5-31-2020

An information theoretic approach to assess perceived audio quality using eeg with reduced number of electrodes

Sansit Das
New Jersey Institute of Technology

ABSTRACT

AN INFORMATION THEORETIC APPROACH TO ASSESS PERCEIVED AUDIO QUALITY USING EEG WITH REDUCED NUMBER OF ELECTRODES

by
Sansit Das

Electroencephalograph(EEG) is a process mainly used in medical and research fields to study the electrical activities in a brain. In this technique, 128 or 256 electrodes are attached to the scalp and the electrical activities of the human brain is recorded with the help of a software. In the global scenario, the EEG responses are studied and analysed to acknowledge any disorders in the brain, such as epilepsy or head injury.

Recent studies performed by researchers, have focused on analysing these electrical activities to access perceived audio quality from users by using information theoretic approaches, such as mutual information. Experiments were conducted by using a total of 128 electrodes, comprised in 8 regions of interests. In the research, each region is considered to have 9 electrodes each. The aim of this thesis was to build on previous research, and to reduce the use of the number of electrodes, so as to reduce the cost and complexity of the setup. In order to achieve this, first both the good and the bad quality audios were analyzed, and a receiver operating characteristic curve is plotted to draw a classification. Furthermore, a combination of different regions were taken and their mutual information were calculated, in order to check which group of regions give us the best result. This test is undertaken for each available combination of 2 or 3 regions. The classification accuracy was obtained by a variance detector approach and the accuracy was verified by computing z scores.

AN INFORMATION THEORETIC APPROACH TO ASSESS PERCEIVED AUDIO QUALITY USING EEG WITH REDUCED NUMBER OF ELECTRODES

by
Sansit Das

A Thesis
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Electrical Engineering

Helen and John C. Hartmann Department of Electrical and Computer Engineering

May 2020

Blank Page

APPROVAL PAGE

AN INFORMATION THEORETIC APPROACH TO ASSESS PERCEIVED AUDIO QUALITY USING EEG WITH REDUCED NUMBER OF ELECTRODES

Sansit Das

Dr. Joerg Kliewer, Thesis Co-Advisor — Date
Professor of Electrical and Computer Engineering, NJIT

Dr. Antje Ihlefeld, Thesis Co-Advisor — Date
Assistant Professor of Bio-Medical Engineering, NJIT

Dr. Ali Abdi, Committee Member — Date
Professor of Electrical and Computer Engineering, NJIT

BIOGRAPHICAL SKETCH

Author: Sansit Das

Degree: Master of Science

Date: May 2020

Undergraduate and Graduate Education:

- Master of Science in Electrical Engineering,
  New Jersey Institute of Technology, Newark, NJ, 2020
- Bachelor of Science in Electronics and Communication Engineering,
  Institute of Technical Education and Research, Bhubaneswar, Odisha, India, 2015

Major: Electrical Engineering

When I was weary, You held me strong
When I broke down, You carried me along
When I was alone, You were there for me
Always in tight embrace and ecstasy

When I was in lonely tears, You cheered me up
When it felt like no one would understand, You lighted me up
You were always there for me
In constant support and prayers

I dedicate this work to my parents for their constant care, guidance and prayers.

ACKNOWLEDGMENT

I would like to express my sincere thanks to my thesis advisors Dr. Joerg Kliewer and Dr. Antje Ihlefeld for their generous support and continuous guidance all the time throughout my research. Without their guidance and persistent help, this thesis would not have been possible.

I take this opportunity to thank my committee member, Dr. Ali Abdi who graciously agreed to serve on my committee.

I also thank Min Zhang for helping me set up EEG equipment for my research.

I would like to express my gratitude to my parents, Er. Bibhudutta Das and Mousumi Sahu, for their unfailing support, encouragement and prayers to pursue my goals not only through the process of research but throughout all my years of study.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

It is a general understanding that the human brain acts as an information processing device like a computer. It is known that the human brain works in a set of sequences, namely receiving the input, processing the information and delivering the output. Humans process the information they receive, rather than merely responding to stimuli. It is this perspective that equates a human brain to a computer, which is responsible for analyzing information from the environment. Information is taken in through the senses, the information is then put through the short-term memory. The information is then encoded to the long term memory, where the information is then stored. The information can be retrieved when necessary.

According to neurobiology, this process is carried out by the neurons or nerve cells present in the human brain. Human brain mainly consists of millions of neurons which are playing an important role for controlling behavior of human body with respect to internal, external motor and sensory stimuli. The neurons interact with each other to communicate and process information based on which we see, hear, move, think, make decisions and generally function. The central nervous system and its network of neurons are at the heart of all the activity that happens in the body. Understanding cognitive behaviour of brain can be done by analyzing either signals or images from the brain.

Electroencephalography(EEG) is among the best techniques used to record, analyse and interpret these signals or brain-wave patterns using an Electroencephalogram. EEG is a well-established approach that enables the noninvasive recording of electrical activity in the human brain. EEG signals refer to voltage fluctuations in the micro-volt range and they are frequently acquired to address clinical as well as research questions. EEG is best suited to record these voltages because it has very high time resolution and captures cognitive processes in the time frame in which cognition occurs, directly measures neural activity, is inexpensive, lightweight, and portable, and monitors cognitive-affective processing in absence of behavioral responses. Data from these experiments are often multivariate, the interactions between the variables are nonlinear, and the landscape of hypothesized or possible interactions between variables is extremely broad.

Information Theory is the best mathematical approach to study these type of time-varying, non-linear brain activities as it includes quantification, storage, and communication of information. The foundation of information theory was laid in 1948 paper by Shannon titled, "A Mathematical Theory of Communication." It is a well suited stochastic statistical tool especially designed to quantify information and to characterize and model neural response [1]-[9]. Key measures in information theory are Entropy and Mutual Information(MI). Entropy quantifies the amount of uncertainty involved in the value of a random variable or the outcome of a random process. MI is a quantity that measures how much one random variables tells us about another. It can be used to analyze EEG which is a non-directional connectivity measure [10]. MI enables the estimation of both linear and non-linear statistical dependencies between time series and can be used to detect functional coupling [11]. Because neural dynamics almost certainly includes many highly nonlinear processes, MI analysis may be helpful in understanding and quantifying the nonlinear transmission of neural information within the brain[10], [12]-[15]. MI has also effectively helped in the research of Abnormal cortical connections in nervous system diseases, such as Alzheimer's, schizophrenia and Parkinson's[16]-[18]. Also, in the cercal sensory system of a cricket, MI transmission from stimulus to response is bound using different approximations of Gaussian mixture models(GMMs)[19]. GMMs are a probabilistic model for representing normally distributed sub-populations within an overall population.

Apart from these works, which mainly focus on information flow in the neural level, this research primarily focuses on high-level information flow between the stimulus and the EEG sensor observations by the electrodes for determining the perceived audio quality of human beings. In this setup, MI and Entropy have been successfully employed to perceive the audio or video quality using EEG[24]-[27].

In [24], MI has been used as a measure to assess audio quality perception by directly measuring the brainwave responses of the human subjects using a high resolution EEG. The subjects are present with audio whose quality varies with time between different possible quality levels, mainly comprising of the worst case scenario, where there are two audio qualities, that is good quality and bad quality. Attempts were made to replicate the results of [24] and it was successful to achieve the above objectives.

More research on this topic, enables to assess the perceived audio quality using EEG with reduced number of electrodes. In order to achieve this, following questions were analyzed:

i. What is the minimum number of electrodes that we need in order to assess the perceived audio quality using EEG?

ii. Is there a good classification between the good and the bad quality of audio perceived after testing with the reduced number of electrodes?

To answer these questions, we analyse the EEG recordings for all our subjects and we divide the huge number of electrodes into specific segments called regions, with each region consisting of 9 electrodes. By using an information theoretic approach and a statistical approach, we take into account some region combinations and determine the minimum number of electrodes. Further, by repeating the experiments with different training and testing sets of the data, thus determine if there is a good classification between the qualities of the audio signal.

The rest of the study is organised as follows. In Section 2, we provide an overview of EEG, the ERP channel and the test sequence used in the analysis. In Section 3, we provide the information theoretic approaches and the algorithms used to determine the entire study. Then the results and the classification plots are shown in Section 4 which is followed by a conclusion in Section 5.

CHAPTER 2

DETAILED DESCRIPTION OF EEG

2.1 Electroencephalogram (EEG)

Electroencephalography is a process of detecting electrical activity in our brain by using electrodes that are attached to the scalp. Figure 2.1 shows the arrangement of 128 EEG electrodes on the scalp of the brain. The brain cells communicate with each other via electric impulses and are active all the time, even when a person is asleep. The EEG records the post-synaptic dendritic currents from cortical pyramidal cells. Not all electrical fields generated by the brain are strong enough to spread all the way through tissue, bone and skull towards the scalp surface. Research studies indicate that it is primarily the synchronized activity of the pyramidal neurons in cortical brain regions which can be measured from the outside (from EEG devices). Their name comes from the fact that they have pyramidal/triangular shaped cell body.

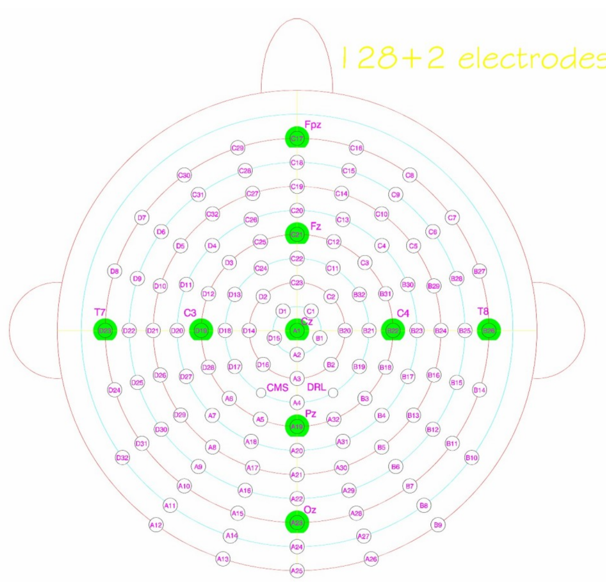

Figure 2.1 The Biosemi EEG cap of 128 electrodes.

Our brain consists of hundreds of thousands of neurons. These neurons have axons, that transmit neural transmitters and dendrites, that receive neuro-transmitters. When the dendrites of a neuron receive neuro-transmitters from the axons of other neurons, it causes an electrical polarity change inside of the neuron. This polarity change is recorded by EEG. The activity of each of the neurons is very difficult to be recorded by the EEG. So, EEG has electrodes with each electrode recording the polarity change in thousands of neurons. An area of a group of neurons working together is called an LFP (Local Field Potential)[28]-[29].

When the EEG signal is time-locked to a stimulus, then the amplitude or potential measured is called **ERP (Event Related Potential)**.The time period before stimulus in an ERP is called baseline. EEG is recorded using the technology of the Differential Amplifier. The differential amplifier takes two electrical inputs and gives output as the difference between the two inputs. Similarly, the EEG also takes the inputs from two electrodes and gives the difference of the potentials from these two electrodes as the output.

The large number of neurons in the human brain have highly complex firing patterns, mixing in a rather complicated fashion. The neural oscillations that are measured with EEG are even visible in raw, unprocessed data. However, the signal is always a mixture of several underlying base frequencies, which are considered to reflect certain cognitive or attentional states. Since these frequencies vary slightly dependent on individual factors, these are classified based on specific frequency ranges, or frequency bands: Delta band (less than 5 Hz), Theta band (4 – 8 Hz), Alpha band (8 – 12 Hz), beta band (more than 12 Hz) and gamma band (more than 25 Hz). The brain wave frequency bands are shown in Figure 2.2[32].

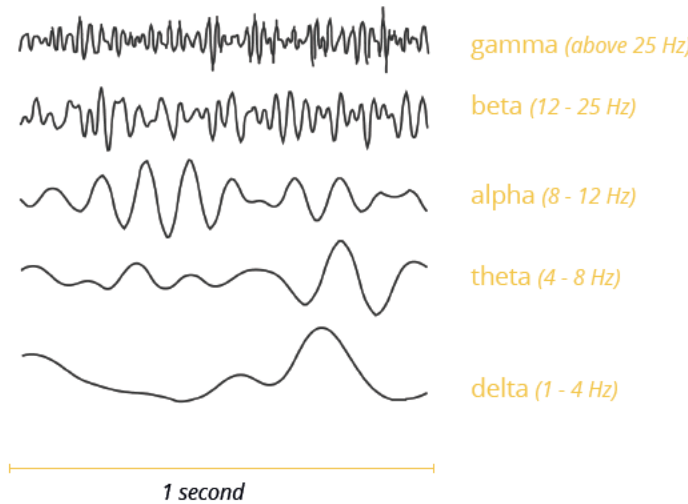

Figure 2.2 Brain Wave Frequency Bands with their respective patterns.

2.2 ERP Channel

An event-related potential (ERP) are the scalp-recorded voltage fluctuations that results from specific sensory, cognitive, or motor event[30]. More formally, it is any stereotyped electrophysiological response to a stimulus. The study of the brain in this way provides a noninvasive means of evaluating brain functioning.

Event-related potentials (ERPs) are very small voltages generated in the brain structures in response to specific events or stimuli. They provide safe and noninvasive approach to study psycho-physiological correlates of mental processes. They are thought to reflect the summed activity of postsynaptic potentials produced when a large number of similarly oriented cortical pyramidal neurons (in the order of thousands or millions) fire in synchrony while processing information. ERPs in humans can be divided into 2 categories. The early waves, or components peaking roughly within the first 100 milliseconds after stimulus, are termed sensory or exogenous as they depend largely on the physical parameters of the stimulus. The ERPs generated in later parts reflect the manner in which the subject evaluates the stimulus and are termed cognitive or endogenous ERPs as they examine information processing[31].

ERP waveforms consist of a series of positive and negative voltage deflections, which are related to a set of underlying components[32]. Though some ERP components are referred to with acronyms (like contingent negative variation is CNV, error-related negativity is ERN), other components are referred to by a letter (N/P) indicating polarity (negative/positive), followed by a number indicating either the latency is in milliseconds or the component's ordinal position in the waveform. For instance, a negative-going peak that is the first substantial peak in the waveform and often occurs about 100 milliseconds after a stimulus is represented as N100 (indicating its latency is 100 ms after the stimulus and that it is negative) or N1 (indicating that it is the first peak and is negative). If it is often followed by a positive peak, then it can be referred to as P200 or P2. The stated latencies for ERP components are often quite variable, particularly so for the later components that are related to the cognitive processing of the stimulus. For example, the P300 component may exhibit a peak anywhere between 250 ms to 700 ms.

For our research, the ERP channel considered here is equivalent to a SIMO(Single Input Multiple Output) channel with unknown characteristics, where the 'single input' refers to the quality of the audio stimulus and the 'multiple output' refers to the observations at the EEG sensor electrodes on the scalp.

2.3 EEG Data and Audio Test Sequence

The data captured from the EEG system on each of the 128 spatial channels or electrodes are sampled at 1kHz. An average re-referencing, baseline removal and high pass filtering was performed to remove any DC noise component resulting from eye blinks, movements, and facial muscle artifacts. It was noticed that for each trial, we have a very large multidimensional data-set. So, in order to analyse the data, the 128 electrodes are grouped into 8 regions of interests[24].

Each and every audio test sequence used here were created from three different base sequences which were sampled at 44.1kHz with a precision of 16 bits per sample. 36 second long test sequences were created with a mixture of both high and low quality of audio. Out of the 36 second long sequence, the first 1 second and the last 5 seconds are fed with no distortions. The remaining 30 second long duration is fed up with either Frequency Truncation distortion or Scalar Quantization distortion[0].

CHAPTER 3

INFORMATION THEORETIC DEFINITIONS AND ALGORITHMS

3.1 Entropy

Entropy is the average rate at which information is produced by a stochastic source of data. Calculating information entropy is a useful tool in machine learning and information theory and is used as the basis for techniques such as feature selection, building decision trees, and, more generally, fitting classification models. Entropy provides a measure of the average amount of information needed to represent an event drawn from a probability distribution for a random variable. The measure of information entropy associated with each possible data value is the negative logarithm of the probability mass function for the value.

$$H(X) = -\sum_{i=1}^{n} P(x_i) log_b P(x_i) \tag{3.1}$$

where H(X) denotes the entropy of random variable X, P(x) is the probability distribution of x and n is the number of samples of X.

Information entropy is typically measured in bits(also called Shannons) or sometimes in natural units (nats). For this analysis, all logarithms to base 2 were taken, thus measuring the entropy and MI in bits.

The dynamic range for Entropy is between 0 and 1. The present experiment reveals the entropy within the dynamic range.

3.2 Mutual Information

Mutual Information quantifies the amount of information obtained about one random variable through observing the other random variable.

Mutual Information is also known as information gain.

$$
\begin{aligned}
I(X;Y) &= H(X) - H(X|Y) \\
&= H(Y) - H(Y|X) \\
&= H(X) + H(Y) - H(X,Y) \\
&= H(X,Y) - H(X|Y) - H(Y|X)
\end{aligned}
$$

Here, X is the ERP channel input and Y is the ERP channel output. Similar to entropy, the dynamic range for the MI value is between 0 and 1.

MI is used to predict the perception of the stimulus. The perception of stimulus is an equiprobable Bernoulli distribution containing two values, either high or low quality. Since, we are only interested in the MI between X and Y, a moderate-to-high value of MI is sufficient enough to indicate that the EEG data output is related to the input stimulus, which means it contains data about the audio quality.

Mutual information has many appealing information theoretic properties. A widely recognized advantage of mutual information is that it allows one to detect non-linear relationships. This can be attractive in particular when dealing with time series data [34]. For categorical variables, mutual information is (asymptotically) equivalent to other widely used statistical association measures such as the likelihood ratio statistic or the Pearson chi-square test. In this case, all of these measures (including MI) are arguably optimal association measures. Interpreting MI as a likelihood ratio test statistic facilitates a straightforward approach for adjusting the association measure for additional covariates [35].

3.3 Gaussian Mixture Models

There has been many research work about Gaussian Mixture Models (GMMs) and Split-Merge Algorithms. Being an extremely powerful probability model, GMM has been widely used in fields of pattern recognition, information processing and data mining. GMMs have also been used for feature extraction from speech data, and used extensively in object tracking of multiple objects, where the number of mixture components and their means predict object locations at each frame in a video sequence. If the number of the Gaussians in a mixture is known, the Expectation-Maximization (EM) algorithm could be used to estimate the parameters in the Gaussian mixture model. However, if the number of Gaussians in the mixture is not known or the computation of the GMMs is complex, then it is very difficult to have an estimation of the parameters in the GMMs.

Here, Taylor Series has been used to calculate the entropy of the Gaussian mixture since the computation of higher order terms is very complex and time taking. Moreover, since we are considering brain activity here, which is non-linear and time varying, Gaussian splitting has become highly important. Also, as we are having two inputs(i.e., good and bad quality audio), namely x_1 and x_2, our probability of the output variable $\underline{y}$ can be written as a sum of two conditional probabilities.

$$p(\underline{y}) = \frac{1}{2}[p(\underline{y}|x_1) + p(\underline{y}|x_2)]$$

By now, for the nonlinear measuring situation, one common view believes that larger co-variance will cause larger error. Therefore, splitting the predicted Gaussian component with large co-variance into Gaussian components with smaller co-variance has become the main trend to alleviate this error.

3.4 Gaussian Splitting

Gaussian Splitting is a process by which a single gaussian component is split into two or more gaussian components in a manner to re-approximate the original component as a mixture itself. A new initialization scheme based on the splitting strategy commonly used in vector quantization has been implemented here. Using this method[36],

i. First the Gaussian component is identified in the mixture with high variance. The mean,variance and weight of Gaussian component are denoted as $\mu_{\mathbf{i}}$, $\sigma_{\mathbf{i}}$ and $\mathbf{w_i}$, respectively.

ii. Then the Gaussian component is split into two by slightly perturbing the mean of the Gaussian along the direction of the standard deviation vector, and re-estimate the model by further EM training[37]. For a four-way split and so on, the standard deviation, mean and the weighting coefficients are reduced in a progressive fashion.

Further:

(a) Attention is given to restrict splitting libraries with equal standard deviations, which means, the standard deviation of the two Gaussian components after split is the same as the original standard deviation.

$$\sigma_1 = \sigma_2 = \sigma_i \text{ (for a two-way split)}$$

where σ_1, σ_2 are the standard deviations of the two Gaussian components

(b) The number of mixture components(L) is restricted to a power of two, which means we can only have a two-way split, four-way split, eight-way split and so on.

$$L = 2, 4, 8, 16, 32, ...$$

(c) For a two-way split, the original weight vector $\mathbf{w_i}$ is equally divided into the weights of the resulting Gaussians, namely w_1 and w_2. For a four way split, the weights of the four consecutive Gaussian components are reduced in a progressive fashion.

$$\text{For a two-way split, } w_1 = w_2 = \frac{w_i}{2} = 0.5$$

For a four way split, two ways of splitting has been found out[24],[37].

Table 3.1 Splitting Library For Weights

Weights	Way1	Way2
w_1	0.35690	0.127380
w_2	0.61042	0.372619
w_3	0.61042	0.372619
w_4	0.35690	0.127380

(d) The means, μ_1 and μ_2 of the Gaussian components after split is slightly changed along the direction of the standard deviation vector.

$$\text{For a two-way split, } \mu_1 = \mu_i - \epsilon$$
$$\mu_2 = \mu_i + \epsilon$$

where ϵ is a positive quantity.

This form of splitting is difficult to apply in practical scenarios, since selection of ϵ is always a compromise. When the two new densities are spaced far apart, ϵ is large and when the two new densities are close to each other, ϵ is small. In our research, ϵ has been set as 0.001. Instead of using ϵ, [38] has proposed $\frac{\sigma_i}{2}$.

3.5 Bootstrap Algorithm

The bootstrap method is a resampling technique used to estimate statistics on a population by sampling a dataset with replacement[39]. Bootstrapping paves a way to calculate standard errors, construct confidence intervals, and perform hypothesis testing for numerous types of sample statistics. Bootstrap methods are alternative approaches to traditional hypothesis testing and are notable for being easier to understand and valid for more conditions.

It can be used to estimate summary statistics such as the mean or standard deviation. It is used in applied machine learning to estimate the skill of machine learning models when making predictions on data not included in the training data. Steps involved are:

i. A dataset of size n is taken.

ii. One single instance is drawn from this dataset and assigned to the jth bootstrap sample. This step is repeated until the bootstrap sample has size n; the size of the original dataset. Each time, samples are drawn from the same original dataset so that certain samples may appear more than once in the bootstrap sample[40].

iii. A model is then fit to each of the b bootstrap samples and the resubstitution accuracy is computed.

iv. The model accuracy is calculated as the average over the b accuracy estimates.

3.5.1 Advantages of Bootstrapping

It is a simple algorithm. It is a straightforward way to derive estimates of standard errors and confidence intervals for complex estimators of complex parameters of the distribution[41].

It allows us to handle outliers without arbitrary cutoffs. Bootstrap is a powerful, computer-based method for statistical inference without relying on too many assumption.

Bootstrap is also an appropriate way to control and check the stability of the results. Although for many problems it is impossible to know the true confidence interval, bootstrap is asymptotically more accurate than the standard intervals obtained using sample variance and assumptions of normality[42].

3.5.2 Disadvantages of Bootstrapping

Bootstrapping is computationally expensive. It cannot be applied to small data sets. Also, the bootstrapping can be time-consuming.

Here t-percentile bootstrap is performed to calculate the estimate of the median of the MI's and therefore a CI estimation is done to find out the upper and lower bound of the MI[24].

CHAPTER 4

ANALYSIS

4.1 Mutual Information Calculation

In our discussion so far, we have noticed that our ERP channel is nonlinear, non-stationary and time varying. So, we begin our analysis by considering the known parameters of our model, which are the input and the output random variables. We follow the same approach as discussed in [24] to calculate the mutual information for a single region. For this first we calculate the conditional entropy of the output random variable $\underline{\mathrm{Y}}$ given the input random variable X.

$$h(\underline{Y}|X) = 0.5(log(2\pi|\underline{C_1}|) + 1) + 0.5(log(2\pi|\underline{C_2}|) + 1)$$

Here $\underline{C_1}$ and $\underline{C_2}$ are the $n * n$ co-variance matrices of the two input distributions. Then we calculate the entropy of the output random variable, by using the law of total probability and taking a mixture of the conditional probabilities.

$$h(\underline{Y}) = -0.5 \int_{\mathbb{R}^n} [p(\underline{y}|x_1) + p(\underline{y}|x_2)] \cdot log(0.5[p(\underline{y}|x_1) + p(\underline{y}|x_2)])$$

where x_1 and x_2 are the two inputs, namely good quality audio and bad quality audio and $p(\underline{y}|x_1)$ and $p(\underline{y}|x_2)$ are the conditional probabilities. As both $p(\underline{y}|x_1)$ and $p(\underline{y}|x_2)$ are multivariate Gaussian distributions, so the output vector $\underline{y}$ is a multivariate Gaussian mixture or a multivariate Gaussian mixture model. Finally we calculate the MI between the EEG output data and the input audio stimulus.

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

In order to approximate the data, all the other steps were carried out as performed in [24] including Gaussian splitting and Taylor series expansion of the Gaussian Mixture Models(GMMs).

To answer the first question, *What is the minimum number of electrodes that we need in order to assess the perceived audio quality using EEG?*, first we compute and analyze the MI results for 4 regions instead of the total 8 regions. This is accomplished using the same equations mentioned above, but instead of taking data samples from 8 different regions, we consider data samples from only 4 regions. All the combinations of 4 different regions were analyzed and the MI results were calculated. While considering four region combinations, we keep in mind that we are considering a total of 36 electrodes. A similar approach was carried out to analyze the MI results for even lesser regions or electrodes. Further, we consider 3 region or 27 electrode combinations and 2 region or 18 electrode combinations and thus calculate the mutual information for each of these combinations. In all the scenarios, a high MI for a combination indicates that the combination has high ability to assess perceived audio quality. Here, we are not considering MI results from single region or 9 electrode observations, as they are too few to determine or assess the quality of audio.

4.2 Variance Detector

To further validate our standing on our MI results and support our analysis, we devise a Variance Detector for each of these region combinations. In order to achieve this, first we partition the data samples of a particular region into nine equal sized folds or blocks, as it is very difficult to analyze thousands of samples region-wise, trial-wise and subject-wise. Then we calculate the variances of these blocks. Here, we are considering a single block to be of 10240 samples or duration wise a 10 second

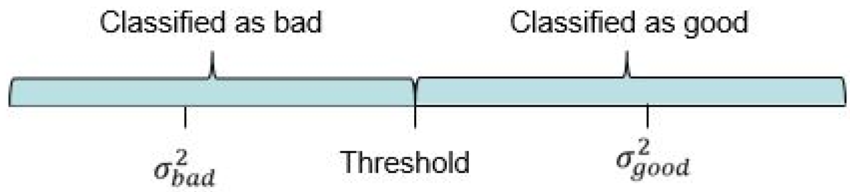

Figure 4.1 Figure showing the two training variances with the threshold and the perceived classifications.

block. Each analysis region-wise, is done by considering the first 10 seconds as the training set and the rest 80 seconds as the test set. While considering a specific region combination, the training set is taken as the first ten seconds from each region present in the combination, while the other remaining samples from each region is taken as the test set. Since, we have two types of input here, namely good quality and bad quality, so we also have two different training variances and two different set of test variances.

Going further, a group of thresholds are assumed between the two training variances. Figure 4.1 shows the Variance Detector approach where the threshold is in between the training variances and classification is drawn to the right and left of the threshold. For a given discrimination threshold value T, the classification rule is such that a subject is allocated to the positive class if and only if its score exceeds the threshold. If the score is less than the threshold, the subject is allocated to the negative class. The true positive probability, also called True Positive Rate(TPR) is the probability that a subject from the positive class is correctly classified as belonging to the positive class:

$$TPR(T) = \frac{\sum_{i=1}^{N_p}(S_i^{(p)} > T)}{N_p}$$

where N_p is the total number of values belonging to the positive class and $S_i^{(p)}$ is a random variable of the positive class. Similarly, the false positive probability, also called False Positive Rate(FPR), is the probability that a subject from the negative class is unclassified or incorrectly classified as belonging to the positive class:

$$FPR(T) = \frac{\sum_{i=1}^{N_n}(S_i^{(n)} < T)}{N_n}$$

where N_n is the total number of values belonging to the negative class and $S_i^{(n)}$ is a random variable of the negative class. So, we check these thresholds on our test data sets. If the variance for a certain block of test data is greater than the threshold, then we assume it belongs to the positive class(good quality) and if it is less than the threshold, we say it belongs to the negative class(bad quality).

To answer the second question, *Is there a good classification between the good and the bad quality of audio perceived after testing with the reduced number of electrodes?*, we follow the same process as discussed in the variance detector approach. A proper Receiver Operating Characteristic(ROC) Curve is plotted by taking the true and false error probabilities. A Curve with a bend towards the upper left corner of the ROC plane, indicates that there is a good classification between the good and the bad quality audio for that region combination.

4.3 Z Score Calculation

To analyze the results achieved by the MI calculations so far, we further calculate the z-scores for the combinations. A higher z-score would indicate a proper classification of good and bad quality audio. Such a z-score is expressed as the inverse error function of the probabilities. It is given by

$$z = \sqrt{2}erf^{-1}((2TPR) - 1) - \sqrt{2}erf^{-1}((2FPR) - 1)$$

where TPR and FPR are the true probability rate and the false probability rate respectively. In order to calculate the TPR and FPR values, we follow the approach of the Variance Detector, where we take the first eighty percent of the total data as the training set, and the remaining twenty percent of the data as the test set. We calculate the variances for both the good and bad quality training data sets and take a threshold which is the midpoint of the training variances. In a fashion similar to that used in Variance detector approach, the TPR and FPR values are computed for each region combinations and using these values, we calculate our z-scores.

4.4 Correlation between Mutual Information and Z Score

To further validate our results, we analyze if there is a good correlation between the Mutual Information and the Z Score results. In order to proceed with this, we try to compute the R^2 test, to check what percent of MI can be verified by the Z score. The R^2 test is basically a statistical measure that represents the proportion of the variance of MI that is explained by the Z Score. In other words, it tries to explain, to what extent the variance of one variable explains the variance of the other variable. Here R denotes the Correlation Coefficient, which can be calculated as,

$$R = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}}$$

where x is the Z Score and y is the Mutual Information. There is a limitation for the R^2 test. A high or a low R^2 is not necessarily good or bad, since it does not convey the reliability of the model.

To recheck our correlation between the MI results and the Z scores, we analyze the correlation by drawing a Linear Regression Model. The Linear Regression Model establishes a relationship between the dependent variable(MI) and the independent variable(Z score).

CHAPTER 5

RESULTS

First the Mutual Information was evaluated for each 3 region combinations where data samples from total 27 electrodes were taken. The mutual information was calculated using the same approach as described in [24], by fourth order Taylor series approximation. While considering a particular region combination, the neural activity occurring between the other regions are considered negligible. It is interesting to note that regions 1,2,6 and regions 2,5,6 have a high mutual information as compared to other region combinations. Simultaneously, the z scores were calculated for these region combinations. In order to showcase all the MI and z-score results, everything was plotted in terms of confusion matrices as done in [43]. Figures 5.1 to 5.8 shows the confusion matrices plotted for both Mutual Information and Z scores for each region combination. Each figure is related to a specific region followed by a 7x7 matrix of other regions. The diagonal elements in the matrices, correspond to 2 region scores which are discussed further.

In order to reduce the number of electrodes further, the variations of Mutual Information and z score results was computed for 2 region combinations. It was interesting to find out that regions 1,2 and 2,6 and 1,6 show a high MI and z score. This is in strong agreement with the results found out from 3 region combinations, where combinations 2,5,6 and 1,2,6 were showing high MI and z-score results. Figure 5.9 shows the confusion matrices plotted for both Mutual Information and Z scores for each region combination. The results were cross validated by taking 80 percent of data as training set to estimate the variances, then the z-score was calculated on the

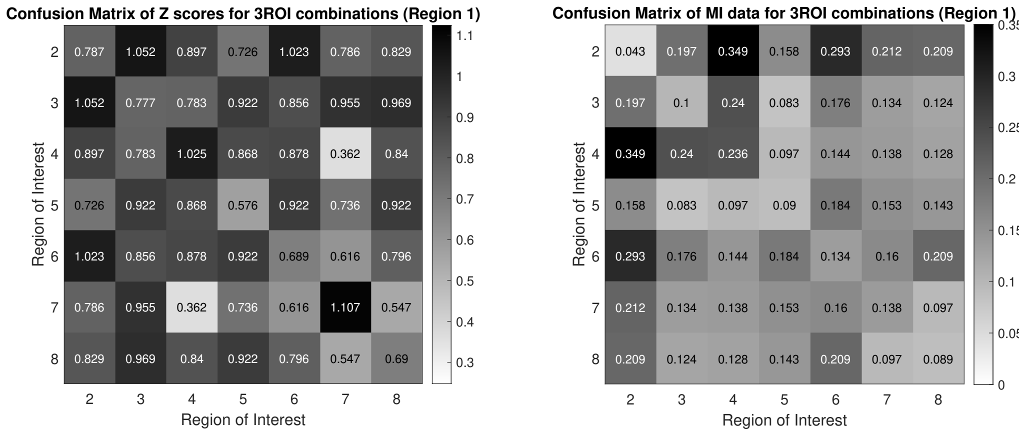

Figure 5.1 Confusion Matrices for Mutual Information and Z scores for 3 region combinations (Region 1).

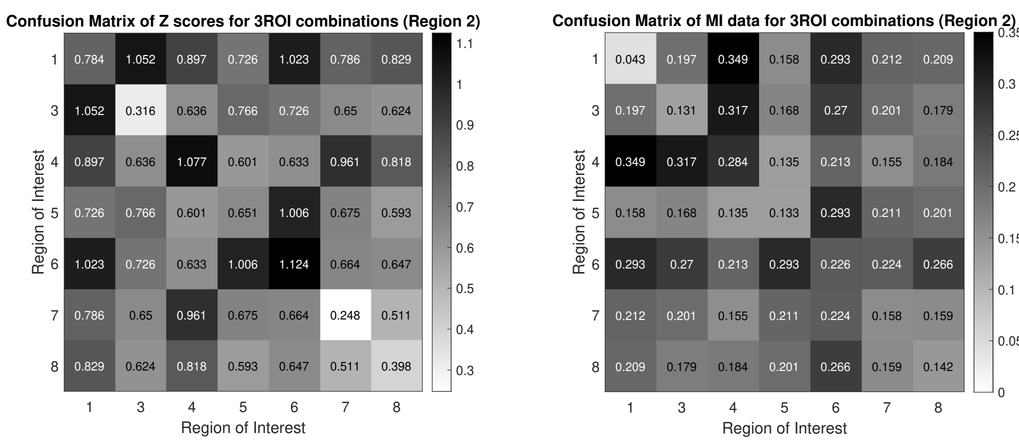

Figure 5.2 Confusion Matrices for Mutual Information and Z scores for 3 region combinations (Region 2).

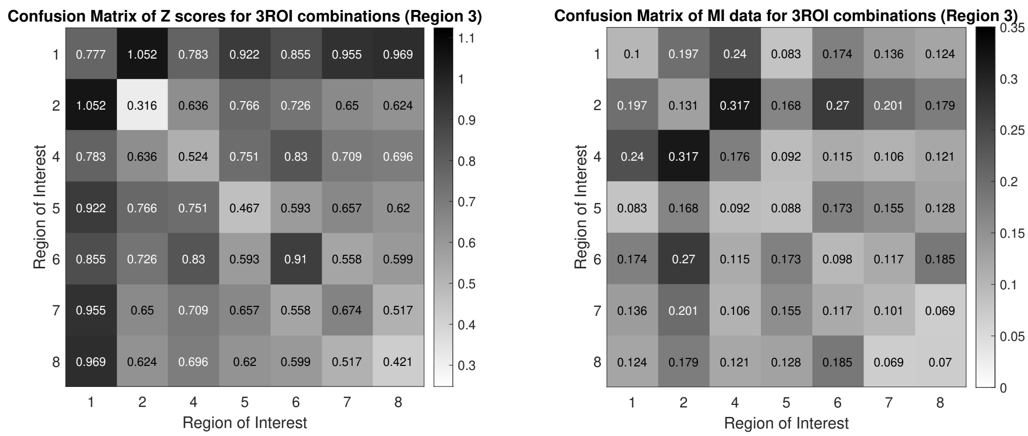

Figure 5.3 Confusion Matrices for Mutual Information and Z scores for 3 region combinations (Region 3).

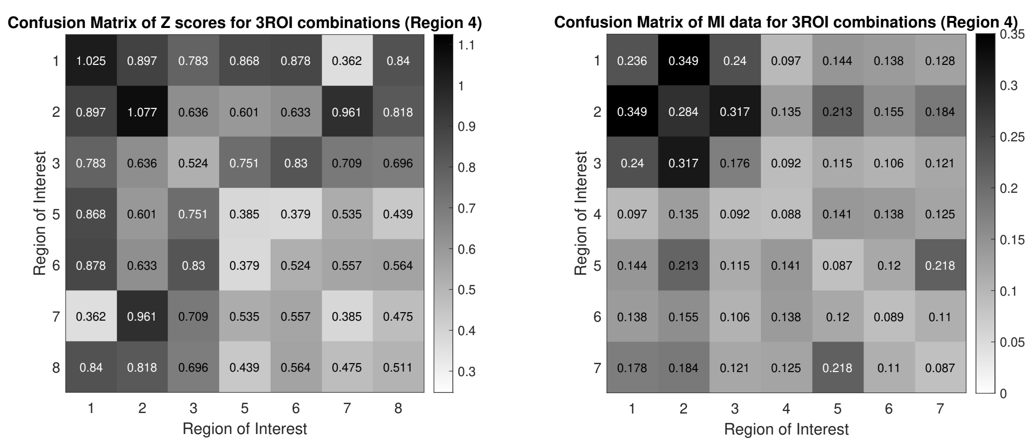

Figure 5.4 Confusion Matrices for Mutual Information and Z scores for 3 region combinations (Region 4).

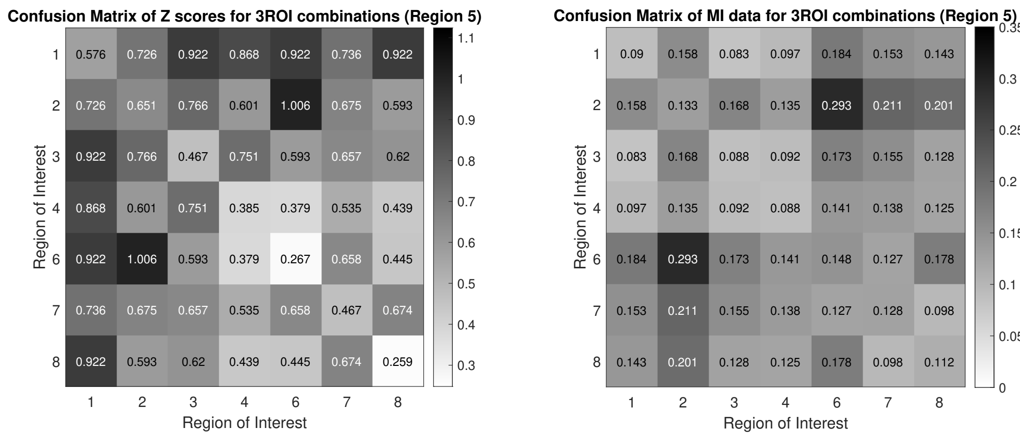

Figure 5.5 Confusion Matrices for Mutual Information and Z scores for 3 region combinations (Region 5).

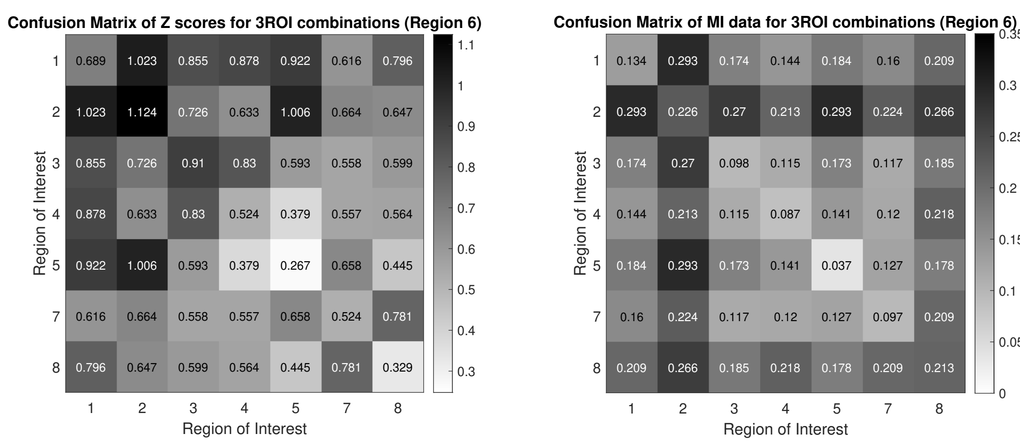

Figure 5.6 Confusion Matrices for Mutual Information and Z scores for 3 region combinations (Region 6).

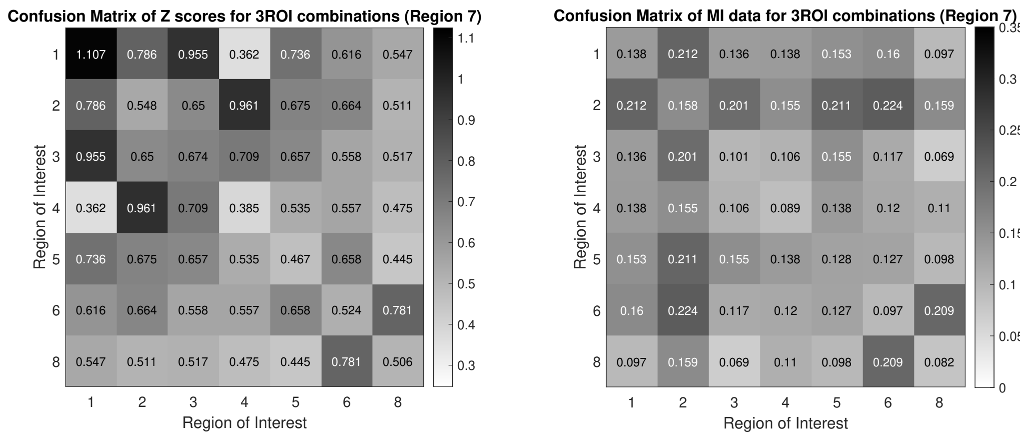

Figure 5.7 Confusion Matrices for Mutual Information and Z scores for 3 region combinations (Region 7).

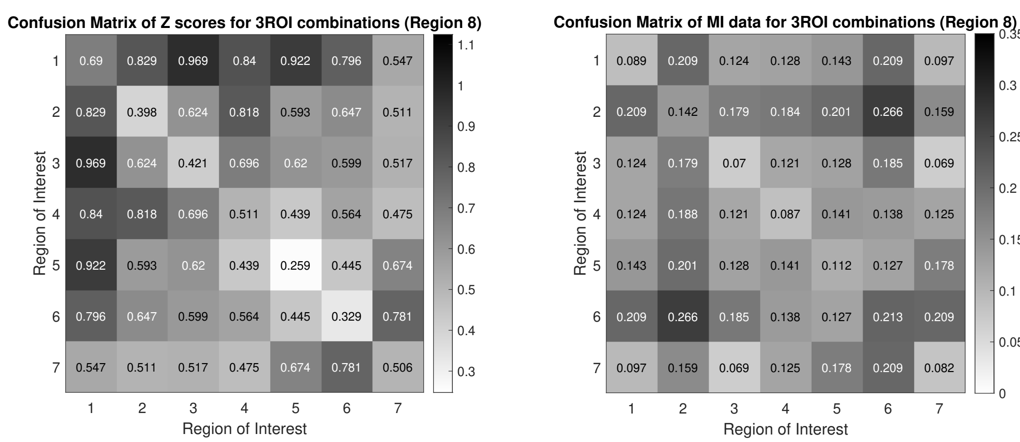

Figure 5.8 Confusion Matrices for Mutual Information and Z scores for 3 region combinations (Region 8).

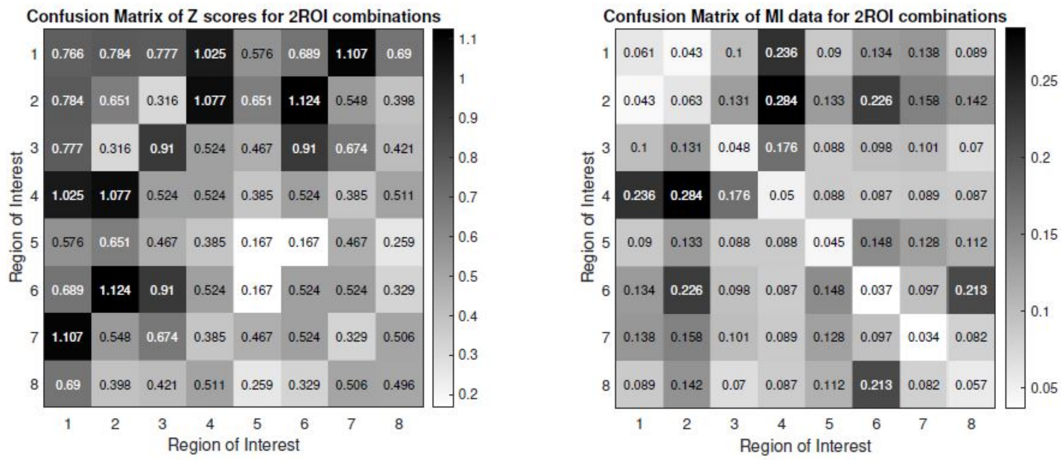

Figure 5.9 Confusion Matrices for Mutual Information and Z scores for 2 region combinations.

remaining 20 percent and this process was repeated 100 times.

After finding out which regions give us a better MI and better z-score, we try to establish a correlation between the two parameters. For this, we plot a scatter plot of both MI and z-score results and calculate the square of the correlation coefficient. We further do a linear regression fit and try to analyze the slope. Figures 5.10 and 5.11 show the scatter plots for 3 region and 2 region combinations. From the values of the square of correlation coefficient(R), it can be seen that there is very less correlation between the two parameters.

Further, we classify the good and bad quality audio using our Variance Detector approach by plotting the ROC curves. The Receiver Operating Characteristic(ROC) curve is a graphical plot by which the degree of discrimination between good and bad quality could be estimated. ROC applies different threshold values to plot the true positive rate (TPR) against the false positive rate (FPR). This was done by using the Contingency Table for different values of threshold. Figures 5.12 and 5.13 show

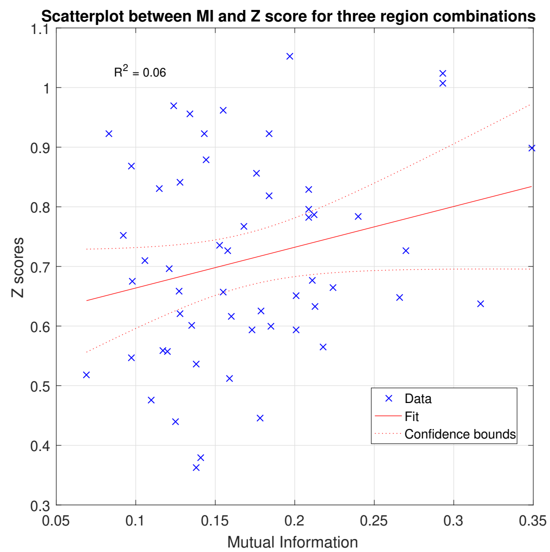

Figure 5.10 Scatter plot between Mutual Information and Z scores for 3 region combinations.

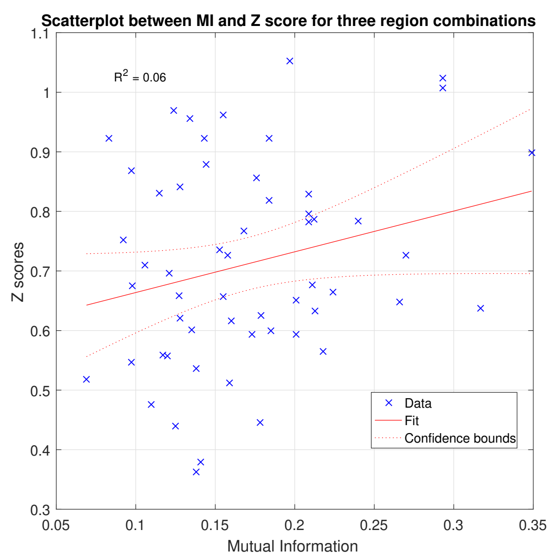

Figure 5.11 Scatter plot between Mutual Information and Z scores for 2 region combinations.

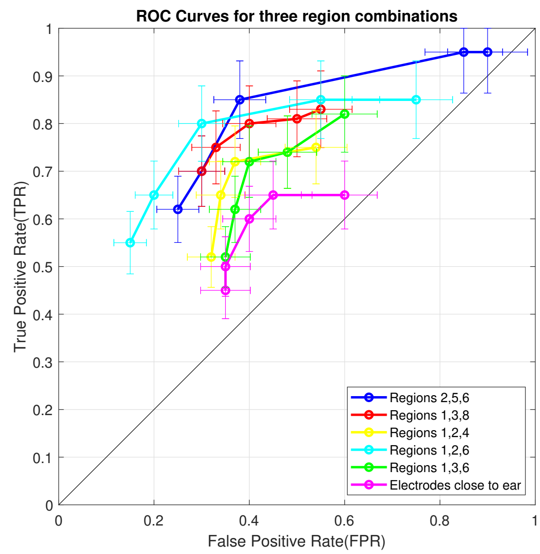

Figure 5.12 Receiver Operating Characteristic(ROC) curves for best 3 region combinations.

the ROC curves for the best five combinations of three and two region combinations. It can be seen from both figures that regions 2 and 6 together shows us a better performance.

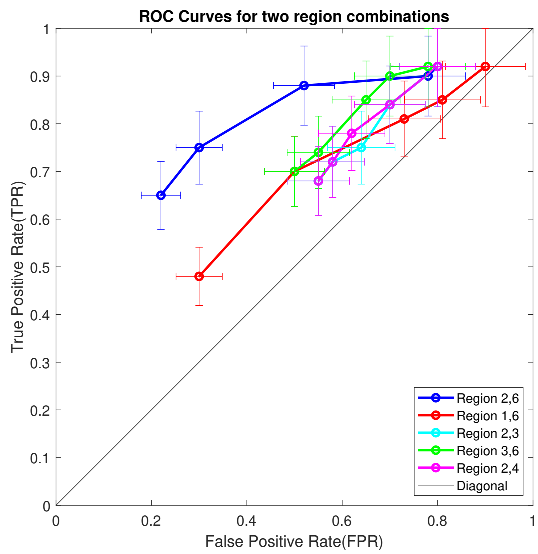

Figure 5.13 Receiver Operating Characteristic(ROC) curves for best 2 region combinations.

CHAPTER 6

CONCLUSION

It has been observed from the above analysis that, a minimum of 2 regions or 18 electrodes may be used to properly assess the perceived audio quality, which has been verified by the Mutual Information, Z score computation and the ROC curves.

Both the graphical and analytical representations of the Mutual Information reveals the information in the data where the low quality is different from the high quality. This has been classified by the variance detector. The ROC curves have been plotted, which shows that there is a good classification between the good and bad quality audio. To analyze this further, we computed the z scores. A squared correlation coefficient test shows that only 17 percent of the variance of the MI can be verified by the z-score. In order to get a significant Z Score and a high R^2 value, Variance Detector approach may be replaced by any other approach in the future.

REFERENCES

[1] Nicholas M. Timme and Christopher Lapish (2018), *A Tutorial for Information Theory in Neuroscience* eNeuro 29 June 2018, 5 (3) ENEURO.0052-18.2018; DOI: https://doi.org/10.1523/ENEURO.0052-18.2018

[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] Huan Gao, Qiguang Miao, Jiachen Yang, and Zhenxin Ma *Image Quality Assessment Using Image Description in Information Theory*

[9] Leslie W. Dalton *Information Theory, Kelly Betting, Risk, Reward, Commission, and Omission: An Example Problem in Breast Cancer*

[10] Changcheng Sun, Fei Yang, Chunfang Wang, Zhonghan Wang, Ying Zhang, Dong Ming and Jingang Du *Mutual Information-Based Brain Network Analysis in Post-stroke Patients With Different Levels of Depression* Front. Hum. Neurosci., 17 July 2018 — https://doi.org/10.3389/fnhum.2018.00285

[11] Wang, Z. J., Lee, P. W.-H., and McKeown, M. J. (2009). *A novel segmentation, mutual information network framework for EEG analysis of motor tasks* Biomed. Eng. Online 8:9. doi: 10.1186/1475-925X-8-9

[12] Jeong, J., Gore, J. C., and Peterson, B. S. (2001). *Mutual information analysis of the eeg in patients with alzheimer's disease* Clin. Neurophysiol. 112, 827–835. doi: 10.1016/S1388-2457(01)00513-2

[13] R. B. Stein, *The information capacity of nerve cells using a frequency code* Biophys. J., vol. 7, no. 6, pp. 797–826, 1967.

[14] 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.

[15] W. S. McCulloch, *An upper bound on the informational capacity of a synapse* in Proc. ACM Nat. Meeting, 1952, pp. 113–117.

[16] Coronel, C., Garn, H., Waser, M., Deistler, M., Benke, T., Dal-Bianco, P., et al. (2017). *Quantitative EEG markers of entropy and auto mutual information in relation to MMSE scores of probable Alzheimer's disease patients* Entropy 19:130. doi: 10.3390/e19030130

[17] Yin, Z., Li, J., Zhang, Y., Ren, A., Von Meneen, K. M., and Huang, L. (2017). *Functional brain network analysis of schizophrenic patients with positive and negative syndrome based on mutual information of EEG time series* Biomed. Signal Process. Control 31, 331–338. doi: 10.1016/j.bspc.2016.08.013

[18] 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

[19] 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.

[20] L. Wu, P. Neskovic, E. Reyes, E. Festa, and H. William, *Classifying nback EEG data using entropy and mutual information features* in Proc. Eur. Symp. Artif. Neural Netw., Bruges, Belgium, 2007, pp. 61–66.

[21] B. Hjorth *An on-line transformation of eeg scalp potentials into orthogonal source derivations* Electroencephalog. Clin. Neurophysiol., 39(5):526–530, 1975

[22] Joyeeta Mitra,John R.Glover, Periklis Y.Ktonas, Eli M.Miizrahi (2009), *A Multistage System for the Automated Detection of Epileptic Seizures in Neonatal Electroencephalography*

[23] 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.

[24] K. Mehta, J. Kliewer *An information theoretic approach towards assing perceptual audio quality using EEG* IEEE Transactions on Molecular, Biological, and Multiscale Communications, pages 176-187, June 2015.

[25] 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.

[26] 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.

[27] 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.

[28] Britton JW, Frey LC, Hopp J L, St. Louis EK, Frey LC, *Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants [Internet]* Chicago: American Epilepsy Society; 2016.

[29] Chris Berka, Daniel J. Levendowski, Milenko M. Cvetinovic, Miroslav M. Petrovic, Gene Davis, Michelle N. Lumicao, Vladimir T. Zivkovic, Miodrag V. Popovic and Richard Olmstead *Real-Time Analysis of EEG Indexes of Alertness, Cognition, and Memory Acquired With a Wireless EEG Headset* Pages 151-170 — https://doi.org/10.1207/s15327590ijhc17023

[30] Luck, Steven J. (2005). *An Introduction to the Event-Related Potential Technique* The MIT Press. ISBN 978-0-262-12277-1.

[31] Shravani Sur and V. K. Sinha *Event-related potential: An overview* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3016705/

[32] Bryn Farnsworth (2019) *EEG (Electroencephalography): The Complete Pocket Guide* https://imotions.com/blog/eeg/

[33] Luck, S.J.; Kappenman, E.S., eds. (2012). *The Oxford Handbook of Event-Related Potential Components* Oxford University Press. p. 664. ISBN 9780195374148

[34] Wiggins C, Nemenman I, *Process pathway inference via time series analysis. Experimental Mechanics* 43(3):361–370. 10.1007/BF02410536

[35] Lin Song, Peter Langfelder Steve Horvath, *Comparison of co-expression measures: mutual information, correlation, and model based indices* BMC Bioinformatics volume 13, Article number: 328 (2012)

[36] A. Sankar, *Experiments with a Gaussian Merging-Splitting Algorithm for HMM Training for Speech Recognition,* SRI International, 1998

[37] Uwe D. Hanebeck, Kai Briechle, Andreas Rauh, *Progressive Bayes: a new framework for nonlinear state estimation,* Proceedings of SPIE Vol.5099, Pg 256-267

[38] GuoQing Yin, Dietmar Bruckner *Split-merge algorithm and Gaussian mixture models for AAL* 2010 IEEE International Symposium on Industrial Electronics

[39] Sebastian Raschka, *Model evaluation, model selection and algorithm selection in machine learning,* https://sebastianraschka.com/blog/2016/model-evaluation-selection-part1.html

[40] D. N. Politis, *Computer-intensive methods in statistical analysis* IEEE Signal Process. Mag., vol. 15, no. 1, pp. 39–55, Jan. 1998.

[41] *A practical introduction to the bootstrap: a versatile method to make inferences by using data-driven simulations* https://garstats.wordpress.com/

[42] DiCiccio TJ, Efron B (1996), *Bootstrap confidence intervals (with Discussion)* Statistical Science 11: 189–228

[43] Hamad Ahmed, Ronnie B Wilbur, Hari M Bharadwaj, Jeffrey Mark *Object classification from randomized EEG trials* Purdue University, West Lafayette IN 47907-2035, USA.