

$$= \sum_{n=1}^N I(X^n; Y_n | Y^{n-1}). \quad (8)$$

Equivalently, the directed information can also be written in terms of the Kullback-Leibler (KL) divergence as

$$DI_1(X^N \rightarrow Y^N) = \sum_{n=1}^N D_{KL}(p(y_n | y^{n-1}, x^n) || p(y_n | y^{n-1})) \quad (9)$$

$$= \sum_{n=1}^N \mathbb{E} \left[ \log \frac{p(Y_n | Y^{n-1}, X^n)}{p(Y_n | Y^{n-1})} \right]. \quad (10)$$

Also in general,

$$DI_1(X^N \rightarrow Y^N) \leq I(X^N; Y^N), \quad (11)$$

with equality if the channel is used without feedback. Directed information therefore not only gives a meaningful notation to the directivity of information, but also provides a tighter characterization than MI on the total information flow over a channel with feedback.

### B. Transfer Entropy

Schreiber [22] introduced a causal measure for the directed exchange of information between two random processes called transfer entropy. Similar to the pioneering work in [19], transfer entropy considers a bidirectional communication channel between  $X^N$  and  $Y^N$ , and measures the deviation of the observed distribution  $p(y_n | y^{n-1}, x^{n-1})$  over this channel from the Markov assumption

$$p(y_n | y^{n-1}, x^{n-1}) = p(y_n | y^{n-1}), \quad n \leq N. \quad (12)$$

In particular, transfer entropy quantifies the deviation of the l.h.s. of (12) from the r.h.s. of (12) using the KL divergence and is defined as

$$TE(X^{n-1} \rightarrow Y^n) \triangleq D_{KL}(p(y_n | y^{n-1}, x^{n-1}) || p(y_n | y^{n-1})) \quad (13)$$

$$= \mathbb{E} \left[ \log \frac{p(Y_n | Y^{n-1}, X^{n-1})}{p(Y_n | Y^{n-1})} \right] \quad (14)$$

$$= H(Y_n | Y^{n-1}) - H(Y_n | Y^{n-1}, X^{n-1}) \quad (15)$$

$$= I(Y_n; X^{n-1} | Y^{n-1}). \quad (16)$$

Correspondingly, for random processes of block length  $N$  we can then define a sum transfer entropy [37], [38] which in

effect calculates and adds the transfer entropy at every history depth  $n$ ,

$$TE^*(X^{N-1} \rightarrow Y^N) = \sum_{n=1}^N TE(X^{n-1} \rightarrow Y^n). \quad (17)$$

Another widely used and closely related measure for causal influence was developed by Granger [39]. Granger causality is a directional measure of statistical dependency based on prediction via vector auto-regression. The relationship between Granger causality and directional information measures has been previously analyzed in [40] and [41]. For the specific case of Gaussian random variables, as it is the case in our EEG scenario, transfer entropy has been shown to be equivalent to Granger causality.

### C. Kamitake's Directed Information

Another variant of directed information is defined by Kamitake [21] and given as

$$DI_2(X^N \rightarrow Y^N) \triangleq \sum_{n=1}^N I(X_n; Y_{n+1}^N | X^{n-1}, Y^n) \quad (18)$$

$$= H(Y_{n+1}^N | X^{n-1}, Y^n) - H(Y_{n+1}^N | X^n, Y^n). \quad (19)$$

We notice that this measure is different from Massey's directed information in that it measures the influence of the current sample  $X_n$  of  $X^N$  at time  $n$  on the future samples  $Y_{n+1}^N$  of  $Y^N$ . Kamitake's information measure is therefore directional, but not causal.

## IV. ASSESSING CORTICAL INFORMATION FLOW VIA DIRECTIONAL MEASURES

### A. Causal Conditioning and Indirect Influences in Multiterminal Networks

A multi-terminal network characterizes the information flow between multiple communicating nodes with several senders and receivers. Let us denote three communicating nodes  $\mathcal{X}$ ,  $\mathcal{Y}$ , and  $\mathcal{Z}$  and the corresponding random processes associated with them as  $X^N$ ,  $Y^N$ , and  $Z^N$  respectively. In our case, the information transfer over the cortex can be considered equivalent to a cortical multi-terminal network, with each ROI taking over the role of a communicating node. In the context of the cortical network the quantities  $X^N$ ,  $Y^N$ , and  $Z^N$ , then correspond to the sampled EEG signals from different ROIs. Also without any loss of generality,  $Z^N$  represents the output of multiple (and potentially all other) ROIs. Our goal here is to identify the connectivity between the processes in the cortical network. In particular, there are two distinct instances of connectivity that can arise as a result of using directional informational measures.

*Definition 3:* A direct connectivity is said to exist from node  $\mathcal{X}$  to node  $\mathcal{Y}$  if there exists a non-zero information flow via a direct path between the nodes.

*Definition 4:* An implied connectivity arises when there is no direct path between two nodes  $\mathcal{X}$  and  $\mathcal{Y}$ , but there is a non-zero information flow between the nodes because of an influence through other nodes in the network.

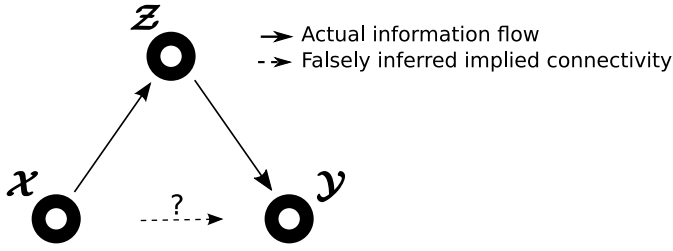


Fig. 3. The relay channel is an example of a network in which simply using directed information can lead to a false inference of *direct* connectivity between the nodes. In the example shown,  $DI(X^N \rightarrow Y^N) > 0$ , even though there is no link between  $X$  and  $Y$ .

Therefore, a positive directed information between the random processes associated with any two nodes in a multi-node network alone does not necessarily equate to a direct connectivity between them [41]–[43]. In Fig. 3 we show an example network topology to illustrate how implied connectivity can lead to false inferences. In the shown relay channel there is no direct information transfer between  $\mathcal{X}$  and  $\mathcal{Y}$ , but the information flow is from  $\mathcal{X}$  to  $\mathcal{Z}$  to  $\mathcal{Y}$ . There is a Markovian influence  $\mathcal{X} \rightarrow \mathcal{Z} \rightarrow \mathcal{Y}$ . This results in a positive value for Massey’s directed information,  $DI_1(X^N \rightarrow Y^N) > 0$ , thereby leading to implied connectivity between  $\mathcal{X}$  and  $\mathcal{Y}$ .

Notice however that in the example presented, the knowledge of  $Z^N$  leads to statistical conditional independence between  $X^N$  and  $Y^N$ . We expand upon this idea and extend the expression for Massey’s directed information to account for the influence of the additional random processes in the network via causal conditioning [35].

*Definition 5:* Causally conditioned Massey’s directed information is defined as the information flowing from  $X^N$  to  $Y^N$  causally conditioned on the sequence  $Z^{N-1}$  as

$$DI_1(X^N \rightarrow Y^N || Z^{N-1}) \triangleq H(Y^N || Z^{N-1}) - H(Y^N || X^N Z^{N-1}) \quad (20)$$

$$= \sum_{n=1}^N D_{KL}(p(y_n | x^n y^{n-1} z^{n-1}) || p(y_n | y^{n-1} z^{n-1})) \quad (21)$$

$$= \sum_{n=1}^N I(X^n; Y_n | Y^{n-1} Z^{n-1}). \quad (22)$$

*Proposition 1* [41], [42]: Assume a network as shown in Fig. 3 with three nodes  $\mathcal{X}$ ,  $\mathcal{Y}$ , and  $\mathcal{Z}$ , with corresponding random processes  $X^N$ ,  $Y^N$ , and  $Z^N$ , respectively. Using causally conditioned Massey’s directed information eliminates implied connectivity, with  $DI_1(X^N \rightarrow Y^N || Z^{N-1}) = 0$  if  $\mathcal{X}$  and  $\mathcal{Y}$  are not directly connected.

Both transfer entropy and Kamitake’s directed information can be extended to incorporate causal conditioning from an additional random process as well.

*Definition 6* [44]–[46]: Causally conditioned transfer entropy and sum transfer entropy are defined respectively as

$$TE(X^{n-1} \rightarrow Y^n || Z^{n-1}) \triangleq D_{KL}(p(y_n | y^{n-1} x^{n-1} z^{n-1}) || p(y_n | y^{n-1} z^{n-1})) \quad (23)$$

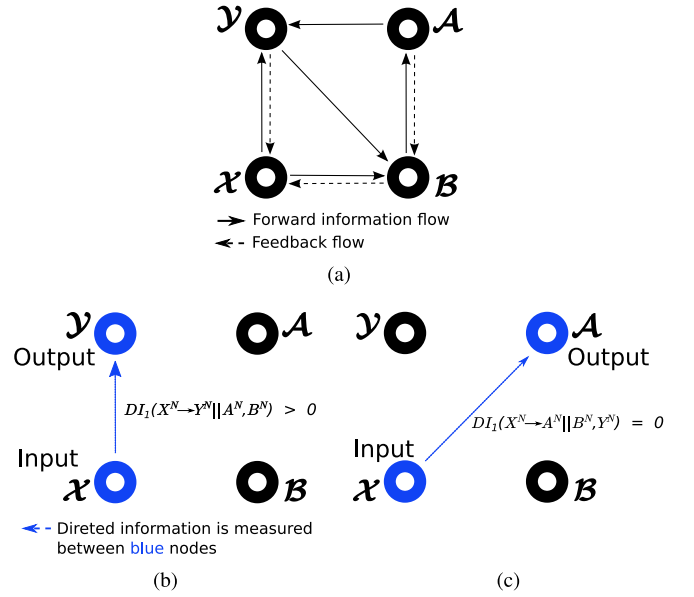


Fig. 4. (a) An example of a multiterminal network with information flow existing between several nodes, for which we estimate the underlying causal connectivity. The solid arrows denote the direction of the forward information flow, while the dashed arrows represent feedback flow between the nodes. (b), (c) Pairwise conditional directed information calculated by choosing an input and output node, while using causal conditioning to account for the influence of all other nodes.

$$= I(Y_n; X^{n-1} | Y^{n-1} Z^{n-1}), \quad (24)$$

$$TE^*(X^{N-1} \rightarrow Y^N || Z^{N-1}) \triangleq \sum_{n=1}^N TE(X^{n-1} \rightarrow Y^n || Z^{n-1}). \quad (25)$$

*Definition 7:* Causally conditioned Kamitake’s directed information is defined as

$$DI_2(X^N \rightarrow Y^N || Z^{N-1}) \triangleq \sum_{n=1}^N I(X_n; Y_{n+1}^N | X^{n-1} Y^n Z^{n-1}). \quad (26)$$

Similar to Massey’s directed information, causally conditioned transfer entropy and Kamitake’s directed information can also be used to eliminate false inferences resulting from implied connectivity [46], [47]. The proof follows along the exact same lines as Proposition 1 and is omitted here. Also, note that despite the conditioning on a causal sequence  $Z^{n-1}$ , Kamitake’s directed information measure is not strictly causal due to the  $Y_{n+1}^N$  term.

Now we discuss how to apply causally conditioned directional information measures in order to estimate the functional connectivity between the ROI network. In Fig. 4(a) we show an example communication network with four interacting nodes. The information flow is depicted using solid arrows and the feedback using dashed arrows, respectively. Also, some nodes do not have a direct link in between them, for example  $\mathcal{X}$  to  $\mathcal{A}$ . In our cortical network model the nodes represent the ROIs, and the quantities  $X^N$ ,  $Y^N$ ,  $A^N$  and  $B^N$ , resp., represent the random processes describing the sampled output of the EEG signals in each ROI. Our goal is then to infer the causal

connectivity between the ROIs given the EEG recordings from each region. Towards this end we propose calculating the pairwise conditional directed information by choosing an input and output node, while using causal conditioning to account for the influence of all other nodes. For example, to estimate the connectivity from  $\mathcal{X}$  to  $\mathcal{Y}$  we calculate  $DI_1(X^N \rightarrow Y^N || A^N B^N)$  as shown in Fig. 4(b). Since there is non-zero information flow between these two nodes we expect the directed information to return a positive value. Similarly in Fig. 4(c), computing  $DI_1(X^N \rightarrow A^N || Y^N B^N)$  yields a zero value since there is no direct connectivity between these two nodes. Repeating this procedure pairwise for all nodes provides a functional connectivity graph representative of the directional information flow over the entire ROI network.

### B. Relationship Between Different Measures

The first relation that we are interested in is between Massey's and Kamitake's directed information measures as examined in [48] and [49]. We extend this analysis to include causal conditioning on  $Z^{N-1}$  and show its connection to causally conditioned MI.

*Proposition 2:* The relation between causally conditioned Massey's and causally conditioned Kamitake's directed information is given by

$$\begin{aligned} & DI_2(Y^N \rightarrow X^N || Z^{N-1}) + DI_1(X^N \rightarrow Y^N || Z^{N-1}) \\ &= DI_2(X^N \rightarrow Y^N || Z^{N-1}) + DI_1(Y^N \rightarrow X^N || Z^{N-1}). \end{aligned} \quad (27)$$

*Proof:* Using the definition in (26) and rewriting in terms of entropy yields

$$\begin{aligned} & DI_2(Y^N \rightarrow X^N || Z^{N-1}) \\ &= \sum_{n=1}^N I(Y_n; X_{n+1}^N | X^n Y^{n-1} Z^{n-1}) \end{aligned} \quad (28)$$

$$\begin{aligned} &= \sum_{n=1}^N \left[ H(X_{n+1}^N | X^n Y^{n-1} Z^{n-1}) - H(X_{n+1}^N | X^n Y^n Z^{n-1}) \right] \end{aligned} \quad (29)$$

$$\begin{aligned} &= \sum_{n=1}^N \left[ H(X^N | Y^{n-1} Z^{n-1}) - H(X^n | Y^{n-1} Z^{n-1}) \right. \\ &\quad \left. - H(X^N | Y^n Z^{n-1}) + H(X^n | Y^n Z^{n-1}) \right] \end{aligned} \quad (30)$$

$$\begin{aligned} &= \sum_{n=1}^N I(X^N; Y_n | Y^{n-1} Z^{n-1}) - DI_1(X^N \rightarrow Y^N || Z^{N-1}), \end{aligned} \quad (31)$$

where we have used the chain rule of entropy in (30). Rearranging yields

$$\begin{aligned} & DI_2(Y^N \rightarrow X^N || Z^{N-1}) + DI_1(X^N \rightarrow Y^N || Z^{N-1}) \\ &= \sum_{n=1}^N I(X^N; Y_n | Y^{n-1} Z^{n-1}) \end{aligned} \quad (32)$$

$$\begin{aligned} &= \sum_{n=1}^N \sum_{i=1}^N I(X_i; Y_n | X^{i-1} Y^{n-1} Z^{n-1}) \end{aligned} \quad (33)$$

$$\begin{aligned} &= \sum_{n=1}^N I(Y^N; X_n | X^{n-1} Z^{n-1}) \\ &= DI_2(X^N \rightarrow Y^N || Z^{N-1}) + DI_1(Y^N \rightarrow X^N || Z^{N-1}), \end{aligned} \quad (34)$$

where (33) follows by using the chain rule of MI, and (34) from interchanging the order of summation. ■

*Definition 8:* Causally conditioned MI is defined as the MI between  $X^N$  and  $Y^N$  causally conditioned on the sequence  $Z^{N-1}$  as

$$\begin{aligned} & I(X^N; Y^N || Z^{N-1}) \triangleq \sum_{n=1}^N I(X^N; Y_n | Y^{n-1} Z^{n-1}) \\ &= \sum_{n=1}^N I(X_n; Y^N | X^{n-1} Z^{n-1}). \end{aligned} \quad (36)$$

The following corollary then expresses causally conditioned MI in terms of directed information and follows directly from (32)-(34) in the proof of Proposition 2.

*Corollary 1:* Causally conditioned MI is the sum of causally conditioned Massey's directed information from  $X^N$  to  $Y^N$ , and causally conditioned Kamitake's directed information in the opposite direction:

$$\begin{aligned} & I(X^N; Y^N || Z^{N-1}) \\ &= DI_1(X^N \rightarrow Y^N || Z^{N-1}) + DI_2(Y^N \rightarrow X^N || Z^{N-1}). \end{aligned} \quad (37)$$

There also exists a connection between Massey's directed information and transfer entropy as shown in [38], [50], and [51], which we extend and state for causal conditioning as follows.

*Proposition 3:* The relation between causally conditioned Massey's directed information and causally conditioned transfer entropy is given by

$$\begin{aligned} & DI_1(X^N \rightarrow Y^N || Z^{N-1}) = \sum_{n=1}^N \left\{ TE(X^{n-1} \rightarrow Y^n || Z^{n-1}) \right. \\ &\quad \left. + I(X_n; Y_n | X^{n-1} Y^{n-1} Z^{n-1}) \right\}. \end{aligned} \quad (38)$$

We observe that the directed information is a sum over the transfer entropy and an additional term describing the conditional undirected MI between  $X_n$  and  $Y_n$ . Massey's directed information as defined in Section III-A was originally intended to measure the dependency between two  $N$  length sequences. If we now relax this constraint to instead suitably measure the flow from a  $N-1$  length sequence  $X^{N-1}$  to a  $N$  length sequence  $Y^N$ , while causally conditioned on  $Z^{N-1}$ , we then have a modified interpretation for causally conditioned Massey's directed information which we denote as  $DI'_1$  and define as follows,

$$DI'_1(X^{N-1} \rightarrow Y^N || Z^{N-1}) = TE^*(X^{N-1} \rightarrow Y^N || Z^{N-1}), \quad (39)$$

where the equality in (39) follows directly from (24). Further, if we assume the sequences to be stationary and infinitely long and that the limit for  $N \rightarrow \infty$  exists, then asymptotically it can be shown [38], [51] that the information rates for causally

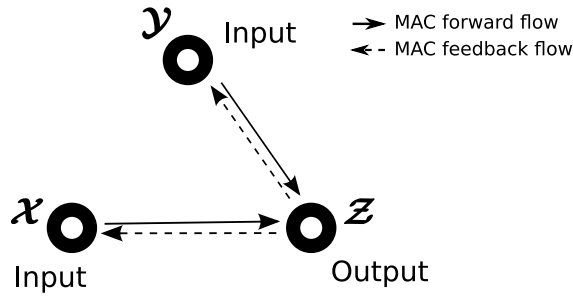


Fig. 5. A multiple access channel with feedback.

conditioned Massey's modified directed information  $DI'_1$  and causally conditioned transfer entropy (23) are in fact equal.

## V. CAUSAL BIDIRECTIONAL INFORMATION FLOW IN EEG

In the following we propose an alternative bidirectional measure for estimating the causal dependency between the ROIs. This measure is motivated by the analysis of the three-terminal multiple access channel (MAC) with feedback, an important canonical building structure in networked communication.

### A. Causal Bidirectional Information (CBI)

In order to derive the CBI measure let us first consider a preliminary result which uses causally conditioned directed information to express the information rate for such a MAC with feedback. Fig. 5 shows a two user MAC with feedback, with channel inputs  $\mathcal{X}$  and  $\mathcal{Y}$ , and corresponding output  $\mathcal{Z}$ . The capacity region for the two user discrete memoryless MAC with feedback can be lower bounded using directed information, in a form similar to the standard cut-set bound for the MAC without feedback [9]. The information rate from  $X^N$  to  $Z^N$  is shown to be [35],

$$R_1 \leq \frac{1}{N} DI(X^N \rightarrow Z^N || Y^N) \text{ for all } p(x_n y_n | x^{n-1} y^{n-1} z^{n-1}) = p(x_n | x^{n-1} z^{n-1}) \cdot p(y_n | y^{n-1} z^{n-1}), \quad n \leq N. \quad (40)$$

The other rate  $R_2$  from  $\mathcal{Y}$  to  $\mathcal{Z}$  and the sum rate  $R_1 + R_2$  can be found in [35] and are not of interest for the following discussion.

Now, consider the scenario of a general multi-terminal network as shown in Fig. 6(a), where each node sends information and receives feedback from every other node in the network. CBI considers the three node network of Fig. 6(a) and measures the information flow in between nodes  $\mathcal{X}$  and  $\mathcal{Y}$  by using the MAC as a reference, as shown in Fig. 6(b). In particular, we point our attention towards the relation in (40) specifying the relation between the joint distribution of the two inputs of the MAC. The conditional independence of the inputs in the factorization of (40) arises due to the causal nature of the feedback structure, where the output at the receiver  $Z^{n-1}$  is available causally at both  $\mathcal{X}$  and  $\mathcal{Y}$  for each time  $n$ . For the case of a MAC with feedback there is no direct connectivity (path) between the inputs. Any violation of (40) creates

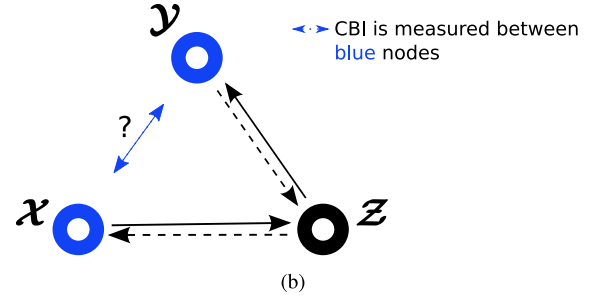
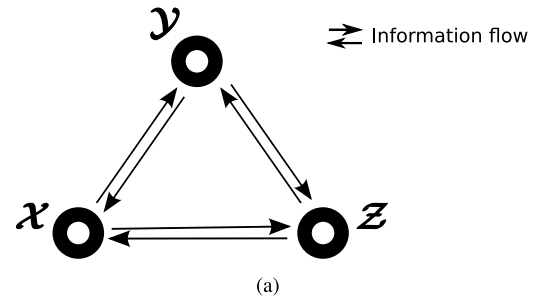


Fig. 6. (a) A general multi-terminal network whose connectivity we are interested in. All three communicating nodes send and receive information from each other. (b) CBI infers the connectivity between  $\mathcal{X}$  and  $\mathcal{Y}$  by calculating the divergence of the observed joint distribution on the network of Fig. 6(a) from the one of a MAC with feedback in (40).

dependencies between  $\mathcal{X}$  and  $\mathcal{Y}$ , and these dependencies can be measured by the KL divergence between the joint distribution on the l.h.s. of (40) and the factorization on the r.h.s. of (40). This result is summarized in the following definition.

*Definition 9:* Consider a multi-terminal network with a source node  $\mathcal{X}$ , a destination node  $\mathcal{Y}$ , and a group of nodes  $\mathcal{Z}$  interacting causally with  $\mathcal{X}$  and  $\mathcal{Y}$  as shown in Fig. 6(a). CBI calculates the KL divergence between the induced distributions of the observed conditional distribution  $p(x_n y_n | x^{n-1} y^{n-1} z^{n-1})$  and the one for an underlying MAC with feedback (40):

$$I(X^N \rightleftharpoons Y^N || Z^{N-1}) = D_{KL} \left( p(x_n y_n | x^{n-1} y^{n-1} z^{n-1}) || p(x_n | x^{n-1} z^{n-1}) p(y_n | y^{n-1} z^{n-1}) \right) \quad (41)$$

$$= \sum_{n=1}^N \mathbb{E} \left[ \log \frac{p(X_n Y_n | X^{n-1} Y^{n-1} Z^{n-1})}{p(X_n | X^{n-1} Z^{n-1}) p(Y_n | Y^{n-1} Z^{n-1})} \right] \quad (42)$$

$$= H(X^N || Z^{N-1}) + H(Y^N || Z^{N-1}) - H(X^N Y^N || Z^{N-1}). \quad (43)$$

Therefore, CBI ascertains the *direct connectivity* (as per Def. 3) between two nodes in a general multi-node network and is zero if and only if the following two conditions are satisfied:

- (i)  $X^n$  and  $Y^n$  are independent for all  $n \leq N$ , i.e., there is no information flow between the two nodes.
- (ii) There is no direct link between  $\mathcal{X}$  and  $\mathcal{Y}$  and all information flows only via an additional node  $\mathcal{Z}$ .

In the following proposition we show that, by definition, CBI is inherently a causal bidirectional modification of Massey's directed information.

*Proposition 4:* Causal bidirectional information (CBI) is the sum of causally conditioned Massey's directed information between  $X^N$  and  $Y^N$ , and the sum transfer entropy in the reverse direction:

$$I(X^N \rightleftharpoons Y^N || Z^{N-1}) \triangleq DI_1(X^N \rightarrow Y^N || Z^{N-1}) + \sum_{n=1}^N \{TE(Y^n \rightarrow X^n || Z^{n-1})\}. \quad (44)$$

*Proof:* We start with

$$\begin{aligned} DI_1(X^N \rightarrow Y^N || Z^{N-1}) + TE^*(Y^{N-1} \rightarrow X^N || Z^{N-1}) \\ = \sum_{n=1}^N \{I(Y_n; X^n | Y^{n-1} Z^{n-1}) + TE(Y^{n-1} \rightarrow X^n || Z^{n-1})\}, \end{aligned} \quad (45)$$

where the equality in (45) follows from (22) in Definition 5. Denoting the term inside the summation on the r.h.s. of (45) by  $\Phi(\cdot)$  and rewriting yields

$$\begin{aligned} \Phi(x^n, y^n, z^{n-1}) \\ = I(Y_n; X^n | Y^{n-1} Z^{n-1}) + TE(Y^{n-1} \rightarrow X^n || Z^{n-1}) \end{aligned} \quad (46)$$

$$\begin{aligned} = \mathbb{E} \left[ \log \frac{p(Y_n | X^n Y^{n-1} Z^{n-1})}{p(Y_n | Y^{n-1} Z^{n-1})} \right] \\ + \mathbb{E} \left[ \log \frac{p(X_n | Y^{n-1} X^{n-1} Z^{n-1})}{p(X_n | X^{n-1} Z^{n-1})} \right] \end{aligned} \quad (47)$$

$$= \mathbb{E} \left[ \log \frac{p(Y_n X_n | Y^{n-1} X^{n-1} Z^{n-1})}{p(Y_n | Y^{n-1} Z^{n-1}) p(X_n | X^{n-1} Z^{n-1})} \right], \quad (48)$$

where in (47) we have made use of (21) and (23), respectively, and (48) follows from the chain rule of joint probability

$$\begin{aligned} p(y_n x_n | x^{n-1} y^{n-1} z^{n-1}) \\ = p(y_n | x_n x^{n-1} y^{n-1} z^{n-1}) \cdot p(x_n | x^{n-1} y^{n-1} z^{n-1}). \end{aligned} \quad (49)$$

Taking the summation on (48) and comparing with (42) proves the claim.  $\blacksquare$

*Corollary 2:* CBI is a symmetric measure

$$\begin{aligned} I(X^N \rightleftharpoons Y^N || Z^{N-1}) \\ = DI_1(X^N \rightarrow Y^N || Z^{N-1}) + TE^*(Y^{N-1} \rightarrow X^N || Z^{N-1}) \end{aligned} \quad (50)$$

$$= DI_1(Y^N \rightarrow X^N || Z^{N-1}) + TE^*(X^{N-1} \rightarrow Y^N || Z^{N-1}). \quad (51)$$

*Proof:* Without any loss of generality, the joint probability distribution in (49) can alternatively be expanded as

$$\begin{aligned} p(y_n x_n | y^{n-1} x^{n-1} z^{n-1}) \\ = p(x_n | y_n y^{n-1} x^{n-1} z^{n-1}) \cdot p(y_n | y^{n-1} x^{n-1} z^{n-1}). \end{aligned} \quad (52)$$

Using the above equation and rewriting (48) then gives us

$$\Phi(x^n, y^n, z^{n-1})$$

$$= \mathbb{E} \left[ \log \frac{p(Y_n X_n | Y^{n-1} X^{n-1} Z^{n-1})}{p(Y_n | Y^{n-1} Z^{n-1}) p(X_n | X^{n-1} Z^{n-1})} \right] \quad (53)$$

$$\begin{aligned} = \mathbb{E} \left[ \log \frac{p(X_n | Y^n X^{n-1} Z^{n-1})}{p(X_n | X^{n-1} Z^{n-1})} \right] \\ + \mathbb{E} \left[ \log \frac{p(Y_n | Y^{n-1} X^{n-1} Z^{n-1})}{p(Y_n | Y^{n-1} Z^{n-1})} \right] \end{aligned} \quad (54)$$

$$= I(X_n; Y^n | X^{n-1} Z^{n-1}) + TE(Y^{n-1} \rightarrow Y^n || Z^{n-1}). \quad (55)$$

Taking the sum on both sides of (55) yields the required result.  $\blacksquare$

We also evaluate the expression for CBI by comparing it to conditional MI. Conditional MI measures the divergence between the actual observations and those which would be observed under the Markovian assumption  $X^N \leftrightarrow Z^{N-1} \leftrightarrow Y^N$ ,

$$\begin{aligned} I(X^N; Y^N | Z^{N-1}) \\ = \mathbb{E} \left[ \log \frac{p(X^N Y^N | Z^{N-1})}{p(X^N | Z^{N-1}) p(Y^N | Z^{N-1})} \right] \end{aligned} \quad (56)$$

$$\begin{aligned} = \sum_{n=1}^N \mathbb{E} \left[ \log \frac{p(X_n Y_n | X^{n-1} Y^{n-1} Z^{n-1})}{p(X_n | X^{n-1} Z^{n-1}) p(Y_n | Y^{n-1} Z^{n-1})} \right] \quad (57) \\ = H(X^N | Z^{N-1}) + H(Y^N | Z^{N-1}) - H(X^N Y^N | Z^{N-1}), \end{aligned} \quad (58)$$

where (57) follows from the chain rule of probability. Conditional MI is zero if and only if  $X^N$  and  $Y^N$  are conditionally independent given  $Z^{N-1}$ . By comparing conditional mutual information (57) with the expression for CBI in (42), we notice that CBI uses causal conditioning as  $Z^{N-1}$  is replaced with  $Z^{n-1}$ .

## VI. INFERRING CHANGE IN FUNCTIONAL CONNECTIVITY

### A. Preliminaries

In our analysis, we separately extract the EEG response sections for each of the two audio quality levels and calculate the information measures individually for each of them. This allows us to compare how the different probability distributions used in each of the presented information measures affect the ability to detect a change in information flow among the ROIs in Fig. 2, between the cases when the subjects listen to high quality audio as opposed to degraded quality audio.

We begin by selecting a source and a destination ROI,  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively. The other six remaining ROIs are considered to represent the side information  $\mathcal{Z}$ . Since all electrodes in a ROI are located within close proximity of one another and capture data over the same cortical region, we consider every electrode in an ROI as an independent realization of the same random process. For example, the sampled EEG data recorded on every electrode within region  $\mathcal{X}$ , in a given time interval  $N$ , is considered a realization of the same random process  $X^N$ . This increases the sample size for the process  $X^N$ , reducing the expected deviation between the obtained empirical distribution of  $X^N$  and the true one. The discussed information



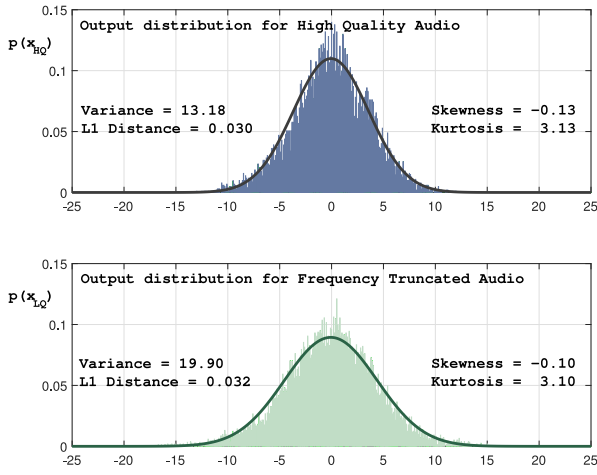


Fig. 7. Sampled EEG output distributions over a single ROI for high quality and frequency distorted audio input-stimulus, respectively. The data used to construct the histogram is the combined output over all sensors in ROI 5 during the period of a single frequency-truncated trial, for Subject S1. The Gaussian fit is obtained by using an estimator that minimizes the  $L_1$  distance between the fitted Gaussian distribution and the histogram data.

measures are therefore calculated between the ROI pairs (and not between individual electrodes) for a total of 2-permutations of 8 ROIs, i.e., 56 combinations of source-destination pairs.

In our earlier work [26] we demonstrated that the output EEG response to the audio quality, over an ROI, 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. Fig. 7 shows the histogram of the sampled EEG data, formed by concatenating the output over all sensors in one ROI for a single subject. The sample skewness and kurtosis of the EEG output distribution are shown in Table II. For a Gaussian distribution the skewness equals 0 and the kurtosis equals 3 [52], [53]. To test Gaussianity of a large sample set, the sample skewness and kurtosis should approach these values, while an absolute skew value larger than 2 [53] or kurtosis larger than 7 [54] may be used as reference values for determining substantial non-normality. By inspecting the sample estimates in Table II and comparing the histogram to a Gaussian distribution in Fig. 7, we observe that the EEG output distribution is indeed strongly Gaussian.

Knowing that the interacting random processes from the ROIs converge to a Gaussian distribution [26] allows us to formulate analytical closed-form expressions for calculating the information measures. The joint entropy of a  $n$ -dimensional multivariate Gaussian distribution with probability density  $p(z_1 \dots z_n)$  is known to be given by [9]

$$H(Z_1 \dots Z_n) = \frac{1}{2} \log(2\pi e)^n |\mathbf{C}(Z_1 \dots Z_n)|, \quad (59)$$

where  $\mathbf{C}(\cdot)$  is the covariance matrix and  $|\cdot|$  is the determinant of the matrix.

If  $X^N$ ,  $Y^N$ , and  $Z^N$  are jointly Gaussian distributed, then using (59) in conjunction with (43) reduces CBI to a function

TABLE II  
SAMPLE SKEWNESS AND KURTOSIS ESTIMATED FOR A SINGLE SUBJECT USING DATA FROM A SINGLE FREQUENCY-TRUNCATED TRIAL. THE SAMPLED EEG DATA FOR A GIVEN ROI WAS OBTAINED BY CONCATENATING THE OUTPUT OF ALL ELECTRODES IN THAT ROI

ROI	HQ			LQ		
	Mean	Skewness	Kurtosis	Mean	Skewness	Kurtosis
1	-0.003	0.180	2.820	-0.014	0.187	2.816
2	0.148	0.057	2.943	0.159	0.214	2.786
3	0.027	0.069	2.931	0.014	0.054	2.945
4	0.216	-0.154	3.155	0.2975	0.131	2.869
5	-0.058	-0.132	3.133	-0.073	-0.103	3.103
6	-0.305	-0.145	3.146	-0.283	-0.259	3.260
7	-0.111	0.016	2.984	-0.101	0.050	2.950
8	0.465	-0.070	3.071	0.445	-0.342	3.342

of their joint covariance matrices,

$$\begin{aligned} I(X^N \rightleftharpoons Y^N || Z^{N-1}) &= \frac{1}{2} \sum_{n=1}^N \log \left\{ \frac{|\mathbf{C}(X^{n-1} Y^{n-1} Z^{n-1})|}{|\mathbf{C}(X^n Y^n Z^{n-1})|} \right. \\ &\quad \left. \times \frac{|\mathbf{C}(X^n Z^{n-1})| \cdot |\mathbf{C}(Y^n Z^{n-1})|}{|\mathbf{C}(X^{n-1} Z^{n-1})| \cdot |\mathbf{C}(Y^{n-1} Z^{n-1})|} \right\}. \end{aligned} \quad (60)$$

In a similar manner, causally conditioned Massey's directed information, Kamitake's directed information, and sum transfer entropy, resp., can be reduced to obtain the following expressions,

$$\begin{aligned} DI_1(X^N \rightarrow Y^N || Z^{N-1}) &= \frac{1}{2} \sum_{n=1}^N \log \frac{|\mathbf{C}(Y^n Z^{n-1})| \cdot |\mathbf{C}(X^n Y^{n-1} Z^{n-1})|}{|\mathbf{C}(Y^{n-1} Z^{n-1})| \cdot |\mathbf{C}(X^n Y^n Z^{n-1})|}, \end{aligned} \quad (61)$$

$$\begin{aligned} DI_2(X^N \rightarrow Y^N || Z^{N-1}) &= \frac{1}{2} \sum_{n=1}^N \log \frac{|\mathbf{C}(X^{n-1} Y^n Z^{n-1})| \cdot |\mathbf{C}(X^n Y^n Z^{n-1})|}{|\mathbf{C}(X^{n-1} Y^n Z^{n-1})| \cdot |\mathbf{C}(X^n Y^n Z^{n-1})|}, \end{aligned} \quad (62)$$

$$\begin{aligned} TE^*(X^{N-1} \rightarrow Y^N || Z^{N-1}) &= \frac{1}{2} \sum_{n=1}^N \log \frac{|\mathbf{C}(Y^n Z^{n-1})| \cdot |\mathbf{C}(Y^{n-1} X^{n-1} Z^{n-1})|}{|\mathbf{C}(Y^{n-1} Z^{n-1})| \cdot |\mathbf{C}(Y^n X^{n-1} Z^{n-1})|}. \end{aligned} \quad (63)$$

## B. Receiver Operating Characteristic (ROC) Curves

In order to evaluate the accuracy with which an information measure and in particular CBI can distinguish between the perceived audio quality we conduct a receiver operating characteristic (ROC) curve analysis [55], [56] on the generated vectors of measurements for the high and degraded quality audio, respectively. The ROC curve serves as a non-parametric statistical test to compare different information rates [57]–[60] and has the advantage that a test statistics can be generated from the observed measurements.

Consider a general binary classification scheme between two classes  $\mathcal{P}$  and  $\mathcal{N}$ , labeled as *positive* and *negative*, respectively. These classes contain samples from the respective class distributions with  $|\mathcal{P}| = N_p$  and  $|\mathcal{N}| = N_n$ . Also, every individual instance from the positive class is associated with a known measurement score  $S_i^{(p)}$  which is a random variable with an unknown distribution and corresponding values  $s_i^{(p)}$ ,  $i = 1, \dots, N_p$ . Similarly, for the negative class we have the random variable  $S_i^{(n)}$  with values  $s_i^{(n)}$ ,  $i = 1, \dots, N_n$ .

For a given discrimination threshold value  $T$ , the classification rule is such that the individual is allocated to the positive class if its score exceeds the threshold, i.e.,  $s_i^{(p)} > T$ . Else,  $s_i^{(p)} \leq T$ , then the instance is allocated to the negative class. The true positive probability [56] is the probability that an individual from the positive class is correctly classified as belonging to the positive class:

$$P_{ip}(T) = \frac{\sum_{i=1}^{N_p} \mathbb{1}(s_i^{(p)} > T)}{N_p}, \quad (64)$$

where  $\mathbb{1}(\cdot)$  denotes the indicator function. Likewise, the false positive probability is the probability that an individual from the negative population is misclassified (incorrectly allocated) as belonging to the positive population:

$$P_{fp}(T) = \frac{\sum_{i=1}^{N_n} \mathbb{1}(s_i^{(n)} > T)}{N_n}. \quad (65)$$

Since in general the best choice for  $T$  is not known, the measurement scores themselves often serve as optimal choices for the discrimination threshold values [55]. The ROC curve is then a graphical representation obtained by plotting the true positive probability  $P_{ip}(T)$  versus the false positive probability  $P_{fp}(T)$  as a function of the discrimination threshold value  $T$ .

The area under the ROC curve, denoted using the variable  $\theta$ , summarizes the results over all possible values of the threshold  $T$  [55], [61], with possible values ranging from  $\theta = 0.5$  (no discriminative ability) to  $\theta = 1.0$  (perfect discriminative ability). In fact, the area under the curve is shown to be equivalent to the Wilcoxon-Mann-Whitney non-parametric test-statistic [52] and can be used to determine whether samples selected from the two class populations have the same distribution [62], [63]. More specifically, the area under the ROC curve is equivalent to the probability that a randomly selected instance from the positive class will have a measurement score that is greater than a randomly selected instance from the negative class.

### C. Performance of Information Measures on Simulated EEG Data

To obtain an initial assessment of the information measures introduced in Sections IV and V, we apply these measures to synthetic EEG data. The advantage of this approach is that the true causal connectivity graph of the simulated network is known and that, in contrast to real EEG data obtained from human subjects, in principle there is no limitation to the number of trials. The EEG data is created using a wavelet transform following the approach described in [64], wherein the EEG

TABLE III  
WAVELET COEFFICIENT SCALE FACTOR FOR HIGH AND LOW ACTIVITY FOR EACH SOURCE-SINK PAIR

Source	Sink	Freq. (Hz)	High	Low
Fz	Cz	0 - 3.91 (delta)	$\mathcal{U}(0.9, 1)$	$\mathcal{U}(0.55, 0.85)$
F5, F6	P5, P6	3.91 - 7.81 (theta)	$\mathcal{U}(0.9, 1)$	$\mathcal{U}(0.55, 0.85)$
P6, F5	Cz	7.81 - 15.62 (alpha)	1	1
T7	T8	15.62 - 31.25 (beta)	1	1

is decomposed as a convolution of a series of basis functions (i.e., wavelets) within selected frequency bands listed in Table III. In [64] the probability distribution of the wavelet coefficients for each frequency band was estimated using real human EEG data sampled at 250Hz, which was shown to be a logistic distribution with heavier tails than a Gaussian distribution. We now generate simulated wavelet coefficients for a particular frequency band by multiplying the associated logistic distribution with a constant scaling factor and then randomly drawing samples from this scaled distribution. The simulated EEG signal is then obtained by an inverse wavelet transform of these randomly drawn samples for all frequency bands in Table III. In this work, two different sets of scaling factors are used to simulate change in network connectivity. Since the magnitude of the scaling coefficient controls the spectral energy (activity) in that band, we here use the activity labels “high” and “low”, respectively, to differentiate between these two sets of connectivities as shown in Table III. For simulation purposes, the wavelet-coefficients for high activity in the delta band (0-3.91 Hz) and theta band (3.91-7.81 Hz) are scaled by a factor chosen from a distribution  $\mathcal{U}(0.9, 1)$ , where  $\mathcal{U}(a, b)$  denotes the uniform distribution within the interval  $(a, b)$ . Low activity is simulated by choosing the scaling factors for the delta and theta bands from  $\mathcal{U}(0.55, 0.85)$ . Note that the uniform distribution and the associated intervals are only chosen to demonstrate a change in delta and theta activity levels, and in theory any suitable set of high-valued and low-valued scaling factors can potentially be used to this effect.

A source and sink electrode location was assigned to each frequency band [64] as shown in Table III. The locations of these electrodes are marked in red in Fig. 2. Simulated EEG signals are generated across the scalp by spherical spline interpolation across neighboring electrodes by using the *eeg\_interp* function in EEGLAB [65]. In essence this function simulates EEG scalp topography by smearing electric potentials through volume conduction [66]. This results in a synthetic EEG waveform at each electrode with energy within each of the four characteristic frequency bands.

We generate 1000 trials of simulated EEG data for each of the two cases of high and low activity. For each trial, a total of 430500 samples (at 250 Hz) are generated for each

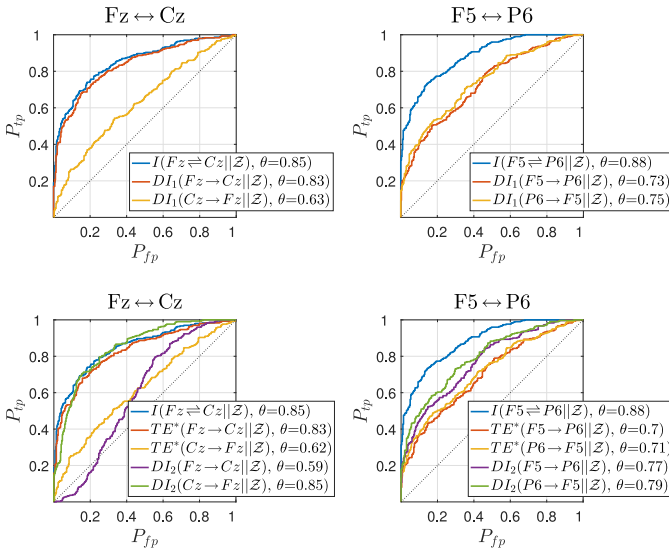


Fig. 8. ROC curves classifying the change in connectivity between high and low activity for the synthetic EEG model.

of the electrodes listed in Table III. The EEG output generated at each of these electrodes is tested for Gaussianity similar to the process described in Section VI-A, including using an estimator that minimizes the  $L_1$  distance between the fitted Gaussian distribution and the generated data, and indeed verified to be near Gaussian. For each trial of high and low activity the source electrode (of a given frequency band for which the activity changed) is designated as  $\mathcal{X}$ , the corresponding destination electrode as  $\mathcal{Y}$ , and all other remaining electrodes in Table III are grouped together as  $\mathcal{Z}$ . The sample space for each trial is obtained by dividing the 430,500 samples into 14,350 sections each of block length  $N = 30$  (120 ms). We then compute the different information measures (60)-(63) over the block length  $N$  and compare how the different probability distributions used in each measure effect the ability to detect change in connectivity. The corresponding covariance matrices,  $\forall n = 1, \dots, N$ , are estimated from the 14,350 sections, assuming stationarity of the EEG signals within each section. While here we always sum up to a depth of  $N = 30$ , we note that the estimate accuracy of the information measures, the computational time and complexity, and the probability of a false negative can all be further improved by rigorously selecting optimal history lengths [38], [67].

The ROC curve analysis is used to test the accuracy with which these information measures can distinguish between the cases of high and low activity. The positive class  $\mathcal{P}$  is constructed by concatenating the information rates for the high activity, and likewise the negative class  $\mathcal{N}$  for low activity. The ROC curve for the information measure is then constructed using the empirical probabilities  $P_{tp}(T)$  and  $P_{fp}(T)$  calculated according to (64) and (65) by varying the threshold  $T$  over all possible values of observed information rates. The resulting ROC curves are shown in Fig. 8. We observe that CBI consistently shows a significant variation in connectivity performing either better ( $F5 \leftrightarrow P6$ ) or at least identical ( $Fz \leftrightarrow Cz$ ) to the other causally conditioned measures, with

the larger area under the ROC curve for CBI indicating superior classification accuracy. Also, since there is no direct link in the direction from  $Cz \rightarrow Fz$  (see Table III), the causally conditioned directional information measures do not show significant discriminability in connectivity. However, an important observation here is that for the electrode pair F5 and P6, the directional measures identify an almost symmetric change in connectivity in both directions. This is because for complex interconnected systems (as modeled by the EEG electrodes in Table III) there is not necessarily a greater directional information transfer from source to destination than from destination to source [22], [68]. Schreiber [22], Wibral *et al.* [38], and Venkatesh and Grover [68] specifically highlight the inherent limitations of applying directed information to infer causal connectivity in an unknown network without *a priori* knowledge of which node is the source and which node is the destination.

#### D. Performance of Information Measures on Real EEG Brain Data

In the following, we show that CBI offers a significant improvement in performance over the other information measures when applied to infer connectivity for real EEG data from human subjects. Towards this end we select a source ROI  $\mathcal{X}$ , a destination ROI  $\mathcal{Y}$ , and group all the other six remaining ROIs as  $\mathcal{Z}$ . We then use the closed form expressions (60)-(63) to calculate the instantaneous information rates for each of the 56 possible combinations of source-destination ROI pairs. The corresponding covariance matrices  $\mathbf{C}(\cdot)$  of the joint Gaussian distributions for each source-destination combination are estimated using 125 millisecond long overlapping sliding windows (having a total length of  $N = 32$  points, with a 8 point overlap on each side) of the trial data. We assume stationarity of the EEG signals within these windows, implying that the functional connectivity does not vary in this segment and also make an implicit Markovian assumption that the current time window captures all past activity between the electrodes. For each sliding window position the covariance matrix is computed as an average over the whole sample space for that specific subject,  $\forall n = 1, \dots, N$ . The sample space for calculating each covariance matrix is created by pooling across segments with the same audio quality from a total of 56 trials across multiple presentations of different distortion and music types. Further, since we consider each electrode in an ROI to be an independent realization of the same random process, we also pool across all electrodes in the ROI. Thus, for 9 electrodes in each ROI and 56 trials with the music quality repeating twice in each trial, we obtain a total of  $56 \times 2 \times 9 = 1,008$  sections (realizations) for each window position per ROI. In order to try and ensure that the number of electrodes in an ROI does not change bias w.r.t. to the increasing sample size, we always use the same number of pooled sections (realizations) to create the effective sample space. The covariance is then computed using sample realizations at the corresponding time across all 1,008 sections, assuming stationarity at that time point across sections. Thus,

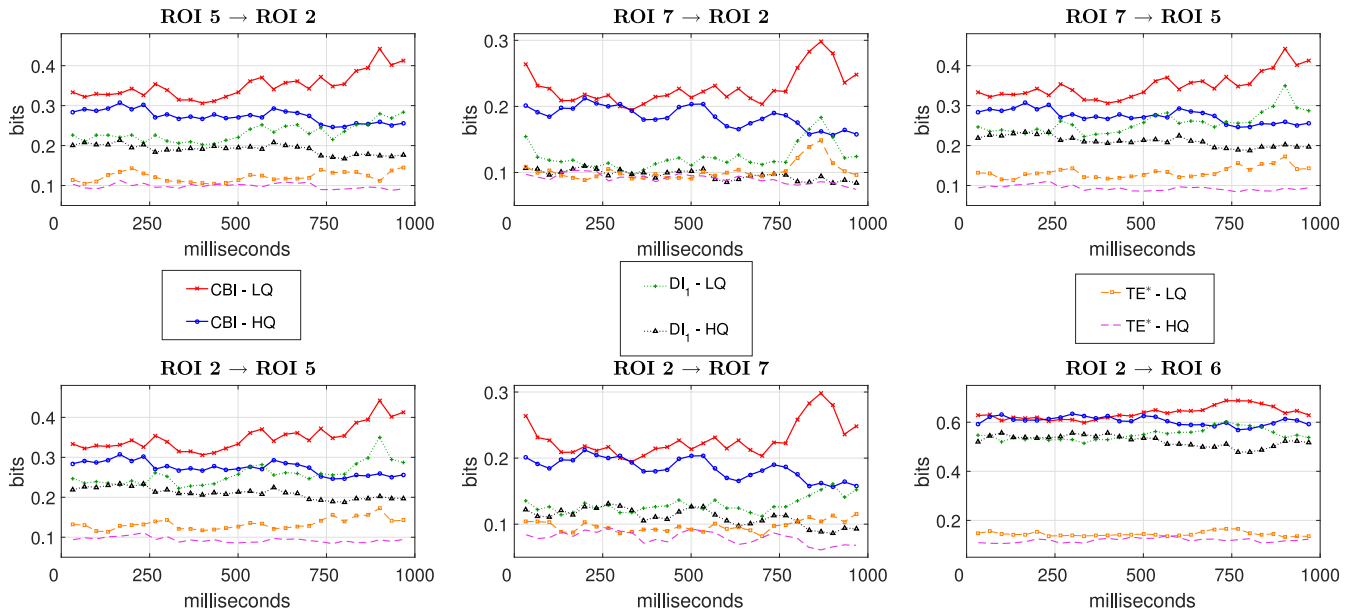


Fig. 9. Instantaneous information transfer rate vectors  $\mathbf{I}$ ,  $\mathbf{DI}_1$  and  $\mathbf{TE}^*$  between six different ROI pairs, averaged over eight human subjects for a second of the trial data after stimulus onset. We observe a significant difference between the transfer rates of the two audio qualities, with the transfer rates being notably higher across all ROIs when the subjects listen to degraded quality audio.

over all sliding window positions we obtain a vector

$$\mathbf{I}(X^N \Rightarrow Y^N || Z^{N-1}) = \left[ I_1(X^N \Rightarrow Y^N || Z^{N-1}), \dots, I_K(X^N \Rightarrow Y^N || Z^{N-1}) \right], \quad (66)$$

where  $K$  is the total number of window positions and  $I_k(X^N \Rightarrow Y^N || Z^{N-1})$  is the instantaneous information rate for the  $k$ -th window computed using (60). The vectors  $\mathbf{DI}_1(X^N \rightarrow Y^N || Z^{N-1})$ ,  $\mathbf{DI}_2(X^N \rightarrow Y^N || Z^{N-1})$  and  $\mathbf{TE}^*(X^{N-1} \rightarrow Y^N || Z^{N-1})$  are constructed similarly. These vectors will be considered further and are hereby denoted as  $\mathbf{I}$ ,  $\mathbf{DI}_1$ ,  $\mathbf{DI}_2$ , and  $\mathbf{TE}^*$  for brevity.

Also, we only select a subset of eight subjects who provide the largest median mutual information values on the event related potential (ERP) channel connecting the audio stimulus and the quantized EEG sensor outputs (see [26] for details). It follows from the discussion in [26] that these were the subjects whose EEG recordings showed the maximum response to the changing audio quality levels. We do not combine data between subjects, and the information rates are calculated separately for each of these eight subjects for all possible combinations of ROI pairs.

1) *Instantaneous Information Results:* Fig. 9 illustrates the vectors for the causal measures  $\mathbf{I}$ ,  $\mathbf{DI}_1$ , and  $\mathbf{TE}^*$  plotted over time for six different combinations of ROI pairs for a second of EEG trial data, averaged over all eight subjects. The 0ms marker on the horizontal axis is the stimulus onset time, i.e., the instant when the audio quality changed. The results indicate a notable difference between the amount of causal information flow for high and degraded audio, in particular for CBI. More specifically, there appears to be a higher amount of information flow between the ROIs when the subject listens to degraded quality audio.

The ROIs we select for plotting in Fig. 9 are based on the direction and order of the auditory sensing pathway in the brain. The primary auditory cortex 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 primary auditory cortex and the other cortical regions, predominantly including the prefrontal cortex [69], [70]. We therefore plot the information rates between the temporal lobes (ROI 5, ROI 7), and the prefrontal cortex (ROI 2) in Fig. 9. We observe here that  $\mathbf{DI}_1$  and  $\mathbf{TE}^*$  detect an almost identical (symmetric) information flow in both directions between ROI pairs 2, 5 and 2, 7, respectively. Other important auditory pathways are the dual dorsal and ventral streams which carry information medially between the prefrontal cortex, the temporal lobes and the parietal lobe [71], [72]. We therefore also include the rates from the prefrontal cortex to the parietal lobe (ROI 2  $\rightarrow$  ROI 6) in Fig. 9.

Note that the first second after stimulus onset can be considered as the *transient response* of the brain's initial perception towards stimulus. Since we are interested in detecting change in connectivity in response to the change in audio quality, for the rest of our analysis here we focus on the information transfer over this initial one second transient period.

2) *ROC Results:* We construct separate classes for each of the different information measures for every ROI pair. Therefore for a given ROI pair,  $\mathcal{P}$  corresponds to one of the  $\mathbf{I}$ ,  $\mathbf{DI}_1$ ,  $\mathbf{DI}_2$ , or  $\mathbf{TE}^*$  rate vectors for degraded quality concatenated for all eight subjects over the first second after stimulus onset, and likewise for  $\mathcal{N}$  with respect to high quality. Then, the observed information rates in each of the vectors are equivalent to the measurement scores.

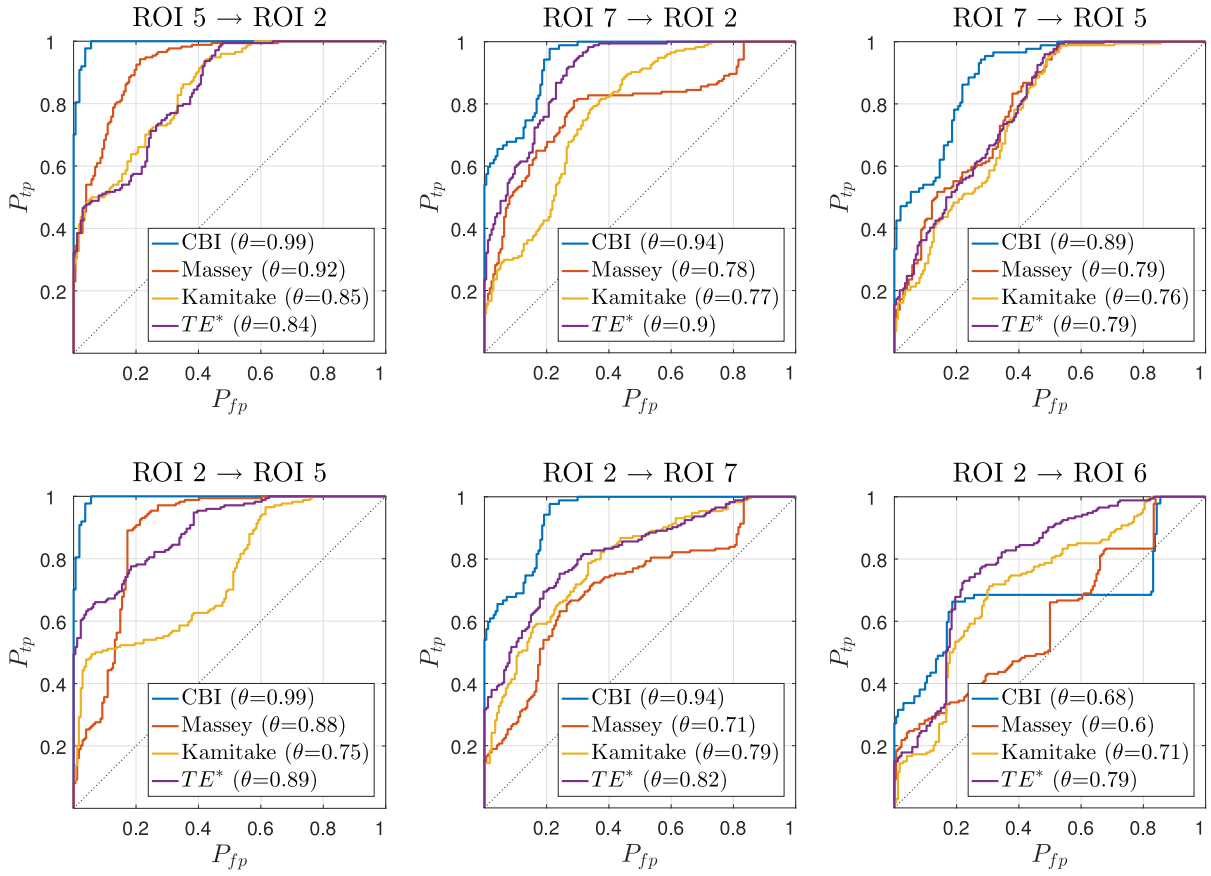


Fig. 10. ROC curves for different information measure and the same six ROI pairs shown in Fig. 9, along with the corresponding area  $\theta$  under the ROC curve. The ROC curve for each information measure is constructed separately by using the corresponding information rates over eight human subjects.

In our setup, given a discrimination threshold value  $T$  for the information rate,  $P_{tp}(T)$  is calculated according to (64) and corresponds to the empirical probability that the observed information rate for degraded quality audio is correctly classified as degraded quality, while  $P_{fp}(T)$  is calculated according to (65) and is the empirical probability that an observed information rate for high quality audio is misclassified as degraded quality.

Fig. 10 shows the ROC curves for each of the information measures, for the same six ROI pairs shown in Fig. 9. For each of the constructed ROC curves in Fig. 10 we also provide the total area under that curve  $\theta$  [61], [63]. We observe from the results in Fig. 10 that the classifier performance of the information measures varies considerably depending on the ROI pair chosen. This is due to the fact that substantial changes are likely to occur between the observed information rates for high quality and degraded quality only for those ROI connections which are actively involved in detecting and processing an auditory stimulus response. For five out of the six ROI pairs depicted in Fig. 10, we notice that CBI performs exceedingly well and has the best discriminability among all given information measures.

3) *Statistical Significance Testing of Area Under the ROC Curve*: The precision of an estimate of the area under a ROC curve is validated by conducting a test for statistical significance [73], [74]. The statistical significance test is used to

determine whether or not the area under the ROC curve is lesser than a specific value of interest  $c$ . The null hypothesis that we are interested in testing can therefore be defined as

$$H_0 : \theta \leq c, \text{ i.e., the area under the ROC curve is less than a specific value of interest } c. \quad (67)$$

Since we do not know the underlying distribution of the observed information rates we employ a non-parametric approach for significance testing by using the method of bootstrapping [73]–[76]. Further, since the observed information rates in the vectors  $\mathbf{I}$ ,  $\mathbf{DI}_1$ ,  $\mathbf{DI}_2$ , and  $\mathbf{TE}^*$  are correlated in time with high probability, we use a modified block bootstrapping approach [77] to preserve this temporal correlation. Given a class  $\mathcal{P}$  with a total of  $N_p$  samples, the block bootstrapping procedure divides this class into overlapping blocks of  $l$  samples each, to create a total of  $N_p - l + 1$  blocks. Bootstrapping is then performed by drawing  $m = N_p/l$  blocks with replacement and concatenating them to form a resampled class. We pick  $l = N_p^{1/3} \simeq 8$  [77]. The significant test is then carried out by performing the following steps:

- i) We perform block bootstrapping on the positive class  $\mathcal{P}$  to create a resampled positive class  $\hat{\mathcal{P}}_b$ , with  $b = 1, 2, \dots, B$ . Likewise  $\mathcal{N}$  is block bootstrapped to create resampled negative class  $\hat{\mathcal{N}}_b$ . Here we use  $B = 2,000$  resamples. Bootstrapping is performed independently for

the positive and negative classes by strictly resampling from only within that particular class.

- ii) We calculate the area under the ROC curve for the pair of bootstrap resampled classes  $\hat{\mathcal{P}}_b$  and  $\hat{\mathcal{N}}_b$ , and denote it as  $\hat{\theta}_b, \forall b = 1, 2, \dots, B$ .
- iii) The standard deviation of the area under the ROC curve can be calculated as

$$\sigma_{\hat{\theta}} = \sqrt{\frac{\sum_{b=1}^B (\hat{\theta}_b - \bar{\hat{\theta}})^2}{B-1}}, \quad (68)$$

where  $\bar{\hat{\theta}} = \sum \hat{\theta}_b / B$  is the estimated mean of the area under  $B$  resampled ROC curves.

- iv) Let  $z_{\hat{\theta}} = (\hat{\theta} - \bar{\hat{\theta}}) / \sigma_{\hat{\theta}}$  denote the normalized random variable corresponding to the bootstrap calculated test statistic  $\hat{\theta}$ . Then  $\lim_{B \rightarrow \infty} p_{z_{\hat{\theta}}} \sim \mathbb{N}(0, 1)$ , where  $p_{z_{\hat{\theta}}}$  is the empirical probability distribution of  $z_{\hat{\theta}}$ , and  $\mathbb{N}(0, 1)$  is the standard Gaussian distribution with zero mean and unit variance.
- v) We then calculate the probability of observing a value for the area under the ROC curve under the null hypothesis, or in other words, the probability of  $\Pr(\hat{\theta} \leq c)$ . This one-sided tail probability is known as the p-value of the test and can be easily calculated using the cumulative distribution function (cdf) of the standard normal distribution,  $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp(-\frac{u^2}{2}) du$ , as  $p_c \triangleq \Pr(\hat{\theta} \leq c) = \Phi\left(\frac{c - \bar{\hat{\theta}}}{\sigma_{\hat{\theta}}}\right)$ . Note that a smaller p-value is evidence against  $H_0$ , see (67).
- vi) If the p-value is smaller than a significance threshold  $\alpha$ , we may safely reject the null hypothesis. Therefore the null hypothesis  $H_0: \theta \leq c$  will be rejected if  $p_c \leq \alpha$ . For our purpose, we set the significance threshold as  $\alpha = 0.05$ .

4) *Inferring Significantly Changing Connections:* A connection between an ROI pair is deemed to *change significantly* if it shows a pronounced difference in the information transfer rates between high and degraded quality, thereby indicating that these ROIs show the greatest change in brain (electrical) activity in response to the change of audio quality.<sup>2</sup> Since a larger area under the ROC curve indicates a greater statistical difference between the information rates, significant connections are identified by testing the null hypothesis  $H_0: \theta \leq c$  for a sufficiently large value of  $c$ . However, the optimal choice of the cutoff threshold for inferring the functional connectivity network in the brain is a rather difficult task [78], [79], and often depends on several criteria including the number of nodes and connections (edges) in the network, the true topology of the underlying functional network, and the nature of the neurophysiological task involved.

<sup>2</sup>We emphasize here that the statistical significant results are not used to identify connections with the largest information flow but rather those with the *largest change* in information flow, where the former is neither checked for, nor required to distinguish between the audio qualities. Note that due to the high dimensionality of the measurement space, assessing statistical significance becomes much more difficult if  $X^N$ ,  $Y^N$ , and  $Z^N$  are not jointly Gaussian distributed.

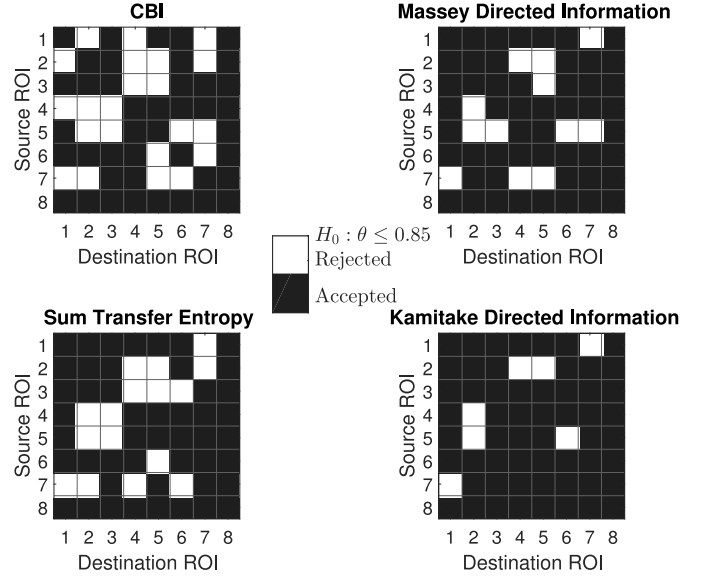


Fig. 11. Inferred connectivity change matrix in response to the change in the perceived audio quality. The null hypothesis tested is  $H_0: \theta \leq 0.85$ , for which the area under the ROC curve is not a significant connection. The white squares indicate significant connections which show the largest change in their information transfer rates in response to changing audio quality.

For the sake of exposition we choose here  $c = 0.85$  and test against the null hypothesis  $H_0: \theta \leq 0.85$ . We perform the significance test via bootstrapping as outlined above and calculate the corresponding p-value  $p_{0.85}$  for each information measure and all ROI pairs. If  $p_{0.85} \leq 0.05$ , then  $H_0$  is rejected, and the ROI pair is declared to be a significantly changing connection. Fig. 11 shows a matrix representation of the net result for all possible combination of ROI pairs. The rows of the matrix represent the source ROI and the columns represent the destination ROI. We observe from these results that CBI is most successful at rejecting the null hypothesis  $\theta \leq 0.85$  across all ROI pairs showing the highest discriminability between the high and degraded quality. We also observe that while Massey's directed information and sum transfer entropy perform much better than Kamitake's directed information, they still perform poorly when compared to CBI. As CBI is a bidirectional symmetric measure the inferred change in connectivity between the ROI pairs is also symmetric. Note that there also appears to be a high degree of bidirectional symmetry in the inferred connectivity change using Massey's and Kamitake's directed information and sum transfer entropy. Finally, we observe from Fig. 11 that every connection identified as significant by Massey's directed information and/or sum transfer entropy is likewise identified as significant by CBI, which is to be expected since CBI is in effect an extension of the two as shown in Proposition 4.

We observe from the inferred connectivity change matrix for CBI that the most active regions are the temporal lobes (ROIs 5 and 7) along with the frontal lobe (ROIs 1, 2 and 3). This is consistent with the literature that the ventral auditory pathway, which in part includes the auditory cortex and the prefrontal cortex, has a role in auditory-object processing and perception [70], [72], [80]. There is also significant information flow from both the left and right temporal lobes to

the parietal lobe (ROI 6). Interestingly, there appears to be no dorsal information flow from the parietal lobe (ROI 6) to the frontal lobes. Also, no significant connections are detected in the occipital lobe (ROI 8), which houses the primary visual cortex and would not be expected to show significant activity for an auditory-discrimination task.

## VII. CONCLUSION

We presented a novel information theoretic framework to assess changes in perceived audio quality by directly measuring the EEG response of human subjects listening to time-varying distorted audio. Causal and directional information measures were used to infer the change in connectivity between EEG sensors grouped into ROIs over the cortex. In particular, Massey's directed information, Kamitake's directed information, and sum transfer entropy were each used to measure information flow between ROI pairs while successfully accounting for the influence from all other interacting ROIs using causal conditioning. We also proposed a new information measure which is shown to be a causal bidirectional modification of directed information applied to a generalized cortical network setting, and whose derivation is strongly related to the classical MAC with feedback. Further, we showed that CBI performs significantly better in being able to distinguish between the audio qualities when compared to the other directed information measures. The connectivity results demonstrate that a change in information flow between different brain regions typically occurs as the subjects listen to different audio qualities, with an overall increase in the information transfer when the subjects listen to degraded quality as opposed to high quality audio. We also observe significant connections with respect to a change in audio quality between the temporal and frontal lobes, which is consistent with the regions that would be expected to be actively involved during auditory signal processing in the brain.

## REFERENCES

- [1] K. Mehta and J. Kliever, "Directed information measures for assessing perceived audio quality using EEG," in *Proc. Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2015, pp. 123–127.
- [2] K. Mehta and J. Kliever, "A new EEG-based causal information measure for identifying brain connectivity in response to perceived audio quality," in *Proc. IEEE Int. Conf. Commun.*, Paris, France, May 2017, pp. 1–6.
- [3] K. J. Friston, "Functional and effective connectivity in neuroimaging: A synthesis," *Human Brain Mapping*, vol. 2, nos. 1–2, pp. 56–78, 1994.
- [4] B. Horwitz, "The elusive concept of brain connectivity," *Neuroimage*, vol. 19, no. 2, pp. 466–470, 2003.
- [5] A. G. Dimitrov, A. A. Lazar, and J. D. Victor, "Information theory in neuroscience," *J. Comput. Neurosci.*, vol. 30, no. 1, pp. 1–5, 2011.
- [6] F. Rieke, D. Warland, R. De Ruyter van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code*. Cambridge, MA, USA: MIT Press, 1999.
- [7] L. Paninski, "Estimation of entropy and mutual information," *Neural Comput.*, vol. 15, no. 6, pp. 1191–1253, 2003.
- [8] D. Golomb, J. Hertz, S. Panzeri, A. Treves, and B. Richmond, "How well can we estimate the information carried in neuronal responses from limited samples?" *Neural Comput.*, vol. 9, no. 3, pp. 649–665, 1997.
- [9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 2012.
- [10] I. E. Ohiorhenuan *et al.*, "Sparse coding and high-order correlations in fine-scale cortical networks," *Nature*, vol. 466, no. 7306, pp. 617–621, 2010.
- [11] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek, "Weak pairwise correlations imply strongly correlated network states in a neural population," *Nature*, vol. 440, no. 7087, pp. 1007–1012, 2006.
- [12] J. Shlens *et al.*, "The structure of large-scale synchronized firing in primate retina," *J. Neurosci.*, vol. 29, no. 15, pp. 5022–5031, 2009.
- [13] J. Xu, Z.-R. Liu, R. Liu, and Q.-F. Yang, "Information transmission in human cerebral cortex," *Physica D Nonlin. Phenomena*, vol. 106, nos. 3–4, pp. 363–374, 1997.
- [14] L. Wu, P. Neskovic, E. Reyes, E. Festa, and H. William, "Classifying n-back EEG data using entropy and mutual information features," in *Proc. Eur. Symp. Artif. Neural Netw.*, Bruges, Belgium, 2007, pp. 61–66.
- [15] J. Jeong, J. C. Gore, and B. S. Peterson, "Mutual information analysis of the EEG in patients with Alzheimer's disease," *Clin. Neurophysiol.*, vol. 112, no. 5, pp. 827–835, 2001.
- [16] S. H. Na, S.-H. Jin, S. Y. Kim, and B.-J. Ham, "EEG in schizophrenic patients: Mutual information analysis," *Clin. Neurophysiol.*, vol. 113, no. 12, pp. 1954–1960, 2002.
- [17] H. Harada, Y. Eura, K. Shiraishi, T. Kato, and T. Soda, "Coherence analysis of EEG changes during olfactory stimulation," *Clin. Electroencephalography*, vol. 29, no. 2, pp. 96–100, 1998.
- [18] B.-C. Min *et al.*, "Mutual information analysis of EEG responses by odor stimulation," *Chem. Senses*, vol. 28, no. 9, pp. 741–749, 2003.
- [19] H. Marko, "The bidirectional communication theory—A generalization of information theory," *IEEE Trans. Commun.*, vol. 21, no. 12, pp. 1345–1351, Dec. 1973.
- [20] J. Massey, "Causality, feedback and directed information," in *Proc. Int. Symp. Inf. Theory Appl.*, Waikiki, HI, USA, Nov. 1990, pp. 303–305.
- [21] T. Kamitake, H. Harashima, and H. Miyakawa, "A time-series analysis method based on the directed transinformation," *Electron. Commun. Japan (I Commun.)*, vol. 67, no. 6, pp. 1–9, 1984.
- [22] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.*, vol. 85, no. 2, pp. 461–464, 2000.
- [23] P.-O. Amblard and O. J. J. Michel, "On directed information theory and Granger causality graphs," *J. Comput. Neurosci.*, vol. 30, no. 1, pp. 7–16, 2011.
- [24] G. Kramer, "Directed information for channels with feedback," Ph.D. dissertation, Dept. Inf. Technol. Elect. Eng., Swiss Federal Inst. Technol., Zürich, Switzerland, 1998.
- [25] C. J. Quinn, N. Kiyavash, and T. P. Coleman, "Directed information graphs," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6887–6909, Dec. 2015.
- [26] K. Mehta and J. Kliever, "An information theoretic approach toward assessing perceptual audio quality using EEG," *IEEE Trans. Mol. Biol. Multi Scale Commun.*, vol. 1, no. 2, pp. 176–187, Jun. 2015.
- [27] *Method for Subjective Assessment of Intermediate Quality Levels of Coding Systems*, Int. Telecommun. Union, Geneva, Switzerland, ITU Recommendation ITU-R BS.1534-1, 2003.
- [28] S. Bosse, K.-R. Müller, T. Wiegand, and W. Samek, "Brain-computer interfacing for multimedia quality assessment," in *Proc. IEEE Int. Conf. Syst. Man Cybern. (SMC)*, Budapest, Hungary, 2016, pp. 002834–002839.
- [29] U. Engelke *et al.*, "Psychophysiology-based QoE assessment: A survey," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 1, pp. 6–21, Feb. 2017.
- [30] C. D. Creusere, J. Kroger, S. R. Siddenki, P. Davis, and J. Hardin, "Assessment of subjective audio quality in Proc. 20th Eur. Signal Process. Conf., Bucharest, Romania, 2012, pp. 2704–2708.
- [31] A. K. Porbadnigk *et al.*, "Using ERPs for assessing the (sub) conscious perception of noise," in *Proc. 32nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Buenos Aires, Argentina, 2010, pp. 2690–2693.
- [32] S. Scholler *et al.*, "Toward a direct measure of video quality perception using EEG," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2619–2629, May 2012.
- [33] M. Mustafa, S. Guthe, and M. Magnor, "Single-trial EEG classification of artifacts in videos," *ACM Trans. Appl. Percept.*, vol. 9, no. 3, pp. 1–15, 2012.
- [34] C. D. Creusere and J. C. Hardin, "Assessing the quality of audio containing temporally varying distortions," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 4, pp. 711–720, May 2011.
- [35] G. Kramer, "Causal conditioning, directed information and the multiple-access channel with feedback," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, Aug. 1998, p. 189.
- [36] J. Pearl, *Causality: Models, Reasoning and Inference*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2009.
- [37] S. Ito, "Backward transfer entropy: Informational measure for detecting hidden Markov models and its interpretations in thermodynamics, gambling and causality," *Sci. Rep.*, vol. 6, Nov. 2016, Art. no. 36831.
- [38] M. Wibral, R. Vicente, and J. T. Lizier, *Directed Information Measures in Neuroscience*. Heidelberg, Germany: Springer, 2014.
- [39] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica J. Econometric Soc.*, vol. 37, no. 3, pp. 424–438, 1969.

- [40] L. Barnett, A. B. Barrett, and A. K. Seth, "Granger causality and transfer entropy are equivalent for Gaussian variables," *Phys. Rev. Lett.*, vol. 103, no. 23, 2009, Art. no. 238701.
- [41] C. J. Quinn, T. P. Coleman, N. Kiyavash, and N. G. Hatsopoulos, "Estimating the directed information to infer causal relationships in ensemble neural spike train recordings," *J. Comput. Neurosci.*, vol. 30, no. 1, pp. 17–44, 2011.
- [42] N. Soltani and A. Goldsmith, "Inferring neural connectivity via measured delay in directed information estimates," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 2503–2507.
- [43] P.-O. Amblard and O. Michel, "Causal conditioning and instantaneous coupling in causality graphs," *Inf. Sci.*, vol. 264, pp. 279–290, Apr. 2014.
- [44] J. T. Lizier, M. Prokopenko, and A. Y. Zomaya, "Local information transfer as a spatiotemporal filter for complex systems," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 77, no. 2, 2008, Art. no. 026110.
- [45] J. T. Lizier, M. Prokopenko, and A. Y. Zomaya, "Information modification and particle collisions in distributed computation," *Chaos Interdiscipl. J. Nonlin. Sci.*, vol. 20, no. 3, 2010, Art. no. 037109.
- [46] V. A. Vakorin, O. A. Krakovska, and A. R. McIntosh, "Confounding effects of indirect connections on causality estimation," *J. Neurosci. Methods*, vol. 184, no. 1, pp. 152–160, 2009.
- [47] O. Sakata, T. Shiina, and Y. Saito, "Multidimensional directed information and its application," *Electron. Commun. Japan (III Fundam. Electron. Sci.)*, vol. 85, no. 4, pp. 45–55, 2002.
- [48] M. Al-Khassawneh and S. Aviyente, "The relationship between two directed information measures," *IEEE Signal Process. Lett.*, vol. 15, pp. 801–804, 2008.
- [49] V. Solo, "On causality and mutual information," in *Proc. IEEE Conf. Decis. Control*, Cancún, Mexico, Dec. 2008, pp. 4939–4944.
- [50] Y. Liu and S. Aviyente, "The relationship between transfer entropy and directed information," in *Proc. IEEE Stat. Signal Process. Workshop (SSP)*, Ann Arbor, MI, USA, 2012, pp. 73–76.
- [51] P.-O. Amblard and O. J. J. Michel, "The relation between Granger causality and directed information theory: A review," *Entropy*, vol. 15, no. 1, pp. 113–143, 2012.
- [52] G. W. Corder and D. I. Foreman, *Nonparametric Statistics: A Step-by-Step Approach*. Hoboken, NJ, USA: Wiley, 2014.
- [53] H.-Y. Kim, "Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis," *Restorative Dentistry Endodont.*, vol. 38, no. 1, pp. 52–54, 2013.
- [54] S. G. West, J. F. Finch, and P. J. Curran, "Structural equation models with nonnormal variables: Problems and remedies," in *Structural Equation Modeling: Concepts, Issues, and Applications*. Thousand Oaks, CA, USA: Sage, 1995, pp. 56–75.
- [55] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.
- [56] W. J. Krzanowski and D. J. Hand, *ROC Curves for Continuous Data*. Boca Raton, FL, USA: CRC Press, 2009.
- [57] T. Bossomaier, L. Barnett, M. Harré, and J. T. Lizier, *An Introduction to Transfer Entropy*. Cham, Switzerland: Springer, 2016.
- [58] S. Ito *et al.*, "Extending transfer entropy improves identification of effective connectivity in a spiking cortical network model," *PLoS ONE*, vol. 6, no. 11, 2011, Art. no. e27431.
- [59] M. Garofalo, T. Nieus, P. Massobrio, and S. Martinoia, "Evaluation of the performance of information theory-based methods and cross-correlation to estimate the functional connectivity in cortical networks," *PLoS ONE*, vol. 4, no. 8, 2009, Art. no. e6482.
- [60] J. T. Lizier and M. Rubinov, "Inferring effective computational connectivity using incrementally conditioned multivariate transfer entropy," *BMC Neurosci.*, vol. 14, Jul. 2013, Art. no. P337.
- [61] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognit.*, vol. 30, no. 7, pp. 1145–1159, 1997.
- [62] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.
- [63] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach," *Biometrics*, vol. 44, no. 3, pp. 837–845, 1988.
- [64] D. A. Bridwell, S. Rachakonda, R. F. Silva, G. D. Pearlson, and V. D. Calhoun, "Spatiospectral decomposition of multi-subject EEG: Evaluating blind source separation algorithms on real and realistic simulated data," *Brain Topography*, vol. 31, no. 1, pp. 47–61, 2018.
- [65] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, 2004.
- [66] P. Nunez and R. Srinivasan, *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford, U.K.: Oxford Univ. Press, 2006.
- [67] R. Vicente, M. Wibral, M. Lindner, and G. Pipa, "Transfer entropy—A model-free measure of effective connectivity for the neurosciences," *J. Comput. Neurosci.*, vol. 30, no. 1, pp. 45–67, 2011.
- [68] P. Venkatesh and P. Grover, "Is the direction of greater Granger causal influence the same as the direction of information flow?" in *Proc. Annu. Allerton Conf. Commun. Control Comput.*, Monticello, IL, USA, Sep. 2015, pp. 672–679.
- [69] L. M. Romanski *et al.*, "Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex," *Nat. Neurosci.*, vol. 2, no. 12, pp. 1131–1136, 1999.
- [70] J. P. Rauschecker and S. K. Scott, "Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing," *Nat. Neurosci.*, vol. 12, no. 6, pp. 718–724, 2009.
- [71] J. Ahveninen *et al.*, "Task-modulated 'what' and 'where' pathways in human auditory cortex," *Proc. Nat. Acad. Sci. USA*, vol. 103, no. 39, pp. 14608–14613, 2006.
- [72] W.-J. Wang, X.-H. Wu, and L. Li, "The dual-pathway model of auditory signal processing," *Neurosci. Bull.*, vol. 24, no. 3, pp. 173–182, 2008.
- [73] C. Cortes and M. Mohri, "Confidence intervals for the area under the ROC curve," in *Proc. Adv. Neural Inf. Process. Syst.*, 2005, pp. 305–312.
- [74] H. Liu, G. Li, W. G. Cumberland, and T. Wu, "Testing statistical significance of the area under a receiving operating characteristics curve for repeated measures design with bootstrapping," *J. Data Sci.*, vol. 3, no. 3, pp. 257–278, 2005.
- [75] B. Efron, "Nonparametric standard errors and confidence intervals," *Can. J. Stat.*, vol. 9, no. 2, pp. 139–158, 1981.
- [76] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, vol. 57. Boca Raton, FL, USA: CRC Press, 1994.
- [77] P. Hall, J. L. Horowitz, and B.-Y. Jing, "On blocking rules for the bootstrap with dependent data," *Biometrika*, vol. 82, no. 3, pp. 561–574, 1995.
- [78] B. C. Van Wijk, C. J. Stam, and A. Daffertshofer, "Comparing brain networks of different size and connectivity density using graph theory," *PLoS ONE*, vol. 5, no. 10, 2010, Art. no. e13701.
- [79] N. Langer, A. Pedroni, and L. Jäncke, "The problem of thresholding in small-world network analysis," *PLoS ONE*, vol. 8, no. 1, 2013, Art. no. e53199.
- [80] J. K. Bizley and Y. E. Cohen, "The what, where and how of auditory-object perception," *Nat. Rev. Neurosci.*, vol. 14, no. 10, pp. 693–707, 2013.



**Ketan Mehta** received the M.S. degree and the Ph.D. degree in electrical engineering from New Mexico State University, Las Cruces, NM, USA, in 2010 and 2017, respectively. From 2012 to 2017, he was a Research Assistant with New Mexico State University. His research interests span information theory, signal processing, and statistical algorithms with interdisciplinary applications in neural signal processing, and cognitive neuroscience.



**Jörg Kliewer** (S'97–M'99–SM'04) received the Dipl.-Ing. (M.Sc.) degree in electrical engineering from the Hamburg University of Technology, Hamburg, Germany, in 1993 and the Dr.-Ing. (Ph.D.) degree in electrical engineering from the University of Kiel, Germany, in 1999, respectively. He was a Research Assistant with the University of Kiel, from 1993 to 1998, and a Senior Researcher and Lecturer with the University of Kiel, from 1999 to 2004. In 2004, he visited the University of Southampton, U.K., for one year, and

from 2005 to 2007, he was with the University of Notre Dame, IN, USA, as a Visiting Assistant Professor. From 2007 to 2013, he was with New Mexico State University, Las Cruces, NM, USA, most recently as an Associate Professor. He is currently with the New Jersey Institute of Technology, Newark, NJ, USA, as an Associate Professor. His current research interests span information and coding theory, graphical models, and statistical algorithms, which includes applications to networked communication and security, data storage, and biology. He was a recipient of the Leverhulme Trust Award and the German Research Foundation Fellowship Award in 2003 and 2004, respectively. He was an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS from 2008 to 2014, and since 2015 serves as an Area Editor for the same journal. He is also an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY since 2017 and is an Editorial Board Member of the IEEE INFORMATION THEORY NEWSLETTER since 2012.