An Information Theoretic Approach Toward Assessing Perceptual Audio Quality Using EEG

Ketan Mehta and Jörg Kliewer, *Senior Member, IEEE*

***Abstract*—In this paper, we propose a novel information theoretic model to interpret the entire "transmission chain" comprising stimulus generation, brain processing by the human subject, and the electroencephalograph (EEG) response measurements as a nonlinear, time-varying communication channel with memory. We use mutual information (MI) as a measure to assess audio quality perception by directly measuring the brainwave responses of the human subjects using a high resolution EEG. Our focus here is on audio where the quality is impaired by time varying distortions. In particular, we conduct experiments where subjects are presented with audio whose quality varies with time between different possible quality levels. The recorded EEG measurements can be modeled as a multidimensional Gaussian mixture model (GMM). In order to make the computation of the MI feasible, we present a novel low-complexity approximation technique for the differential entropy of the multidimensional GMM. We find the proposed information theoretic approach to be successful in quantifying subjective audio quality perception, with the results being consistent across different music sequences and distortion types.**

***Index Terms*—Mutual information, perception, audio quality, electroencephalography (EEG), Gaussian mixture model (GMM).**

I. Introduction

AN OFTEN used metaphor for the human brain is that of an information processing device. Our perception of the world around is received as input to the brain via sensory nerves. Detecting, interpreting, and responding to a stimulus is a multistage process that results in the activation and hierarchical interaction of several different regions in the brain. In these hierarchies, sensory and motor information is represented and manipulated in the form of neural activity, where neurons transfer information about stimulus features by successively transmitting impulse trains of electrical or chemical signals. Information in the brain is therefore inherently encoded as sequences of neural activity patterns. The superposition of these activity patterns can be non-invasively recorded via electroencephalography (EEG) by placing sensors on the scalp. In the following we provide a novel information theoretic model to analyze the overall transmission chain comprising of input stimulus generation, processing by the human brain, and the output EEG brain activity, as a time-varying, nonlinear communication channel with memory.

Information theory (IT) provides a stochastic framework which is also well suited to characterize and model neural response [2]–[7]. In fact, soon after Shannon's initial work [8], the concept of information was applied to calculate the capacity and bounds of neural information transfer [9]–[11]. In particular, entropy and mutual information (MI) [12] can be used to quantify how much information a neural response contains about a stimulus [2], [13]. In [14] the author discusses local expansions for the nonparametric estimation of entropy and MI useful for neural data analysis. Further, [15], [16] present a power-series-based component expansion for calculating the MI in order to analyze information conveyed by the spike timing of neuron firing in the somatosensory cortex of a rat. In [17] different approximations of Gaussian mixture models are used to bound the MI transmission from stimulus to response in the cercal sensory system of a cricket.

In contrast to these works which assess information transfer on the neural level, we focus on the "high-level" information flow between stimulus and EEG sensor observations. In such a setup MI has also has been successfully employed in the past for EEG feature extraction and classification purposes [18]–[21]. Notably in [22], MI is used to describe information transmission among different cortical areas during waking and sleep states using EEG measurements. Other researchers have similarly used MI to analyze EEG data to investigate corticocortical information transmission for pathological conditions such as Alzheimer's disease [23] and schizophrenia [24], for odor stimulation [25], [26], and even for motor disorders such as Parkinson's disease [27].

For subjective quality testing of audio the current state-of-the-art approach is Multi Stimulus with Hidden Anchor (MUSHRA) [28]. One characteristic that MUSHRA and most of the other existing audio and video testing protocols have in common is that each human participant assigns a single quality-rating score to each test sequence. Such testing suffers from a subject-based bias towards cultural factors in the local testing environment and tends to be highly variable. In contrast, by using EEG to directly analyze the brainwave patterns we capture fast, early brain responses that depend only on the perceived variation in signal quality. Because of these reasons there has recently been a growing interest in using EEGs to classify human perception of audio [29], [30] and visual quality [31]–[34]. For example, [30] investigates the use of a time-space-frequency analysis to identify features in EEG brainwave responses corresponding to time-varying

Manuscript received October 14, 2014; revised May 28, 2015; accepted August 17, 2015. Date of publication November 18, 2015; date of current version January 14, 2016. This work was supported by the NSF under Grant CCF-1065603, and presented in part at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Florence, Italy, May 2014 [1].

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 (e-mail: jkliewer@njit.edu).

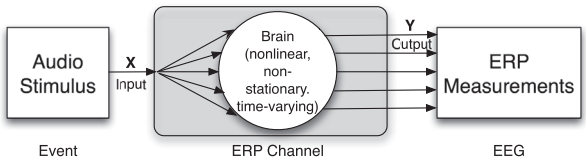
Digital Object Identifier 10.1109/TMBMC.2015.2501744

audio quality. Further, [29], [31] propose to use linear discriminant analysis (LDA) classifiers to extract features based on the P300 ERP component [35], [36] for classifying noise detection in audio signals and to assess changes in perceptual video quality, respectively. Similarly, in [34] the authors employ a wavelet-based approach for an EEG-classification of commonly occurring artifacts in compressed video, using a single-trial EEG.

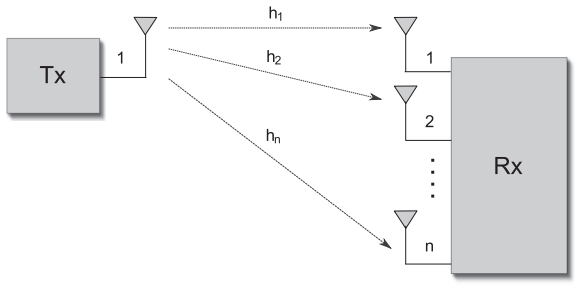
Our approach here is different compared to above-mentioned studies in that we use MI to quantitatively measure how accurately the quality of the audio stimulus is transmitted over the brain response channel, where the observation at the channel output is given by the EEG measurements. In other words, we ask how useful EEG measurements are to assess subjective audio quality. How well can the audio quality be blindly estimated (i.e., without knowing the original audio quality) from the EEG measurements? This could be useful for example for designing custom hearing aids optimized based on perceived audio quality or for a subjective assessment of audio streams which slowly vary their quality with time. For the sake of simplicity and analytical tractability we restrict ourselves to the worst case scenario of only two audio quality levels (high quality and bad quality).

We show that the EEG measurements can be modeled as a multidimensional Gaussian mixture model (GMM). A direct computation of the MI from the EEG measurements is not feasible due to the high dimensionality of the data captured by using a large number of EEG sensors. Instead, we present a low-complexity approximation technique for the differential entropy of the multidimensional GMM based on a Taylor series expansion, which is essential for carrying out the MI computation problem in acceptable time. Additionally, we also provide novel analytical differential entropy bounds for the one-dimensional GMM case. To the best of our knowledge, this is the first time that an information theoretic characterization has been used to evaluate human perception of audio quality by directly measuring the brain activity of the subjects using a high resolution EEG.

The rest of the paper is organized as follows. In Section II we provide an overview of EEG, the ERP channel, the experiment, and the stimulus audio sequences. In Section III we analyze the ERP channel using an information theoretic framework and calculate the MI over this end-to-end channel. The results are presented in Section IV, which is followed by a discussion in Section V. We conclude with a summary of our study and future directions in Section VI.



(a) Communicating over the ERP channel.



(b) Single-input multiple-output channel (SIMO) channel.

Fig. 1. The end-to-end perceptual processing channel is analogous to a SIMO communication channel. The input to the channel is the audio quality level and the output is the EEG response.

II. Background

A. Electroencephalography (EEG)

EEG is an observation of the electrical activity in the brain recorded over a period of time using electrodes placed on the scalp. As mentioned earlier, a single EEG electrode measures the sum electric potential resulting from the synchronous activity of several hundred million neurons averaged over tissue masses. The large spatial scale of EEG recordings make them unsuitable for studying the activity of neurons or small cell assemblies. Instead, EEG provides a robust measure of cortical brain activity and is associated with cognition and behavior.

While EEG is affected by stimulus, it does not require one and occurs even in the absence of external stimulus, for, e.g., brain activity recorded during different sleeping stages. Event-related potentials (ERP) typically represent time averaged EEG data recorded specifically in response to a stimulus like light strobes, audio tones, etc. For this reason ERPs are associated with state-dependent brain processing like selective-attention, task context, memory (recognition), and other changes in mental state [37].

Typically, EEG studies attempt to tie specific cognitive processes to averaged data. While these extracranial recordings provide estimates of large-scale synaptic activity and are closely related to the endogenous brain state, the most appropriate way in which to combine the data in order to achieve this goal is not clear. Common methods of analyzing and processing recorded EEG include epoching, averaging, time-frequency analysis and linear discriminant analysis, correlation and coherence rely on linear dependencies and assume the EEG signal to be stationary. In contrast, employing mutual information (MI) is well suited for modeling nonlinear, non-stationary phenomena like EEGs as it is based on empirically observed probability distributions and also performs an implicit averaging.

B. ERP Channel

Our goal here is to use MI to quantify the information transfer between the input audio stimulus and the recorded EEG. In order to achieve this goal, we model the overall transmission chain comprising of stimulus generation, processing by the human brain, and the EEG sensors as a communication channel (Fig. 1(a)). In the neuroscience literature such a channel in general is referred to as the ERP channel [38].

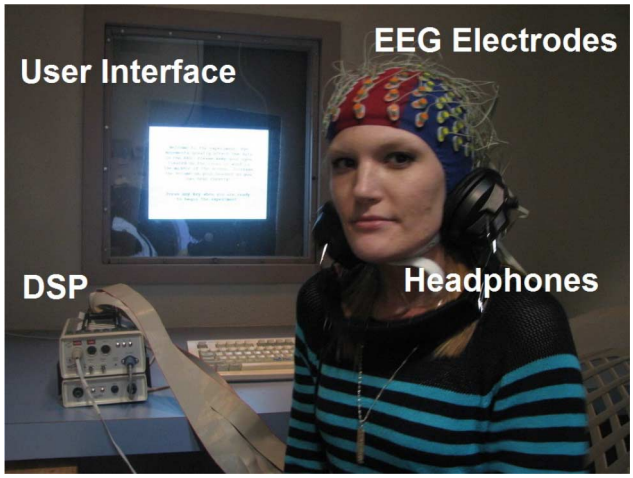


Fig. 2. The EEG experiment setup.

In our case, the ERP channel is equivalent to a single-input multiple-output (SIMO) channel with unknown characteristics. A SIMO channel model is often used in wireless communications to represent a system with a single transmitting antenna and multiple receiving antennas as shown in Fig. 1(b). In particular, for the ERP channel under consideration, the quality of the audio stimulus represents the (single) input, and the observation at the EEG sensor electrodes on the scalp is the (multiple) output of this channel, respectively.

C. Experiment

To collect the data required for the study in our experiments, test subjects were presented with a variety of different audio test sequences whose qualities varied with time. The EEG activity of the subjects was recorded as they passively listened to these test sequences via a pair of high fidelity headphones (see Fig. 2).

Each trial consisted of a computer controlled presentation of an audio test sequence. At the end of the trial, the user interface prompted the subject to continue on to the next trial. A 128-channel ActiveTwo Biosemi EEG capture system monitored and recorded the brain activity of the subjects during the trial. The EEG system comprised of electrodes, connecting wires, and a DSP. The DSP re-referenced and amplified the weak signals picked up by the electrodes on the scalp and then converted the analog signal into a digital signal for storage. The DSP was also connected to the computer and used to synchronize the recordings at the start and end of each trial.

The trials were conducted in a room specifically built for recording EEGs with a RF shielded testing chamber. A total of 28 test subjects, all with normal hearing capability, participated in the experiment with the majority of them being male.

D. Stimulus

All the audio test sequences used as stimulus were created from three fundamentally different base-sequences. The base-sequences were sampled at 44.1 kHz with a precision of 16 bits per sample. Two different types of distortions were implemented – scalar quantization and frequency band truncation – with each distortion type in itself having three quality levels of impairment "Q1", "Q2", and "Q3", respectively (see Table I). Thirty second long test sequences were created from the base-sequence by selecting one of the two distortion types and by applying the three possible impairment levels in a time-varying pattern over the whole duration of the sequence. Fig. 3 shows one of the many possible test sequences used in the study. The criterion and use of these test sequences and distortion quality levels as a metric for quality perception is described in detail in [39].

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 (MDCT) AND EITHER THE FREQUENCY TRUNCATION OR THE SCALAR QUANTIZATION WAS APPLIED TO THE COEFFICIENTS PRIOR TO RECONSTRUCTION [39]

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

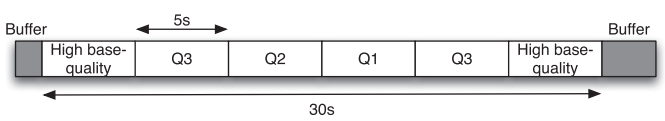


Fig. 3. A 30 second test sequence where the audio quality changes in a time-varying pattern over the whole duration of the sequence. Different possible combinations of quality changes and two distortion types were presented to each subject in a randomized fashion.

Note that despite the subjects were presented with all different quality levels in our listening tests, in our analysis we focus only on the high base-quality and the "Q3" degraded quality audio as inputs to the ERP channel. This addresses the worst-case quality change and keeps the problem analytically and numerically tractable. This choice is also supported from a subjective quality assessment viewpoint as it easier to distinguish between 'good' and 'bad' audio, as opposed to 'medium-good' or 'medium-bad' qualities.

Typically, responses to a stimulus observed by EEG measurements are averaged over different trials to remove inconsistencies between these trials [40]. Evaluating the response to repeated trials provides a measure of discriminability on the given task, allowing to reliably judge which state changes. Multiple presentations of both the impaired test sequences and the unimpaired reference base sequences were made in a randomized fashion. Also, to remove any possible bias, all possible combinations of quality changes, distortion types, and base sequences were presented to each subject.

E. EEG Data

The EEG system captured data on each of the 128 spatial channels, sampled at 1024 Hz. An average re-referencing and baseline removal was performed to remove any DC noise component, which is often introduced over time into the EEG data usually as a result of steady brain activity, but also due

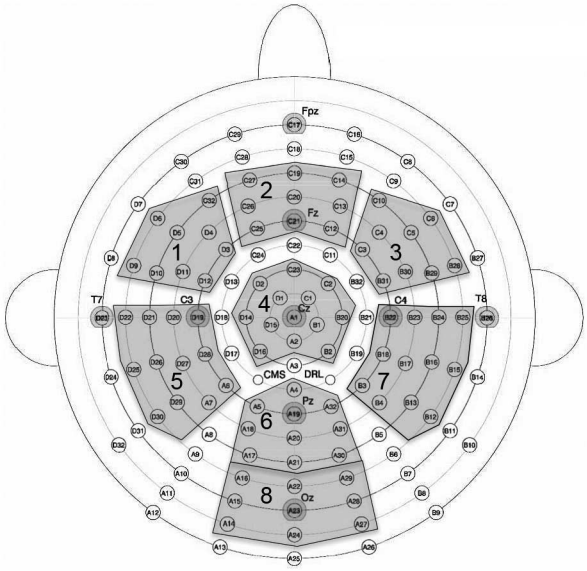


Fig. 4. Schematic representation of the 128-channel EEG system. The electrodes are grouped into eight regions of interest (ROI) to effectively map the activity across different regions of the brain.

to muscle tension, signal fluctuation, or other noise sources. Finally, the data was passed through a high-pass filter with a cut-off frequency at 1 Hz, and a transition bandwidth of 0.2 Hz.

We notice that for each trial we have a very large multidimensional data-set from which to attempt and extract signals corresponding to changes in the human perception of audio quality. To better manage the large amount of collected data while effectively mapping the activity across different regions of the brain, we suggest grouping the 128 electrodes of the EEG-system into specific regions of interest (ROI). The grouping scheme we use is shown in Fig. 4, and is similar to the one used in [41]. 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. For example, regions 5 and 6 cover the temporal lobes, while region 7 spans the parietal lobe. Also, the nearly symmetrical distribution of the ROIs in this scheme allows for an easy mapping and comparison of EEG activity between different cortical areas.

III. Information Theoretic Analysis

Notation: Throughout this paper, random variables are denoted by uppercase letters X and their (deterministic) outcomes by the same letter in lowercase x. Probability distributions are denoted by $p(x)$. A continuous random process is denoted using a time index $Y(t)$. Multivariate random variables are denoted with an underlined uppercase letters $\underline{Y}$. Deterministic vectors are represented as smallcase underlined letters $\underline{\mu}$, while matrices are denoted with boldface $\mathbf{C}$. The expected value of a function is denoted as $\mathbb{E}[\cdot]$.

A. ERP Channel

In our discussion so far, we have shown that the ERP channel under consideration (see Fig. 1) is analogous to a SIMO communication channel. The capacity of SIMO channels has been previously analyzed in [42]. However, since determining the capacity achieving input distribution for the ERP channel is virtually impossible to obtain in an experimental setup, we focus on analyzing the MI for a uniform input distribution (symmetric capacity) as outlined below. Further, since the ERP channel is non-linear, non-stationary, potentially time varying, and does not have an explicitly defined channel characteristic, the expressions stated in [42] for the direct computation of its MI are not applicable. We begin our analysis therefore by taking a closer look at the only known parameters of our model, *viz.* the input and the output random variables.

The input random variable X of the ERP channel is uniformly distributed over a set of class indices $\mathcal{X}$ which describe the quality of the stimulus sequences at any given time interval. As discussed in Section II-C we restrict ourselves to $|\mathcal{X}| = 2$. Then, the audio quality of the input sequence can be represented as an equiprobable Bernoulli distribution

$$X = \begin{cases} x_1, & \text{if the input stimulus is of high quality,} \\ x_2, & \text{if the input stimulus is of degraded quality level 'Q3',} \end{cases}$$

with probabilities

$$p(x_1) = p(x_2) = 1/2. \tag{1}$$

The output of the channel is given by the set $\mathcal{Y}^n \subset \mathbb{R}^n$ containing all possible values of the EEG potential at any given time interval. For a total of n ROIs we therefore get a (multivariate) output vector of random processes $\underline{Y}(t) = (Y_1(t), \ldots, Y_n(t))$. To reduce the variance of the ERP channel measurements we consider every electrode in the i-th ROI as an independent realization of the same random process $Y_i(t)$, $i = 1, \ldots, n$. This is supported by our grouping scheme wherein all electrodes of a ROI are located within close proximity of one another and capture data over the same cortical region. Further, for tractability we assume the random process to be stationary within the five second non-overlapping blocks of a trial with constant audio-quality. The probability function $p_{\underline{Y}}(\cdot)$ is therefore not dependent on time and can be written as $p_{\underline{Y}}\left(\underline{y}(t)\right) = p_{\underline{Y}}(\underline{y})$ without any loss of generality. Note that this assumption does not rule out any non-stationary behavior between sections of different audio quality within the same trial.

The relationship between the input and output of the ERP channel is characterized by its conditional probability distribution. Since the input can take on two distinct values, there are two marginal conditional distributions $p(y_i|x_1)$ and $p(y_i|x_2)$ corresponding to any given ROI. Fig. 5 shows the marginal conditional distributions obtained via histogram measurements of a single ROI output over time. A detailed inspection using different subjects, input sequences, and ROIs allows us to assert two important facts about this distribution. *First*, the conditional distribution converges to a Gaussian with zero mean. 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

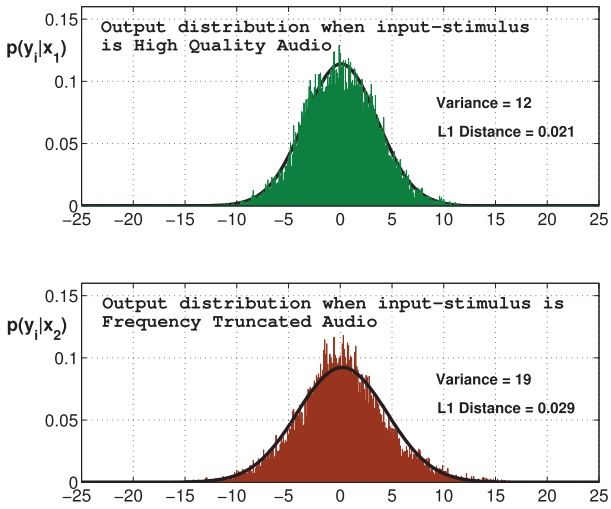


Fig. 5. Marginal conditional distributions $p(y_i|x_1)$ and $p(y_i|x_2)$ of subject S4 over a single ROI, for high quality and frequency distorted audio input-stimulus, respectively. The Gaussian fit is obtained by using an estimator that minimizes the L_1 distance between the fitted Gaussian distribution and the histogram data.

Gaussian distribution as a result of the Central Limit Theorem (CLT). It then follows directly from the CLT that the probability distribution for n ROIs will also converge to a n-dimensional multivariate Gaussian distribution. *Second*, we observe from Fig. 5 that there is a difference between the variances of the distributions $p(y_i|x_1)$ and $p(y_i|x_2)$. This indicates that Y_i is dependent on X, i.e., the EEG data is related to and contains "information" about the input stimulus, which is characterized by the MI $I(X;Y_i)$ as a measure of the information transfer over the ERP channel.

B. Entropy and Mutual Information

The entropy[1] of a (continuous or discrete) random variable X with a probability distribution $p(x)$ quantifies the amount of information in X, and is defined as [12]

$$h(X) = \mathbb{E}[-\log p(x)]. \tag{2}$$

Unless otherwise specified we will take all logarithms to base 2, thus measuring the entropy and MI in bits. For two random variables X and Y with a joint probability density function $p(x, y)$ the MI between them is defined as [12]

$$I(X;Y) = \mathbb{E}_{XY}\left[\log \frac{p(x,y)}{p(x)p(y)}\right]. \tag{3}$$

In this work we are interested in the MI between the ERP channel input X and output $\underline{Y}$. Therefore, a moderate-to-high value for $I(X;\underline{Y})$ indicates that the EEG data output is related to the input stimulus, i.e., contains information about the audio quality. If we consider the conditional probability distributions corresponding to n ROIs chosen simultaneously, then each of the distributions $p(\underline{y}|x=0)$ and $p(\underline{y}|x=1)$ are n-dimensional

[1]In the following, we simply use the denomination "entropy" even when referring to differential entropy for the case of continuous random variables.

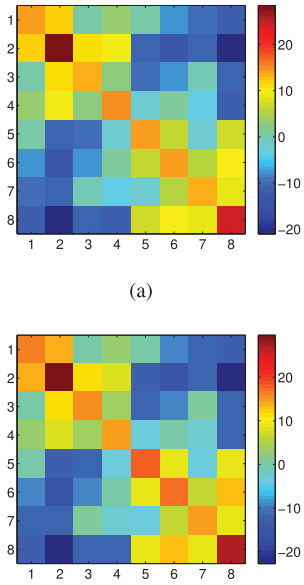


Fig. 6. The covariance matrices corresponding to the mulitvariate Gaussian conditional distributions (a). $p(\underline{y}|x_1)$ and (b). $p(\underline{y}|x_2)$ of subject S4 over n=8 ROIs, indicating the high correlation between the different regions.

Gaussian distributions, with $\mathbf{C_1}$ and $\mathbf{C_2}$ being the $n \times n$ covariance matrices of each of the distributions respectively. Fig. 6 shows the covariance matrices of a subject over $n = 8$ regions for a single trial. The conditional entropy $h(\underline{Y}|X)$ is then given by

$$\begin{aligned} h(\underline{Y}|X) &\triangleq -\sum_{x\in X} p(x) \int_{-\infty}^{\infty} p(\underline{y}|x)\log p(\underline{y}|x)\, d\mathbf{y} \\ &= \frac{1}{4}\left\{\log(2\pi e)^n|\mathbf{C_1}| + \log(2\pi e)^n|\mathbf{C_2}|\right\}, \end{aligned} \tag{4}$$

where we have used the closed-form expressions for the differential entropy of a multivariate Gaussian distribution [12, Theorem 8.4.1]. Using the law of total probability we can write $p(y)$ in terms of the conditional probabilities as

$$\begin{aligned} p(\underline{y}) &= \sum_{x\in X} p(\underline{Y}=\underline{y}|x)p(x) \\ &= \frac{1}{2}\left\{p(\underline{y}|x_1) + p(\underline{y}|x_2)\right\}. \end{aligned} \tag{5}$$

As both $p(\underline{y}|x_1)$ and $p(\underline{y}|x_2)$ are Gaussian distributed, the resulting random vector $\underline{y}$ is a multivariate Gaussian mixture with a distribution $p(\underline{y})$ given in (5). Therefore, the entropy of $\underline{Y}$ is given by

$$\begin{aligned} h(\underline{Y}) = -\frac{1}{2}\int_{\mathbb{R}^n} \Big\{\Big[p(\underline{y}|x_1) + p(\underline{y}|x_2)\Big] \\ \cdot \log\left(\frac{1}{2}\Big[p(\underline{y}|x_1) + p(\underline{y}|x_2)\Big]\right)\Big\}\, d\underline{y}. \end{aligned} \tag{6}$$

The MI between the output EEG data and the input audio stimulus can then be calculated using

$$I(X;\underline{Y}) = h(\underline{Y}) - h(\underline{Y}|X). \tag{7}$$

C. Entropy Approximation

The entropy of the output random variable (6) is the entropy of a mixture of two Gaussian distributions. It turns out, to the best of our knowledge, that there is no closed form solution for the entropy of a Gaussian mixture model (GMM) which is also a recurring open problem in the literature. The expression for the entropy of a GMM consists of a multiple integral over a logarithm of a sum of exponential functions which makes it difficult to formulate a general closed-form analytical solution. In the absence of an analytical solution it is usually common to use a numerical method to calculate a sufficiently accurate estimate of the solution. However, as the dimensionality of the Gaussian multivariate random variable increases, it becomes computationally infeasible to evaluate the n-order multiple integral using numerical integration. This drawback becomes even more apparent in our case where several test subjects each with multiple test sequences must be processed. In keeping with our primary goal of evaluating the overall information transfer over the ERP channel, we take a closer look at this open problem in the following subsections and propose two novel low-complexity methods to estimate the differential entropy of a GMM.

1) Single Region Entropy Bounds: In the following we derive new, tight upper and lower bounds for the entropy of a single ROI. The critical factor in the entropy calculation of the Gaussian mixture is the logarithm of the sum of individual distributions. By employing the Jacobian logarithm we obtain an alternate expression for the logarithm of a sum

$$\ln(e^a + e^b) = \max(a, b) + \ln(1 + e^{-|a-b|}), \tag{8}$$

where a and b are constants. The expression in (8) can be bounded as

$$\max(a, b) < \ln(e^a + e^b) \leq \max(a, b) + \ln 2\,, \tag{9}$$

where equality on the upper bound is achieved when $a = b$. Substituting

$$e^a \rightarrow \frac{1}{\sigma_1}\exp\frac{-y^2}{2\sigma_1^2}$$
$$\text{and} \quad e^b \rightarrow \frac{1}{\sigma_2}\exp\frac{-y^2}{2\sigma_2^2}\,, \tag{10}$$

we obtain the following bounds

$$\max\left(\frac{-y^2}{2\sigma_1^2} - \ln\sigma_1, \frac{-y^2}{2\sigma_2^2} - \ln\sigma_2\right) \tag{11}$$
$$< \ln\left(\frac{e^{\frac{-y^2}{2\sigma_1^2}}}{\sigma_1} + \frac{e^{\frac{-y^2}{2\sigma_2^2}}}{\sigma_2}\right) \tag{12}$$
$$\leq \max\left(\frac{-y^2}{2\sigma_1^2} - \ln\sigma_1, \frac{-y^2}{2\sigma_2^2} - \ln\sigma_2\right) + \ln\left(1 + \frac{\sigma_2}{\sigma_1}\right). \tag{13}$$

Here σ_1^2, σ_2^2 are the variances of the two Gaussian distributions $p(y|x_1)$ and $p(y|x_2)$, respectively, chosen such that $\sigma_2 > \sigma_1$. We can now use the lower bound in (11) to calculate the upper bound for the maximum entropy $h(Y)$ for a single ROI,

$$\begin{aligned} h(Y) &< -\frac{1}{2}\int_{-\infty}^{\infty}\left\{[p(y|x_1) + p(y|x_2)]\right. \\ &\qquad \left.\cdot\max\left(\frac{-y^2}{2\sigma_1^2} - \ln 2\sqrt{2\pi}\sigma_1\,,\ \frac{-y^2}{2\sigma_2^2} - \ln 2\sqrt{2\pi}\sigma_2\right)\right\} dy \\ &= \int_0^{\lambda}[p(y|x_1) + p(y|x_2)]\cdot\left(\frac{y^2}{2\sigma_1^2} + \ln 2\sqrt{2\pi}\sigma_1\right) dy \\ &\qquad + \int_{\lambda}^{\infty}[p(y|x_1) + p(y|x_2)]\cdot\left(\frac{y^2}{2\sigma_2^2} + \ln 2\sqrt{2\pi}\sigma_2\right) dy, \end{aligned} \tag{14}$$

where $\pm\lambda$ are the points of intersection of the two component Gaussian densities of the GMM. The points of intersection can be found by equating the two Gaussian densities and solving,

$$\begin{aligned} (1/\sqrt{2\pi}\sigma_1)e^{\frac{-\lambda^2}{2\sigma_1^2}} &= (1/\sqrt{2\pi}\sigma_2)e^{\frac{-\lambda^2}{2\sigma_2^2}} \\ \Longrightarrow \lambda &= \frac{\sigma_1^2\sigma_2^2\cdot\left(\ln\sigma_1^2 - \ln\sigma_2^2\right)}{\sigma_1^2 - \sigma_2^2}. \end{aligned} \tag{15}$$

The two integrals in (14) have a closed form solution and can be solved using the expressions for truncated normal distributions as

$$\begin{aligned} h(Y) &< \frac{1}{2}\left(\frac{-\lambda}{\sigma_1}Q\left(\frac{\lambda}{\sigma_1}\right) + \frac{1}{c_1}\right) + \ln(2\sqrt{2\pi}\sigma_1)\cdot\left[\frac{1}{c_1} + \frac{1}{c_2}\right] \\ &\quad + \frac{\sigma_2^2}{2\sigma_1^2}\left(\frac{-\lambda}{\sigma_2}Q\left(\frac{\lambda}{\sigma_2}\right) + \frac{1}{c_2}\right) \\ &\quad + \frac{\sigma_1^2}{2\sigma_2^2}\left(-\frac{\lambda}{\sigma_1}Q\left(\frac{\lambda}{\sigma_1}\right) + \frac{1}{d_1}\right) \\ &\quad + \frac{1}{2}\left(-\frac{\lambda}{\sigma_2}Q\left(\frac{\lambda}{\sigma_2}\right) + \frac{1}{d_2}\right) + \ln(2\sqrt{2\pi}\sigma_1)\cdot\left[\frac{1}{d_1} + \frac{1}{d_2}\right], \end{aligned} \tag{16}$$

where

$$\begin{aligned} c_1 &= \frac{1}{\Phi(\frac{\lambda}{\sigma_1}) - \Phi(0)}\,, \quad c_2 = \frac{1}{\Phi(\frac{\lambda}{\sigma_2}) - \Phi(0)}\,, \\ d_1 &= \frac{1}{1 - \Phi(\frac{\lambda}{\sigma_1})}\,, \qquad d_2 = \frac{1}{1 - \Phi(\frac{\lambda}{\sigma_2})}\,, \end{aligned} \tag{17}$$

and $\Phi(\cdot)$ is the standard normal cumulative density function, and $Q(\cdot)$ is the standard normal probability density function. A similar approach can be used to calculate a lower bound on the minimum entropy of $h(Y)$ by using the upper bound in (11).

This method is successful in providing tight bounds for the the entropy of a mixture of two scalar normal random variables, i.e., the entropy for an individual ROI. A limitation of this method however is that it is not effective for evaluating multivariate normal distributions and is therefore unsuitable for evaluating the entropy of multiple ROIs chosen simultaneously.

In particular, a mixture of a n-dimensional normal distribution will have intersections in a n-dimensional space and the integral in (14), if it even exists, will be extremely complex.

2) Multiple Region Entropy Approximation: In order to address these shortcomings for higher dimensional EEG outputs $\underline{Y}$ we propose an approximation for the entropy by performing a component-wise Taylor series expansion of the logarithmic function [43]. This approximation makes no prior assumptions and is suitable in general for estimating the entropy of any given multidimensional GMM.

Let $p_i(\underline{z}) \sim \mathcal{N}(\underline{z}; \underline{\mu}_i, \mathbf{C}_i)$ be the probability distribution for the i-th component of the GMM associated with the n-dimensional Gaussian distribution $\mathcal{N}(\underline{z}; \underline{\mu}_i, \mathbf{C}_i)$ with mean $\underline{\mu}_i \in \mathbb{R}^n$ and covariance matrix $\mathbf{C}_i \in \mathbb{R}^{n\times n}$. The probability distribution of the GMM is then given by

$$p(\underline{z}) = \sum_{i=1}^{L} w_i\, p_i(\underline{z}), \tag{18}$$

where L is the number of mixture components, and w_i denotes the weight of the i-th component of the GMM, respectively. Also, let

$$f(\underline{z}) = \log p(\underline{z}) = \log\left(\sum_{j=1}^{L} w_j\, p_j(\underline{z})\right). \tag{19}$$

If we then use the Taylor series to expand the function $f(\underline{z})$ around $\underline{\mu}_i$, we obtain

$$f(\underline{z}) = f(\underline{\mu}_i) + \frac{f'(\underline{\mu}_i)}{1!}(\underline{z}-\underline{\mu}_i) + \frac{f''(\underline{\mu}_i)}{2!}(\underline{z}-\underline{\mu}_i)^2 + \ldots \tag{20}$$

The odd central moments of a Gaussian distribution are zero [44], [45]. Therefore, all the odd order terms in the Taylor series expansion of the entropy of the Gaussian mixture are also zero, resulting in the following expansion

$$\begin{aligned} h(\underline{z}) &= -\int_{\mathbb{R}^n} p(\underline{z}) \log p(\underline{z})\, d\underline{z} = -\int_{\mathbb{R}^n} p(\underline{z}) f(\underline{z})\, d\underline{z} \\ &= -\sum_{i=1}^{L} \int_{\mathbb{R}^n} w_i\, p_i(\underline{z}) \cdot \left\{ f(\underline{\mu}_i) + \frac{f''(\underline{\mu}_i)}{2!}(\underline{z}-\underline{\mu}_i)^2 + \ldots \right\} d\underline{z}. \end{aligned} \tag{21}$$

The Gaussian mixture is a smooth function and the derivatives of the function $f(\underline{z})$ will therefore always exist. The Taylor series approximation is exact only if an infinite order of terms are considered, and the accuracy of the approximation therefore depends on the number of expansion terms used. We provide closed-form expressions for the zeroth-order, second-order, and fourth order terms of the Taylor series expansion in the Appendix at the end of this paper.

The Taylor series in (20) is expanded around the mean of the distribution. As the variance of the distribution increases, the data set is spread further away from the mean and as a result a larger number of higher-order terms are required to reduce the approximation error. Also, if one of the distributions has a relatively high variance compared to the other distributions in the mixture, then the corresponding n-dimensional pdf is skewed heavily along that particular axis.

Calculating a large number of higher order terms is complex and computationally demanding. In order to obtain a high accuracy approximation while keeping the number of expansion terms constant, we propose using a *variance splitting* approach [43], [46]. In this approach we split and replace the high variance Gaussian component with a mixture of Gaussians, each with a substantially lower variance than the original. The i-th component with $i = 1, 2, \ldots, L$ for splitting is identified and the corresponding covariance matrix $\mathbf{C}_i$ is diagonalized as

$$\mathbf{D}_i = \Sigma_i^T \mathbf{C}_i \Sigma_i\,, \tag{22}$$

such that

$$\mathbf{D}_i = \operatorname{diag}\left(\lambda_1^{(i)}, \lambda_2^{(i)}, \ldots, \lambda_n^{(i)}\right) \tag{23}$$

is the diagonal matrix containing the eigenvalues of the matrix $\mathbf{C}_i$, and Σ_i is the diagonalization matrix whose columns are the eigenvectors of $\mathbf{C}_i$. The i-th mixture component is then split along the principal axis of its longest ellipsoid into K sub-components as

$$w_i \cdot p_i(\underline{z}) = \sum_{k=1}^{K} w_k^{(i)} \cdot w_i \cdot p_k^{(i)}(\underline{z}), \tag{24}$$

where $\bar{w}_k, \bar{\mu}_k$, and $\bar{\sigma}_k$ are splitting parameters. Further, we define

$$\begin{aligned} &w_k^{(i)} \triangleq \bar{w}_k w_i \quad \text{with} \quad \sum_{k=1}^{K} w_k^{(i)} = w_i, \\ &\underline{\mu}_k^{(i)} \triangleq \underline{\mu}_i + \sqrt{\lambda_d^{(i)}} \cdot \bar{\mu}_k \cdot [0, \ldots, 1, \ldots, 0]^T, \\ &\mathbf{D}_k^{(i)} \triangleq \operatorname{diag}\left(\lambda_1^{(i)}, \ldots, \lambda_{d-1}^{(i)}, \bar{\sigma}_k^2, \lambda_{d+1}^{(i)}, \ldots, \lambda_n^{(i)}\right), \end{aligned} \tag{25}$$

where $\lambda_d^{(i)}$ is the largest eigenvalue of $\mathbf{C}_i$. The new covariance matrices, $\mathbf{C}_k^{(i)}$, for the sub-components are obtained by rotating back

$$\mathbf{C}_k^{(i)} = \Sigma_i \mathbf{D}_k^{(i)} \Sigma_i^T, \tag{26}$$

defining the distributions $p_k^{(i)}(\underline{z}) \sim \mathcal{N}(\underline{z}; \underline{\mu}_k^{(i)}, \mathbf{C}_k^{(i)})$. The parameters $\bar{w}_k, \bar{\mu}_k$ and $\bar{\sigma}_k$ are directly estimated using the splitting library shown in Table I. The library was generated for the standard Gaussian density (zero mean, unit variance) using the splitting technique used in [46], [47]. The chosen Gaussian component is first split into two mixture components starting with an initial guess

$$\begin{aligned} w \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\} \approx\ & \bar{w}_1 \exp\left\{-\frac{1}{2}\frac{(x-\bar{\mu}_1)^2}{\bar{\sigma}_1^2}\right\} \\ & + \bar{w}_2 \exp\left\{-\frac{1}{2}\frac{(x-\bar{\mu}_2)^2}{\bar{\sigma}_2^2}\right\}, \end{aligned} \tag{27}$$

TABLE II
SPLITTING LIBRARY FOR THE STANDARD GAUSSIAN DENSITY WITH ZERO MEAN AND UNIT VARIANCE [46], [47]

k	$\bar{w}_k$	$\bar{\sigma}_k$	$\bar{\mu}_k$
Two Way Split ($M = 2$)			
1	0.5000	0.7745125	-0.56520
2	0.5000	0.7745125	0.56520
Four Way Split ($M = 4$)			
1	0.127380	0.5175126	-1.41312
2	0.372619	0.5175126	-0.44973
3	0.372619	0.5175126	0.44973
4	0.127380	0.5175126	1.41312

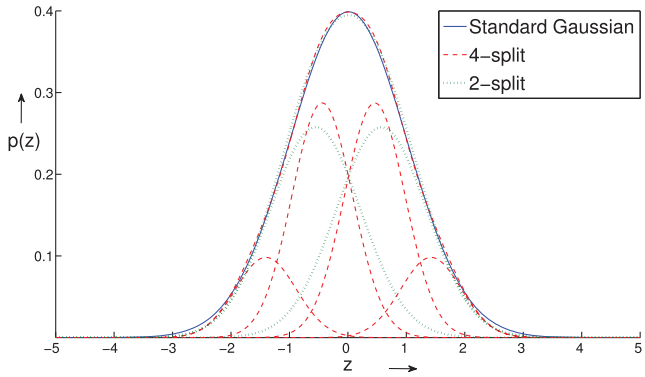


Fig. 7. Variance Split of a Standard Gaussian Using the Splitting Library in Table II.

where

$$\bar{\mu}_1 = \mu - \epsilon, \bar{\mu}_2 = \mu + \epsilon, \bar{\sigma}_1 = \bar{\sigma}_2 = \sigma, \bar{w}_1 = \bar{w}_2 = w/2, \tag{28}$$

and $\epsilon = 0.001$ is a small constant. A least-squares error minimization is then performed wherein the standard deviation of the split components are recursively reduced while adapting the means and weights accordingly. When no further optimization is possible, the resulting two mixture densities can again be split according to (27) yielding four mixture components. The results of using this library for both a two-way and a four-way-split of the standard Gaussian are depicted in Fig. 7. While maintaining a high degree of accuracy, this split is not perfect and introduces a marginal amount of error. In theory, it would require an infinite sum of Gaussian distributions on the r.h.s. of (27) to exactly represent the single Gaussian distribution on the l.h.s. of (27).

Fig. 8 exemplarily shows the simulation results for the entropy approximation of a sample GMM with $n = 2$ as its variance is increased, calculated by using a fourth-order Taylor series expansion, with and without the variance split. The results indicate an increase in the overall accuracy of the final entropy estimate when using the variance split. We have observed that the splitting approach is especially helpful at higher variances and, if required, can be performed repeatedly to further refine the approximation.

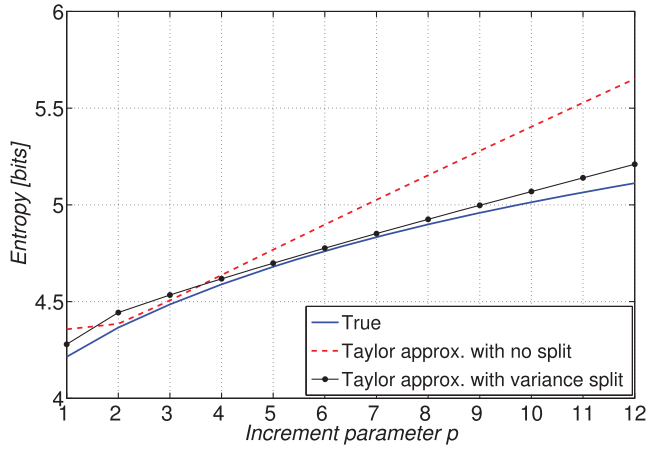


Fig. 8. Entropy approximation using a fourth-order Taylor expansion for a sample synthetic GMM with zero mean vectors, equal weights, and $\mathbf{C}_1 = [4, 2; 2, 4]$, $\mathbf{C}_2 = [5 + p, 2; 2, 5 + p]$, where the variance of the second distribution is incremented using the parameter p. The true entropy was calculated using numerical integration which in this case for $n = 2$ is still computationally tractable.

TABLE III
SINGLE REGION ENTROPY BOUNDS AND MUTUAL INFORMATION FOR SUBJECT S4

ROI	Entropy	Lower Bound	Upper Bound	MI
1	3.2512	2.9676	3.5549	0.0464
2	3.4257	3.1781	3.7633	0.0484
3	3.2650	3.0493	3.6128	0.0727
4	3.0395	2.7990	3.3828	0.0498
5	3.3441	3.1937	3.7105	0.1481
6	3.1016	2.8841	3.4354	0.0427
7	3.3437	3.1089	3.6840	0.0890
8	3.5481	3.2253	3.8623	0.0121

IV. MUTUAL INFORMATION RESULTS

The MI can be trivially upper bounded as $I(X; \underline{Y}) \leq H(X) = 1$ bit, where we have used the fact that X is drawn from an equiprobable Bernoulli distribution. This upper bound depends only on the quality of the audio stimulus, independently of the subject, the audio sequence, the pre-processing of the EEG data, or the number of ROIs considered. Therefore, the maximum information that can be transferred over the ERP channel for the given input is 1 bit.

A. Single Region

We first evaluate the MI over a single region. Choosing only one ROI at a time corresponds to $\underline{Y}$ being univariate, i.e., $n = 1$. Table III lists the MI results calculated for each ROI for one test subject. Also shown are the lower and upper entropy bounds obtained using the Jacobian Log bound introduced in Section III-C1.

The output of the ERP channel maps the activity spread over the entire cortex distributed over all 8 ROIs. Considering only a single region then ignores the fact that the ROIs are coupled due to the shift of neural activity between the regions. Therefore,

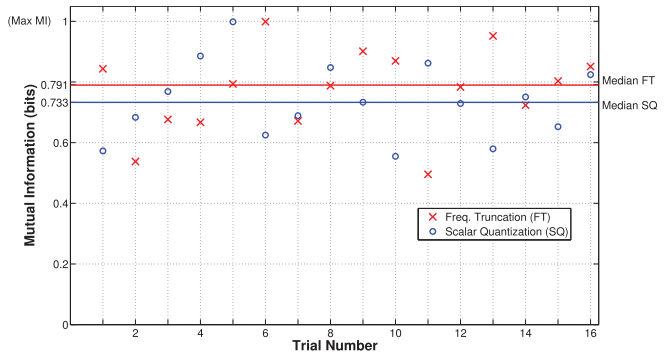


Fig. 9. MI estimates for subject S4 for the two considered distortion types, conducted using different combinations of base sequences, and time-varying distortion patterns.

the available information is severely limited when only a single ROI is considered. It is interesting to note however, that the temporal regions 5 and 7 show a higher MI than the other regions. This is in strong agreement with the fact that the the primary auditory cortex which is involved with the perception and processing of auditory information, is located in the temporal lobe.

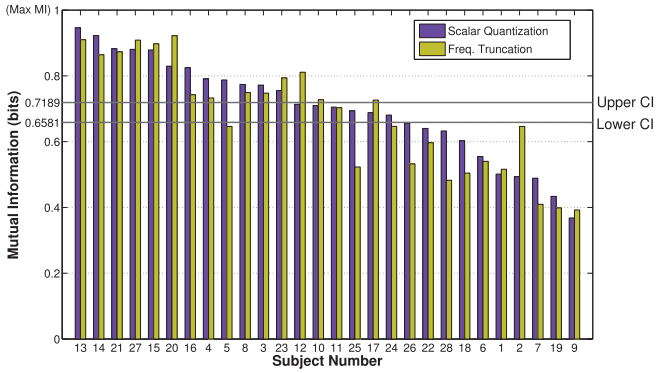
B. Multiple Regions

To calculate the entropy over multiple regions we use the fourth-order Taylor series approximation presented in Section III-C2. A four-component variance split is performed twice to further refine the approximation result. Fig. 9 shows the final MI estimates of a single subject for each trial using different combinations of the base music sequences and time-varying distortion patterns. Also, calculated and shown in the same figure is the median MI across all trials for each of the two distortion types. The median is chosen here to remove any outliers among the trials. We observe that the MI over the ERP channel for a given trial is in general moderate to high. Further, as one would expect, there is a significant increase in the MI calculated over multiple ROIs, when compared to the single ROI case.

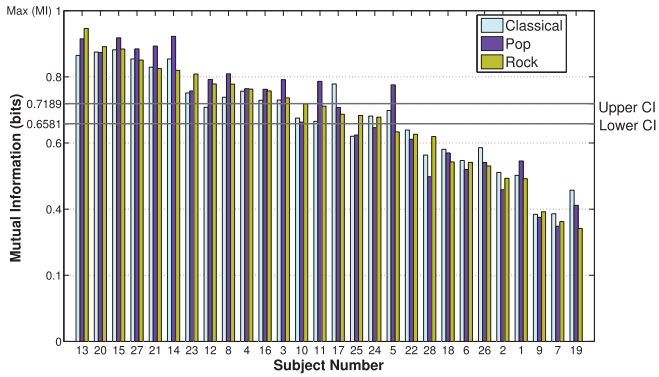
The median individual MI estimates for different distortions and each subject are summarized in Fig. 10(a). We notice that in general for a given subject there is only a minor variation in the median MI between the different distortion types. Further, Fig. 10(b) shows the median MI of each subject calculated across trials for different base music sequences used. By comparing Fig. 10(a) and Fig. 10(b) we observe a similar trend, and subjects that have a high MI in one experiment also have a high MI in the other. Overall, the MI is found to be consistent across the entire pool of test subjects across different trials, distortion-types, and music-types.

C. Confidence Intervals for the MI Using Bootstrapping

The bootstrap [48], [49] is a powerful resampling technique for assessing the accuracy of a parameter estimate (in this case the median). It is especially useful in making asymptotically-consistent statistical inferences when only a limited number



(a) Different distortion types



(b) Different music types

Fig. 10. Ordered median MI estimates for each of the 28 test subjects when presented with the same set of trial-sequences using a combination of different distortion types, patterns, and music sequences. Confidence intervals (CI) for the median are obtained using bootstrapping [48], [49].

of sample realizations are available. We derive confidence intervals (CI) for the median of the MI estimates using the bootstrap-t percentile method [50]. It has been shown [51] that the bootstrap-t percentile intervals have a higher order accuracy as compared to regular bootstrapping by better accounting for the skewness and other errors in the original population.

The original sample set consists of the actual MI estimates calculated for each of the trials. The median of this sample is denoted by $\overline{M}$. Bootstrap resamples are then created by drawing samples, with replacement, from a population of this original sample set. The median for each of this resampled set is calculated and denoted as $\overline{M}_b^*$ with $b = 1, 2, \ldots, B$. Generally, B is a large number in the order of several hundred. For our calculations here we use $B = 1000$ resamples.

The t-parameter is defined by normalizing the difference between the original sample statistics (in this case the median) and the resampled statistics over the standard error

$$t_b^* = \frac{\overline{M} - \overline{M}_b^*}{\sigma^*/\sqrt{s}}, \tag{29}$$

where σ^{*2} is the sample variance and s is the number of elements in each resampled set. The empirical standard deviation σ^* is itself calculated by performing a nested bootstrap from resamples of $\overline{M}_b^*$.

TABLE IV
MI CONFIDENCE INTERVALS ON THE ERP CHANNEL IN BITS, OBTAINED USING BOOTSTRAPPING

Type	Lower CI	Upper CI
Distortion		
Freq. Truncation	0.6162	0.6937
Scalar Quant.	0.6625	0.7087
Music		
Rock	0.6594	0.7286
Pop	0.6352	0.7321
Classic	0.6509	0.7101
Overall	**0.6581**	**0.7189**

We therefore get a total of B values of the t-parameter $(t_1^*, \ldots, t_b^*, \ldots, t_B^*)$, from which we calculate the required percentile cutoffs of our CIs. For our analysis, we take the 2.5th and 97.5th percentile cutoffs and calculate the lower and upper CI, respectively, as

$$\begin{aligned} \mathrm{CI}_{\mathrm{low}} &= \overline{M} - t_{25}\sigma^*, \\ \mathrm{CI}_{\mathrm{high}} &= \overline{M} - t_{975}\sigma^*. \end{aligned} \tag{30}$$

Table IV lists the confidence intervals for different distortion and music types. Both upper and lower CIs are also included in Fig. 10.

V. DISCUSSION

MI quantifies the amount of information the recorded EEG contains about the quality of the input audio-stimulus. By measuring how aligned the brain activity is in response to the stimulus we thereby directly quantify the subject's perception with respect to the change in audio quality with time. A high MI therefore indicates a large correlation between the actual and the perceived audio quality. For example, the fact that subjects 13 and 27 have a high median MI in Fig. 10 implies that these subjects have a strong perception towards quality change irrespective of the music or distortion type used in the test sequence.

However, brain dynamics are fundamentally multiscale in nature and it is often difficult to draw a one-to-one relation between the neural electrical field dynamics and the cognitive/behavioral state [52]. The EEG activity is dependent on the psychological state of the subject including concentration, motivation, and fatigue. As a consequence, the recorded EEG activity varies considerably not only over the duration of a trial, but also from one trial to another, which could result in low MI, for example as it is the case with subjects 7, 9, and 19.

It is also important to note the significance of choosing the ROIs. The output EEG activity recorded over the ERP channel is dependent on the number of regions, the size (i.e., the number of individual electrodes within each ROI), and the location of the ROIs on the scalp. Increasing the number of regions increases complexity whereas considering too few oversimplifies the problem, as demonstrated by the single ROI results in Table III. In our work we select the regions based on the cortical lobes (see Section II-E). The ROI selection could possibly be improved by identifying the distribution of virtual sources inside the brain which generate the EEG activity on the scalp. A source localization of the EEG signal is commonly referred to as the *inverse problem* and while several solutions have been suggested to varying degrees of accuracy [53] it continues to be an undetermined problem [52].

VI. CONCLUSION

In this work we have presented a new framework for audio quality assessment based on observed EEG data for subjects listening to time-varying distorted audio. By modeling the end-to-end perceptual processing chain as a discrete-input time-varying nonlinear SIMO channel with memory, we are able to quantify how the subjective perception changes if the observed audio quality changes with time. We have focused on the simplest binary-input SIMO case where the audio quality can only change between distorted and undistorted levels. The MI results demonstrate the practical applicability of this approach and provide a performance measure for the information transfer over the ERP channel, which can be computed via a new low complexity algorithm. This algorithm is based on a fast approximation technique for the differential entropy of a multidimensional GMM. The proposed approach can be extended to video quality perception and to other neuroimaging assessment techniques like MEG, albeit with a higher complexity. As potential future work the proposed information theoretic framework can be exploited to develop an inference model for blind audio quality classification based on perception.

APPENDIX

In this appendix we derive the Taylor series expansion of the entropy up to its fourth order component. The Taylor series expansion of the entropy for the GMM can be written as

$$\begin{aligned} h(\underline{z}) &= -\sum_{i=1}^{L} \int_{\mathbb{R}^n} w_i\, p_i(\underline{z}) \left\{ f(\underline{\mu}_i) + \frac{f''(\underline{\mu}_i)}{2!}(\underline{z} - \underline{\mu}_i)^2 + \ldots \right\} d\underline{z} \\ &= h_0 + h_2 + h_4 + R\,, \end{aligned} \tag{31}$$

where h_0, h_2, h_4 are the zeroth, second and fourth order expansion terms respectively, and R is the residual sum term representing all the remaining higher order terms.

A. Zeroth-Order Term

The zeroth-order Taylor series expansion term h_0 is given by

$$\begin{aligned} h_0 &= -\sum_{i=1}^{L} \int_{\Re^n} w_i\, p_i(\underline{z})\, f(\underline{\mu}_i)\, d\underline{z} \\ &= -\sum_{i=1}^{L} w_i \log p(\underline{\mu}_i). \end{aligned} \tag{32}$$

which follows from the fact that the integral over the entire probability distribution is 1.

B. Second-Order Term

The second-order expansion term requires us to calculate the gradient g_i and Hessian H_i of the Gaussian probability distribution $p_i(\underline{z}) \sim \mathcal{N}(\underline{z}; \underline{\mu}_i, \mathbf{C}_i)$, which are respectively defined as

$$\mathsf{g}_i \triangleq \nabla p_i(\underline{z}) = p_i(\underline{z})\mathbf{C}_i^{-1}(\underline{\mu}_i - \underline{z}), \tag{33}$$

$$\mathsf{H}_i \triangleq (\nabla\nabla^T)p_i(\underline{z}) = \frac{\mathsf{g}\mathsf{g}^T}{p_i(\underline{z})} - p_i(\underline{z})\mathbf{C}^{-1}, \tag{34}$$

where ∇ denotes the vector differential operator and $i = 1, 2, \ldots, L$. Accordingly, the gradient $\mathfrak{g}_i$ and Hessian $\mathbf{H}_i$ for the log-density function $f(\underline{z})$ can be defined as

$$\mathfrak{g}_i \triangleq \nabla f(\underline{z}) = \nabla \log p_i(\underline{z}) = \frac{1}{p_i(\underline{z})}\mathsf{g}_i, \tag{35}$$

$$\mathbf{H}_i \triangleq (\nabla\nabla^T) f(\underline{z}) = -\frac{1}{p_i(\underline{z})^2}\mathsf{g}_i\mathsf{g}_i^T + \frac{1}{p(\underline{z})}\mathsf{H}_i. \tag{36}$$

We can then calculate the second-order term as

$$\begin{aligned} h_2 &= -\sum_{i=1}^{L}\int_{\mathbb{R}^n} w_i\, p_i(\underline{z}) \cdot \frac{1}{2}\mathbf{H}_i \cdot (\underline{z} - \underline{\mu}_i)(\underline{z} - \underline{\mu}_i)^T\, d\underline{z} \\ &= -\frac{1}{2}\sum_{i=1}^{L} w_i \cdot \mathbf{H}_i \circ \mathbf{C}_i|_{\underline{z}=\underline{\mu}_i} \quad . \end{aligned} \tag{37}$$

The symbol $\circ$ denotes the Frobenius product $\mathbf{A} \circ \mathbf{B} = \sum_i \sum_j a_{ij} \cdot b_{ij}$.

C. Fourth Order Term

The third order partial derivative with respect to $p_i(\underline{z})$, $i = 1, 2, \ldots, L$, is given by

$$\mathsf{T}_i \triangleq \mathsf{H}_i \otimes \left(\mathbf{C}_i^{-1}(\underline{z} - \underline{\mu}_i)\right) - \mathsf{g}_i \otimes \mathbf{C}_i^{-1} - \#\mathbf{C}_i^{-1}\mathsf{g}_i^T, \tag{38}$$

where $\mathbf{A}\otimes\mathbf{B}$ is the Kronecker product of the matrices $\mathbf{A}$, $\mathbf{B}$, and $\#$ is the matrix vectorization operator. The vectorization of a matrix is the linear transformation of converting a $m \times n$ matrix into a $mn \times 1$ column vector by stacking the columns of the matrix on top of each other. Further, partial differentiating the Hessian twice yields the fourth order derivative with respect to $p_i(\underline{z})$,

$$\begin{aligned} \mathsf{F}_i &\triangleq (\nabla\nabla^T)\mathsf{H}_i \\ &= \mathsf{H}_i \otimes \mathbf{C}_i^{-1} + \mathbf{C}_i^{-1}(\underline{z} - \underline{\mu}_i) \otimes \mathsf{T}_i^T \\ &\quad + \mathbf{C}_i^{-1} \otimes \mathsf{H}_i + \mathbf{C}_i^{-1} \otimes \mathsf{H}_i^T. \end{aligned} \tag{39}$$

Similarly then, the third and fourth order partial differential matrix for the log-density $f(\underline{z})$ are respectively defined as

$$\begin{aligned} \mathbf{T}_i \triangleq\ & \frac{1}{p_i(\underline{z})^3}\left\{\mathsf{g}_i \otimes \mathsf{g}_i\mathsf{g}_i^T\right\} - \frac{1}{p_i(\underline{z})^2}\left\{\mathsf{H}_i \otimes \mathsf{g}_i + \mathsf{g}_i \otimes \mathsf{H}_i\right\} \\ &+ \frac{1}{p_i(\underline{z})}\left\{\mathsf{T}_i - \#\mathbf{C}_i^{-1}\mathsf{g}_i^T\right\}, \end{aligned} \tag{40}$$

$$\begin{aligned} \mathbf{F}_i \triangleq\ & \frac{1}{p_i(\underline{z})^4}\left\{\mathsf{g}_i^T\mathsf{g}_i \otimes \mathsf{g}_i\mathsf{g}_i^T\right\} \\ &- \frac{1}{p_i(\underline{z})^3}\left\{\mathsf{H}_i \otimes \mathsf{g}_i\mathsf{g}_i^T + \mathsf{g}_i \otimes \mathsf{H}_i \otimes \mathsf{g}_i^T\right\} \\ &+ \frac{1}{p_i(\underline{z})^2}\left\{2\mathsf{g}_i \otimes \mathsf{T}_i^T - (\#\mathbf{C}_i^{-1}\mathsf{g}_i^T) \otimes \mathsf{g}_i^T\right\} \\ &+ \frac{1}{p_i(\underline{z})}\left\{\mathsf{F}_i - 2\mathsf{H}_i \otimes \mathbf{C}_i^{-1} - \mathsf{T}_i \otimes \mathsf{g}_i^T + \mathbf{C}_i^{-1} \otimes \mathsf{H}_i^T\right\}. \end{aligned} \tag{41}$$

The fourth order term is then given by

$$\begin{aligned} h_4 &= -\sum_{i=1}^{L}\int_{\mathbb{R}^n} w_i\, p_i(\underline{z}) \frac{f''''(\underline{\mu}_i)}{4!}(\underline{z} - \underline{\mu}_i)^4\, d\underline{z} \\ &= -\frac{1}{24}\sum_{i=1}^{L} w_i \cdot \mathbf{F}_i \circ (\mathbf{C}_i \otimes \mathbf{C}_i^T)|_{\underline{z}=\underline{\mu}_i} \quad . \end{aligned} \tag{42}$$

ACKNOWLEDGMENT

The authors would like to thank Dr. James Kroger from the Department of Psychology, New Mexico State University (NMSU) for his expertise and assistance with EEG data acquisition and analysis, and also Dr. Charles Creusere from the Klipsch School of Electrical Engineering, NMSU for his insight and stimulating discussions that greatly assisted the research.

REFERENCES

[1] K. Mehta and J. Kliewer, "Assessing subjective perception of audio quality by measuring the information flow on the brain-response channel," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Florence, Italy, May 2014, pp. 5884–5888.
[2] F. Rieke, D. Warland, R. De Ruyter van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code*. Cambridge, MA, USA: MIT Press, 1999.
[3] 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.
[4] D. H. Johnson, "Information theory and neural information processing," *IEEE Trans. Inf. Theory*, vol. 56, no. 2, pp. 653–666, Feb. 2010.
[5] M. Aghagolzadeh, S. Eldawlatly, and K. Oweiss, "Synergistic coding by cortical neural ensembles," *IEEE Trans. Inf. Theory*, vol. 56, no. 2, pp. 875–889, Feb. 2010.
[6] D. Ostwald, C. Porcaro, and A. P. Bagshaw, "An information theoretic approach to EEG–fMRI integration of visually evoked responses," *Neuroimage*, vol. 49, no. 1, pp. 498–516, 2010.
[7] D. Ostwald, C. Porcaro, S. D. Mayhew, and A. P. Bagshaw, "EEG-fMRI based information theoretic characterization of the human perceptual decision system," *PLoS ONE*, vol. 7, no. 4, p. e33896, 2012.
[8] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 623–656, 1948.
[9] R. B. Stein, "The information capacity of nerve cells using a frequency code," *Biophys. J.*, vol. 7, no. 6, pp. 797–826, 1967.
[10] A. Rapoport and W. J. Horvath, "The theoretical channel capacity of a single neuron as determined by various coding systems," *Inf. Control*, vol. 3, no. 4, pp. 335–350, 1960.
[11] W. S. McCulloch, "An upper bound on the informational capacity of a synapse," in *Proc. ACM Nat. Meeting*, 1952, pp. 113–117.
[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2012.
[13] A. Borst and F. E. Theunissen, "Information theory and neural coding," *Nat. Neurosci.*, vol. 2, no. 11, pp. 947–957, 1999.
[14] L. Paninski, "Estimation of entropy and mutual information," *Neural Comput.*, vol. 15, no. 6, pp. 1191–1253, 2003.
[15] S. Panzeri and S. R. Schultz, "A unified approach to the study of temporal, correlational, and rate coding," *Neural Comput.*, vol. 13, no. 6, pp. 1311–1349, 2001.

[16] S. Panzeri, R. S. Petersen, S. R. Schultz, M. Lebedev, and M. E. Diamond, "The role of spike timing in the coding of stimulus location in rat somatosensory cortex," *Neuron*, vol. 29, no. 3, pp. 769–777, 2001.
[17] A. G. Dimitrov, J. P. Miller, T. Gedeon, Z. Aldworth, and A. E. Parker, "Analysis of neural coding through quantization with an information-based distortion measure," *Netw. Comput. Neural Syst.*, vol. 14, no. 1, pp. 151–176, 2003.
[18] S. Aviyente, L. A. Brakel, R. K. Kushwaha, M. Snodgrass, H. Shevrin, and W. J. Williams, "Characterization of event related potentials using information theoretic distance measures," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 5, pp. 737–743, May 2004.
[19] R. K. Kushwaha and W. J. Williams, "An information flow technique for category related evoked potentials," *IEEE Trans. Biomed. Eng.*, vol. 39, no. 2, pp. 165–175, Feb. 1992.
[20] 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.
[21] A. Temko, G. Boylan, W. Marnane, and G. Lightbody, "Speech recognition features for EEG signal description in detection of neonatal seizures," in *Proc. 32nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. Eng. Med. Biol. Soc., Buenos Aires, Argentina*, 2010, pp. 3281–3284.
[22] J. Xu, Z. Liu, R. Liu, and Q. Yang, "Information transmission in human cerebral cortex," *Phys. D Nonlinear Phenom.*, vol. 106, pp. 363–374, 1997.
[23] 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, pp. 827–835, 2001.
[24] S. H. Na, S. Jina, S. Y. Kima, and B. Hamb, "EEG in schizophrenic patients: Mutual information analysis," *Clin. Neurophysiol.*, vol. 113, pp. 9154–1960, 2002.
[25] H. Harada, Y. Eura, K. Shiraishi, T. Kato, and T. Soda, "Coherence analysis of EEG changes during olfactory stimulation," *Clin. Electroencephalogr.*, vol. 29, no. 2, pp. 96–100, 1998.
[26] B.-C. Min *et al.*, "Mutual information analysis of EEG responses by odor stimulation," *Chem. Senses*, vol. 28, pp. 741–749, 2003.
[27] Z. J. Wang, P. W. Lee, and M. J. McKeown, "A novel segmentation, mutual information network framework for EEG analysis of motor tasks," *Biomed. Eng. Online*, vol. 8, no. 9, pp. 1–19, 2009.
[28] ITU-R, "Method for subjective assessment of intermediate quality levels of coding systems," *Question ITU-R 220/10, Recommendation ITU-R BS.1534-1, Geneva, Switzerland*, 2001–2003.
[29] 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.
[30] C. D. Creusere, J. Kroger, S. R. Siddenki, P. Davis, and J. Hardin, "Assessment of subjective audio quality from EEG brain responses using time-space-frequency analysis," in *Proc. 20th Eur. Signal Process. Conf., Bucharest, Romania*, 2012, pp. 2704–2708.
[31] 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.
[32] L. Lindemann, S. Wenger, and M. Magnor, "Evaluation of video artifact perception using event-related potentials," in *Proc. ACM SIGGRAPH Symp. Appl. Percept. Graph. Visual., Toulouse, France*, 2011, pp. 53–58.
[33] M. Mustafa, L. Lindemann, and M. Magnor, "EEG analysis of implicit human visual perception," in *Proc. SIGCHI Conf. Human Factors Comput. Syst., Austin, TX, USA*, 2012, pp. 513–516.
[34] M. Mustafa, S. Guthe, and M. Magnor, "Single-trial EEG classification of artifacts in videos," *ACM Trans. Appl. Percept.*, vol. 9, no. 3, pp. 12:1–12:15, 2012.
[35] T. W. Picton, "The P300 wave of the human event-related potential," *J. Clin. Neurophysiol.*, vol. 9, no. 4, pp. 456–479, 1992.
[36] S. Sutton, P. Tueting, and J. Zubin, "Information delivery and the sensory evoked potential," *Science*, vol. 155, pp. 1436–1439, 1967.
[37] P. Nunez and R. Srinivasan, *Electric Fields of the Brain: The Neurophysics of EEG*. London, U.K.:: Oxford Univ. Press, 2006.
[38] J. J. Vidal, "Real-time detection of brain events in EEG," *Proc. IEEE*, vol. 65, no. 5, pp. 633–641, May 1977.
[39] C. D. Creusere and J. C. Hardin, "Assessing the quality of audio containing temporally varying distortions," *IEEE Trans. Audio Speech Lang. Process.*, vol. 19, no. 4, pp. 711–720, 2011.
[40] W. Freeman and R. Q. Quiroga, *Imaging Brain Function With EEG: Advanced Temporal and Spatial Analysis of Electroencephalographic Signals*. New York, NY, USA: Springer, 2012.
[41] T. Gruber, N. J. Trujillo-Barreto, C.-M. Giabbiconi, P. A. Valdés-Sosa, and M. M. Müller, "Brain electrical tomography (BET) analysis of induced gamma band responses during a simple object recognition task," *NeuroImage*, vol. 29, no. 3, pp. 900–888, 2006.
[42] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, 1999.
[43] M. F. Huber, T. Bailey, H. Durrant-Whyte, and U. D. Hanebeck, "On entropy approximation for Gaussian mixture random vectors," in *Proc. IEEE Int. Conf. Multisensor Fusion Integr. Intell. Syst., Seoul, Soth Korea*, 2008, pp. 181–188.
[44] J. J. Shynk, *Probability, Random Variables, and Random Processes: Theory and Signal Processing Applications*. Hoboken, NJ, USA: Wiley, 2012.
[45] I. Reed, "On a moment theorem for complex Gaussian processes," *IRE Trans. Inf. Theory*, vol. 8, no. 3, pp. 194–195, Apr. 1962.
[46] U. D. Hanebeck, K. Briechle, and A. Rauh, "Progressive Bayes: A new framework for nonlinear state estimation," in *Proc. SPIE Multisensor Multisource Inf. Fusion Archit. Algorithms Appl.*, Apr. 2003, pp. 256–267.
[47] A. Sankar, "Experiments with a Gaussian merging-splitting algorithm for HMM training for speech recognition," in *Proc. DARPA Speech Recognit. Workshop*, 1998, pp. 99–104.
[48] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. Boca Raton, FL, USA: CRC Press, 1994, vol. 57.
[49] B. Efron, "Nonparametric standard errors and confidence intervals," *Can. J. Statist.*, vol. 9, no. 2, pp. 139–158, 1981.
[50] A. M. Zoubir and B. Boashash, "The bootstrap and its application in signal processing," *IEEE Signal Process. Mag.*, vol. 15, no. 1, pp. 56–76, Jan. 1998.
[51] D. N. Politis, "Computer-intensive methods in statistical analysis," *IEEE Signal Process. Mag.*, vol. 15, no. 1, pp. 39–55, Jan. 1998.
[52] S. Makeig, C. Kothe, T. Mullen, N. Bigdely-Shamlo, Z. Zhang, and K. Kreutz-Delgado, "Evolving signal processing for brain–computer interfaces," *Proc. IEEE*, vol. 100, pp. 1567–1584, May 2012.
[53] R. Grech *et al.*, "Review on solving the inverse problem in EEG source analysis," *J. Neuroeng. Rehabil.*, vol. 5, no. 1, p. 25, 2008.

Ketan Mehta received the B.E. degree in electronics and communication from Visvesvaraya Technological University, Belgaum, India, and the M.S. degree in electrical engineering from New Mexico State University (NMSU), Las Cruces, NM, USA, in 2008 and 2010, respectively. He is currently pursuing the Ph.D. degree at NMSU. Since 2012, he has been a Research Assistant with NMSU. His research interests include information theory and signal processing with included applications in compression, optical communication, and cognitive neuroscience.

Jörg Kliewer (S'97–M'99–SM'04) received the Dipl.-Ing. (M.Sc.) degree in electrical engineering from Hamburg University of Technology, Hamburg, Germany, and the Dr.-Ing. degree (Ph.D.) in electrical engineering from the University of Kiel, Kiel, Germany, in 1993 and 1999, respectively. From 1993 to 1998, he was a Research Assistant with the University of Kiel, and from 1999 to 2004, he was a Senior Researcher and a Lecturer with the same institution. In 2004, he visited the University of Southampton, Southampton, U.K., for 1 year, and from 2005 to 2007, he was with the University of Notre Dame, 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 now with the New Jersey Institute of Technology, Newark, NJ, USA, as an Associate Professor. His research interests include coding and information theory, graphical models, and statistical algorithms, which includes applications to networked communication and security, data storage, and biology. 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 member of the Editorial Board of the IEEE Information Theory Society Newsletter since 2012. He was the recipient of the Leverhulme Trust Award and the German Research Foundation Fellowship Award in 2003 and 2004, respectively.