

IMPERIAL COLLEGE LONDON

DEPARTMENT OF MATHEMATICS

APPLIED MATHEMATICS MSc PROJECT

Information-theoretic analysis of EEG from a musical performance

Author:
Joshua SOUTHERN

Supervisor:
Prof. Henrik JENSEN

September 14, 2018

College Identifier: 01430811

**Imperial College
London**

Declaration of own work

The work contained in this thesis is my own work unless otherwise stated.

Signature: Joshua Southern

Abstract

This project is motivated by EEG measurements taken from musicians and the audience during a performance of classical music. The measurements are taken during both improvisation and non-improvisation states. The aim of this study is to use the recently developed causality measure, partial mutual information with mixed embedding (PMIME), to construct a cross-brain network between the people in the concert and then to identify differences in community structure between the two different states of performance. This was supplemented by a complexity analysis on the signals at individual electrodes, so that both local and macro differences in brain function between the modes of playing were explored. For the complexity analysis, permutation Lempel-Ziv was used as a measure to complement previous work done using ordinary Lempel-Ziv complexity. In both our complexity analysis and our causality-network analysis, we have found neuronal differences between the musicians improvising and when they are playing a mechanical rendition of music. The differences can be difficult to interpret, but there is a quantifiable difference in the information flow in the whole system of people, information flow between musicians and the local complexity of signals at electrodes.

Acknowledgements

I would like to express my sincere gratitude to:

- My supervisor, Henrik Jeldtoft Jensen, for his guidance and support, introducing me to the research area and giving me freedom to do my own research.
- My second supervisor, Hardik Rajpal, for giving me ideas for avenues of research and always providing guidance whenever it was required as well as technical advice.

Contents

1	Introduction	6
1.1	The brain as a complex system	6
1.2	Music and the brain	7
1.3	Structure of the thesis	8
2	Understanding complex systems	10
2.1	From time-series to interactions	10
2.2	From interactions to networks	11
3	Information Theory, Causality and PMIME	12
3.1	Information	12
3.2	Entropy	13
3.3	Mutual information rate	14
3.4	Transfer entropy	15
3.5	Partial Mutual Information with Mixed Embedding	16
4	Networks and Community detection	18
4.1	Networks and representations	18
4.2	Directed networks	19
4.3	Weighted networks	19
4.4	Overview of community detection	19
4.5	The map equation	21
5	Kolmogorov complexity and its approximation	22
5.1	Kolmogorov complexity	22
5.2	Lempel-Ziv complexity	23
5.3	Permutation Lempel-Ziv complexity	25
6	EEG analysis on music improvisation	26
6.1	Experimental setup	26
6.2	Methods	27
6.2.1	Causality analysis	27
6.2.2	Cross-brain network analysis	28
6.2.3	Complexity analysis	29
6.3	Results	32
6.3.1	Causality and network analysis	32
6.3.2	Complexity analysis	35
6.4	Discussion	36
6.5	Conclusion	38

List of Figures

1	The Internet Network as found by Bell Labs during the Internet mapping project [8]. Colors represent a community.	20
2	The 6 different possible motifs when the number of data points, $m = 3$	25
3	Histogram of the weights in all of the causal matrices for each mode. The number of bins is set at 100.	28
4	A directed network indicating information flow in the first 4 seconds of the strict mode. The color of the node represents the community to which it belongs . . .	29
5	The normalized autocorrelation function for the electrical activity at the Flutist's F3 electrode	30
6	The effect size for different numbers of motifs, $m!$. Stars indicate the level of significance of the effect size from a permutation test.	31
7	The number of communities as a function of time for the the two cross-brain networks. The differences in mean between the two modes is also shown. . . .	32
8	Plots of the size of the largest community over time for the the two cross-brain networks. A plot of the differences in mean between the two modes is also shown.	33
9	Plots of the entropy of the community structure over time for the the two cross-brain networks. A plot of the differences in mean between the two modes is also shown.	34
10	The average complexity over all electrodes as a function of time in both the strict and the let-go modes for the pianist	35
11	The effect size for the different people between the means of the Let-go mode – means in the Strict mode. Stars indicate the level of significance of the effect size from a permutation test.	35

1 Introduction

1.1 The brain as a complex system

A complex system is most often defined as a system containing a large number of interacting components, which produce emergent structures whose evolution is very sensitive to initial conditions [26]. There is a consensus that the brain is one such complex system [37] - the neurons are the individual components and they interact giving rise to emergent properties such as consciousness [5]. It has been found that some neurons have increased interaction with particular others depending on the performed function [9][37] and so it is believed that the brain can be decomposed into functional parts [13]. Instead of analyzing the interaction of individual neurons, we can gain an understanding of the brain by understanding the functional parts and how these different components cooperate to perform cognitive tasks. This is very difficult given that different parts of the brain are activated at different times whilst it also appears as if information is reaching all neurons at all times. Fortunately, advances in technology have given us the opportunity to analyze brain activity. Neuroimaging techniques such as functional magnetic resonance imaging (fMRI); single-photon emission computed tomography (SPECT) and positron emission tomography (PET) are often used to visualize brain activity. However, although these techniques often produce useful and clear images as well as having a good spatial resolution, they suffer from a slow temporal resolution [6]. This is not so useful if we are wanting to understand cooperation in the brain at neuronal timescales. One method that can be used to trace the information flow in the brain at such a high temporal resolution is an Electroencephalograph (EEG). An EEG was first demonstrated by Hans Berger in 1924 to measure human brain waves [40]. Electrodes are placed at positions around the scalp and the electrical activity is measured at each electrode and then grounded to a reference electrode. As well as being non-invasive and relatively cheap compared to other neuroimaging techniques, the temporal resolution of the EEG is of the order milliseconds, similar to what we believe to be the neuronal timescale [18]. If we can spatially isolate functional parts of the brain with an electrode, then at a neuronal timescale we can hopefully understand functional processes from the electrical activity at that electrode. From the EEG we can also map a flow of information between different functional parts by quantifying the interactions between the different time-

series of electrical activity. Describing this flow of information between functional parts of the brain as well as understanding the local electrical activity could lead to an understanding of both the brain and other complex systems, and how emergent properties arise through small-scale cooperation. Previously, work has been done modeling intra-brain information flow with the study on relating intelligence to the magnitude of causal drivers in an EEG [38], and with an examination of information flow in an EEG to distinguish human awake, meditation and drowsiness states [11]. Rather than just looking at one EEG and understanding how information flows through the brain, this thesis explores the use of a causality analysis to describe information flow between electrodes in multiple EEGs of different people during a musical performance. Hopefully, some insight will be gained on the intra and cross-brain interactions between musicians and audience members. Additionally, a complexity analysis to quantify the activity at local electrodes will be done. The extent to which these analyses can distinguish between different neural activities is looked at by trying to differentiate EEG data recorded when the musicians play a mechanical rendition of music, a so-called strict mode, and a let-go mode, where the musicians are improvising. This will perhaps give some insight into the brain's complex behavior and how it operates in different modes, as well as how people interact during a concert.

1.2 Music and the brain

Previously, there have been a large number of studies looking at the complex behavior of the brain during musical performance due to music's natural relevance to cognition, and the fact that it provides an intimate relationship between production, perception, experience and emotion [7]. An fMRI study was done on performance in pianists [25], a study on beat perception has been undertaken [22], and a study on musical memory has been done, [14] along with many others. Recently, some research has also looked into the effect of improvisation, one of the most complex forms of creative behavior. An improvising musician faces the unique challenge of managing several processes all at once; generating and evaluating melodic and rhythmic sequences; executing fine-motor movements and co-ordinating performance with other musicians in the ensemble - all with the overall goal of creating aesthetically appealing music. Understanding improvisation is not only relevant to the psychol-

ogy of music, but also to the psychology of creativity, and how acquired expertise shapes brain structure and function [3]. A majority of studies looking at the effects of musical performance and improvisation on the brain have focused on finding brain regions which are most activated. For instance, in an fMRI study of melodic and rhythmic improvisation [30], results show that the dorsal premotor cortex is mainly responsible for melodic improvisation, while the pre-supplementary motor area is related to rhythmic improvisation. In a PHD thesis by Xiaogeng Wan, information flow is explored in EEG data taken from musicians and audience members during an improvised performance and a mechanical rendition of music [43]. Rather than look at sites with a high activation, a network of interactions was constructed between electrodes in the EEG and sites that had a high degree centrality were noted, i.e. sites that were important in the flow of information. Additionally, cross-brain networks were created to understand information flow between musicians and also with the audience. Not only does this give insight into a musicians creative cognition but also how musicians co-ordinate during improvisation and the audiences shared experience with the performers. Alike with this thesis, I shall explore cross-brain networks. However, rather than look at meaningful brain regions for information flow or highly activated areas, I will look at community structure between brain regions in both musicians and the audience shedding light on the ‘connectedness’ of the group.

1.3 Structure of the thesis

In this thesis, I shall use the information based causality measure called partial mutual information with mixed embedding (PMIME) in order to construct a cross-brain directed network of information flow from EEG data. This EEG data is taken from musicians and audience members during a musical performance where there is both a strict and a let-go mode. Community detection will then be employed to try and quantify the difference in network structure between the two modes. Additionally, I will use a complexity measure to try and quantify the local process at an electrode and see if this is able to differentiate between the two modes. As well as supplement the growing research area exploring musical performance and the brain, the general analysis used in this thesis can be extended to any complex system containing multivariate time-series. Chapter 2 will give a general overview of this process - Interactions between mul-

tivariate time-series are quantified and then this information is used to construct a network to represent the complex system.

Chapters 3 and 4 provide the mathematical background to the causality and network analysis I use on the EEG measurements. This requires some basic information and network theory.

Chapter 5 gives the background necessary for the permutation Lempel-Ziv measure which I use in the complexity analysis.

Chapter 6 then outlines the practical aspects of my analysis on the EEG data as well as the obtained results. A discussion of these results as well as any potential future work is also given.

2 Understanding complex systems

2.1 From time-series to interactions

Today, large datasets of multivariate time series exist in fields such as physiology, genetics, meteorology and finance; representing complex systems such as the climate, human body, the genome and the global economy. We are often interested in interactions between these time-series. We may want to understand the coupling mechanism between two subprocesses, or the interactions between multiple subprocesses in order to identify how a complex interaction mechanism is mediated. Connections between time-series in these systems are directed, as there is a transfer of information. For instance, the interaction of two stock prices in the FTSE 100 at a particular time would imply that one price has an effect on the other. The most basic approach, and the most common, to quantify interactions is to estimate all Pearson correlations at lag zero between each pair of time-series. This has been done by Tsonis and Roebber in studying the architecture of the climate [39], Onnela et al for quantifying interactions between financial companies [27] and many others wanting to understand interactions between a pair of agents in a complex system. However, this measure signifies that the interactions are instantaneous and thus they cannot be interpreted in a directional way as information transfer implies. Therefore, a measure involving some time-asymmetry must be used. This has previously been done with the use of lagged correlation [34]. The cross correlation lag function is used to assess the time delay and to quantify the strength of a link mediated by a certain mechanism. Although this measure factors in a time-asymmetry, the measure itself is still symmetric. A positive lag correlation implies that time-series X has an influence on time-series Y in the future but also vice versa and therefore it can be hard to interpret physically. Additionally, correlation doesn't imply causation [29]. The industrial revolution caused both a rise in the number of births and the number of chimneys. If we were to just look at time-series of the number of chimneys and births, we would falsely assume that increasing the number of chimneys causes an increase in the number of births as they are positively correlated. Therefore, the measure used to interpret interactions between the time-series should be asymmetric, and also only quantify unique information one variable has about another. Granger's formulation of causality encompasses this idea. He states that if we want to test whether variable X causes variable Y , the

first step would be to predict the current value of Y with all previous information, other than variable X . Then in a second step predict variable Y with all previous information including X . If the second prediction is judged to be better than the first one, then one can conclude that X causes Y . A measure to test for this Granger causality would be perfect in our attempt at quantifying interactions between time-series. However, finding such a measure is difficult. It is unrealistic to be able to account for all previous information, as the number of time-series in our system is finite. Additionally, previous measures such as the Granger causality test, or its extension to non-linear dependencies, require models to be fit to the time-series, and thus make assumptions about the underlying process. In order to analyze interactions in the multivariate EEG data, I shall use tools from information theory which I believe are perfectly suited for measuring this form of causality.

2.2 From interactions to networks

Each time-series within the complex system represents a specific quantity and can be viewed as a node. The quantified interaction between each pair of time-series can then be interpreted as an edge between the nodes. This gives us a network representing dependencies in the system. Many complex systems have been modeled as a network; such as the World Wide Web, Social friendships and Power Grids, as a network representation offers a powerful way of seeing the manner in which complex systems are interconnected [31]. Network models of complex systems have shed light on a variety of complex empirical phenomena, including the frequencies of protein-protein interactions, the social causes of obesity and the propagation of viruses through the Internet to name a few [31]. Looking at the degree distribution, dynamics and the community structure of the network as well as many other network analysis tools that have been developed, allows us to understand the complex system at a much deeper level and how local pairwise interactions contribute to the global structure of the system.

3 Information Theory, Causality and PMIME

This chapter shall discuss the basic concepts of information theory such as information, entropy and mutual information rate. Then transfer entropy, which is based on these information-theoretic tools, will be explored as well as its natural relation to Granger causality. Based on the strengths and weaknesses of transfer entropy, a new improved practical measure is introduced named PMIME which shall be the measure used for the EEG causality analysis.

3.1 Information

In 1948 Claude Shannon introduced the concept of information in his pioneering work 'A mathematical theory of communication' [36]. Shannon information, as it is known, describes the amount of uncertainty or randomness in a source signal.

Definition 3.1. (Shannon information) Let X denote a set of stochastic events and $p(x_i)$ the probability distribution of each $x_i \in X$. The information of x_i is defined as the negative log value of the probability of x_i :

$$I(x_i) = -\log p(x_i).$$

The Shannon information describes the uncertainty of a single event. When we want to describe properties between two or more events, we use the joint, conditional and mutual information. Let X and Y be two stochastic event sets.

Definition 3.2. (Joint information) Let XY be a two-dimensional joint event set. For any element $x_i y_i \in XY$, the joint information for the product event $x_i y_i$ is given by

$$I(x_i y_i) = \log p(x_i y_i)$$

where $p(x_i y_i)$ is the two-dimensional joint probability of $x_i y_i$.

Definition 3.3. (Conditional information) Let x_i and y_j be elements of X and Y respectively. The conditional mutual information of x_i given y_j is defined as

$$I(x_i | y_j) = -\log p(x_i | y_j)$$

Definition 3.4. (Mutual information) The amount of information an event $y_j \in Y$ provides to another event $x_i \in X$ is defined as the mutual information between x_i and y_j :

$$I(x_i; y_j) = \log \frac{p(x_i | y_j)}{p(x_i)}.$$

The joint information measures the uncertainty when all members of the product events simultaneously occur. The conditional information gives the amount of uncertainty of one event conditioned on another and the mutual information measures the common uncertainty shared between two events. These tools are very useful in describing single events and their relationship. However, more often we are interested in describing entire event sets. To do this we use entropy.

3.2 Entropy

Entropy measures the average uncertainty for an entire event set.

Definition 3.5. (Information entropy) Let X be a set of stochastic events and x_i be an element in X . The information entropy of X is the expectation of $I(x_i)$:

$$H(X) = E[I(x_i)] = E[-\log p(x_i)] = - \sum_{i=1}^q p(x_i) \log p(x_i).$$

The information entropy describes the average amount of uncertainty of an event set. If an event has zero probability, it has no contribution to the entropy as by definition $0 \cdot \log 0 = 0$. For a discrete event set, entropy is non-negative as each term in the summation is non-negative. Additionally, entropy reaches its minimum when $p_i = 1$ for some i and $p_j = 0$ (if $j \neq i$), and it reaches its maximum ($H(X) = \log n$) for a uniform probability distribution, i.e. $p_i = \frac{1}{n}, i = 1, 2, \dots, n$.

Definition 3.6. (Joint entropy) In the joint events set XY , the expectation of the joint information $x_i y_j$ is defined as the joint entropy:

$$H(x, Y) = \sum_{XY} p(x_i y_j) I(y_j | x_i) = - \sum_{XY} p(x_i y_j) \log p(x_i y_j)$$

Definition 3.7. (Conditional entropy) Let XY be the joint stochastic events set, the expectation of the conditional information $I(y | x)$ is defined as the conditional entropy of Y on X :

$$H(Y | X) = \sum_{XY} p(x_i y_j) I(y_j | x_i) = - \sum_{XY} p(x_i y_j) \log p(y_j | x_i).$$

Joint entropy is a measure of the average amount of information shared by both X and Y , whereas conditional entropy describes the average amount of information remained in one event set providing the other.

3.3 Mutual information rate

Definition 3.8. (Mutual information rate) Consider the arbitrary joint events set XY , the mutual information rate is defined as the joint probability expectation of the mutual information $I(x_i; y_j)$:

$$\begin{aligned} I(X; Y) &= \sum_{XY} p(x_i y_j) I(x_i; y_j) = \sum_{XY} p(x_i y_j) \log \frac{p(x_i | y_j)}{p(x_i)} \\ &= \sum_{XY} p(x_i y_j) \log \frac{p(y_j | x_i)}{p(y_j)} \end{aligned}$$

when X and Y are mutually independent, $I(X; Y) = 0$.

The mutual information rate measures the amount of information exchanged per unit time between the sets of stochastic events. If we have many nodes in a network representing time-series data, the mutual information rate between each pair of nodes can be a powerful tool in analyzing the complex system and the interaction between nodes. However mutual information rate cannot be used to determine the predominant direction of information flow given that it is a static, symmetric property. In order to analyze dynamical properties such as driving and responding, quantities based on transition probabilities have to be considered, leading to the introduction of transfer entropy.

3.4 Transfer entropy

Transfer entropy (TE) is an information-based causality measure, introduced by Schreiber to describe the information transfer between coupled systems [35]. A discrete process X , approximated by a k th order stationary Markov process, satisfies

$$p(x_{n+1} | x_n^{(k)}) = p(x_{n+1} | x_n^{(k+1)}) \quad (3.1)$$

where $x_n^{(k)} = (x_n, \dots, x_{n-k+1})$ is a k -dimensional delay embedding vector. Extending the system X to two systems X and Y , the generalized Markov property gives:

$$p(x_{n+1} | x_n^{(k)}) = p(x_{n+1} | x_n^{(k)}, y_n^{(l)}) \quad (3.2)$$

which means that in the absence of information flow from $Y \rightarrow X$, the state of Y has no influence on the transition probabilities of X . Transfer entropy is defined by a Kullback entropy to evaluate the generalized Markov property:

$$TE_{Y \rightarrow X} = \sum p(x_{n+1}, x_n^{(k)}, y_n^{(l)}) \log \frac{p(x_{n+1} | x_n^{(k)}, y_n^{(l)})}{p(x_{n+1} | x_n^{(k)})} \quad (3.3)$$

TE can both be expressed in terms of conditional entropies and in terms of mutual information rates.

$$TE_{Y \rightarrow X} = H(x_{n+1} | x_n^{(k)}) - H(x_{n+1} | x_n^{(k)}, y_n^{(l)}) \quad (3.4)$$

and

$$TE_{Y \rightarrow X} = I(x_{n+1}; \mathbf{v}_n) - I(x_{n+1}; x_n^{(k)}) \quad (3.5)$$

where $\mathbf{v}_n = (x_n^{(k)}, y_n^{(l)})$ is the embedding vector for the past of both X and Y . Therefore, TE can be viewed as an asymmetric extension to mutual information rate which uses conditional probabilities to measure the information transfer between systems. Hence, transfer entropy can be interpreted as the average uncertainty eliminated from x_{n+1} when the past of Y is presented providing the past of X . In order to discount the possibility that another random variable Z drives both Y and X , the partial transfer entropy (PTE) can be used. This can be expressed as:

$$PTE_{Y \rightarrow X|Z} = H(x_{n+1} | x_n^{(k)}, z_n^{(m)}) - H(x_{n+1} | x_n^{(k)}, y_n^{(l)}, z_n^{(m)}) \quad (3.6)$$

where Z is a discrete process representing the rest of the system not accounted for by X and Y and is approximated by an m th order stationary

Markov process. PTE clearly incorporates the idea of Granger causality. It provides a measure of the unique information contained in Y for the prediction of X and thus is positive when Y Granger causes X . TE and PTE have been used in a variety of applications from exploring information flow between financial time-series to providing feedback to improve the performance of artificial neural networks [44][19]. However, TE and PTE suffer from slow computation. With the same delay embedding with embedding dimension m (and delay τ) for X and Y , $TE_{Y \rightarrow X}$ requires the estimation of a joint probability distribution of dimension $2m + 1$ (m for X , m for Y and 1 for the future of X). For PTE, when K variables are observed, the dimension becomes $Km + 1$, and eventually PTE will fail for large m or K . This is the case in a lot of practical situations, including for an EEG. Therefore, a measure that addresses dimensionality reduction is required in our analysis. Such a measure was introduced by Kugiumtzis and is called partial mutual information with mixed embedding [21].

3.5 Partial Mutual Information with Mixed Embedding

Let $\{x_t, y_t, z_{1,t}, \dots, z_{K-2,t}\}_{t=1}^n$ be a multivariate time series of K variables given by $X, Y, Z_1, \dots, Z_{K-2}$, and we want to estimate the effect of Y on X conditioning on $Z = \{Z_1, \dots, Z_{K-2}\}$. The future of X at each time step t is represented by a vector of T feature values, $\mathbf{x}_t^T = [x_{t+1}, \dots, x_{t+T}]$. A maximum lag is set for each variable, e.g. L_x for X and L_y for Y . In our EEG analysis, as all variables are the same type, we make the assumption that the maximum lag, L , is the same for all variables. Let us denote the set of all lagged variables at time t as W_t , containing the components $y_t, y_{t-1}, \dots, y_{t-L_y}, x_t, x_{t-1}, x_{t-L_x}, z_{1,t}, \dots$ and the same for the other variables in the system with their range given by the maximum lag.

An iterative scheme is used to form a mixed embedding vector $\mathbf{w}_t \in W_t$ starting with an empty embedding vector, $\mathbf{w}_t^0 = \emptyset$ [42]. In the first iteration, the component in W_t being most correlated to \mathbf{x}_t^T is found by the K-nearest neighbors (kNN) estimate of mutual information. This means $w_t^1 = \operatorname{argmax}_{w \in W_t} I(\mathbf{x}_t^T; w)$, and we have $\mathbf{w}_t^1 = [w_t^1]$. In the second iteration, the component in $W_t \setminus \mathbf{w}_t^1$ that gives the most information about \mathbf{x}_t^T is found. This is done by finding $w_t^2 = \operatorname{argmax}_{w \in W_t} I(\mathbf{x}_t^T; w \mid w_t^1)$, where the conditional mutual information is again estimated by kNN. The mixed embedding vector then becomes $\mathbf{w}_t^2 = [w_t^1, w_t^2]$. This progressive scheme

continues until iteration j when the additional information of w_t^j is not large enough. This is quantified with the stopping criterion

$$I(\mathbf{x}_t^T; \mathbf{w}_t^{j-1})/I(\mathbf{x}_t^T; \mathbf{w}_t^j) > A. \quad (3.7)$$

Here, $A \in (0, 1)$ is a significance threshold near 1. The empirical optimum choice for A has been found to be 0.95, which not only allows the inclusion of a new component in the embedding vector even if it contains only a small amount of unique information about \mathbf{x}_t^T but it also prevents false positives [42]. The obtained mixed embedding vector \mathbf{w}_t can contain any of the lagged variables $X, Y, Z_1, \dots, Z_{K-2}$. To test whether Y causes X we look at how much of Y is contained in \mathbf{w}_t . Let us denote the components of Y in \mathbf{w}_t as \mathbf{w}_t^y , for X as \mathbf{w}_t^x and for the other variables in Z as \mathbf{w}_t^z . To quantify the causal effect of Y on X conditioned on the other variables in Z , we define partial mutual information with mixed embedding (PMIME) as

$$PMIME_{Y \rightarrow X|Z} = \frac{I(\mathbf{x}_t^T; \mathbf{w}_t^y | \mathbf{w}_t^x, \mathbf{w}_t^z)}{I(\mathbf{x}_t^T; \mathbf{w}_t)}. \quad (3.8)$$

The normalization means $PMIME_{Y \rightarrow X|Z} \in [0, 1]$. It is zero if there are no driving components in the mixed embedding vector ($\mathbf{w}_t^y = \emptyset$), meaning there is no direct causal effect from Y to X , and it is 1 if the mixed embedding vector is totally dominated by the driving variable ($\mathbf{w}_t^x = \mathbf{w}_t^z = \emptyset$). The measure is of a very similar form to PTE, but in PTE the uniform delay embedding vectors of X, Y and Z are used and the delay parameters have to be set. Therefore, PMIME addresses the issue of high dimensionality in PTE and so will hopefully be a suitable measure for capturing the interdependencies in an EEG. Software to compute PMIME has been written for MATLAB [21]. Inputting K multivariate time-series over a specified length of time, will produce a $K \times K$ causal matrix with the (i, j) entry indicating the causal strength from time-series, i to time-series j , where $i, j \in K$.

4 Networks and Community detection

This chapter shall introduce basic concepts from network theory such as representation, directed networks and networks with an associated strength between links. After this, an overview of community detection methods within these networks shall be given followed by a description of one of these methods known as the map equation. For the EEG analysis, having constructed a directed network using PMIME, the map equation is used to try and understand the flow of information in the brain and between persons.

4.1 Networks and representations

A network is simply a collection of connected objects. The objects are known as nodes or vertices, and the connections between the nodes are edges. A simple network cannot have edges connecting a node to itself (self-edge) and there is no more than one edge between a pair of nodes. On the other hand, a multi-network can have both self-edges and multi-edges. In this thesis, only simple networks are considered. The nodes in a simple network can be uniquely indexed by a list of integers, e.g. i stands for the i th node, $i = 1, 2, \dots, N$ where N is the total number of nodes. In a simple undirected network, the link between nodes indexed by i and j is represented by an unordered pair (i, j) . Therefore, a simple undirected network $G = (N, L)$ can be specified by the number of nodes N and a list of unordered pairs, L . Additionally, a simple undirected network can be presented by an $N \times N$ adjacency matrix $A = (a_{ij})$ with entries of the form

$$a_{ij} = \begin{cases} 1 & \text{if there is a link between nodes } i \text{ and } j, \\ 0 & \text{otherwise} \end{cases} \quad (4.1)$$

Here, 1 denotes the existence of a link between nodes i and j and 0 represents the absence of such a link. The adjacency matrix of a simple undirected network is symmetric with vanished diagonal elements. However, in a lot of real-world networks such as the World Wide Web, Food webs and Economic networks, the interaction between nodes i and j goes one way and not the other. For example, in a food web, there is an edge between whales, i and plankton, j due to i eating j , however plankton do not eat whales! So the edge is directed. These are known as directed networks and need to be specified slightly differently.

4.2 Directed networks

A simple directed network $G = (N, L)$ is a simple network consisting of N nodes and a set of **ordered** pairs of nodes, L . The set of ordered pairs of nodes, is a list of direct links, in which each (i, j) indicates a direct link from node j to node i . A direct network can also be represented by a $N \times N$ adjacency matrix with entries

$$a_{ij} = \begin{cases} 1 & \text{if there is a link from node } j \text{ to } i, \\ 0 & \text{otherwise} \end{cases} \quad (4.2)$$

The adjacency matrix of a simple directed network is asymmetric with vanished diagonal elements. In many real networks, there is also a strength of connection between nodes. In order to describe systems with this property, we use weighted networks.

4.3 Weighted networks

As well as specifying links between nodes, a simple weighted network additionally assigns each link a real number, i.e. a weight or strength. A simple weighted network can be represented by assigning the elements of the adjacency matrix as the weights of connections. For example, the adjacency matrix of a simple weighted, directed network is given by

$$a_{ij} = w \quad (4.3)$$

where w is the weight of the link from node j to i . Most frequently, w is given a value of zero when there is no directed edge between the nodes.

4.4 Overview of community detection

Real-world networks are often extremely large and difficult to visualize. This can be seen with a network of the Internet in Figure 1. The mapping consists of frequent trace-route style path probes, one to each registered Internet entity. From this, the graph is built showing the paths to most of the nets on the Internet. The challenge is to extract useful information buried in the large amount of nodes and edges. We want to find tools that can simplify and highlight important structures in these networks so that we can understand their organization. Such tools are called community detection methods and they are designed to identify groups of strongly connected nodes. These groups of strongly connected nodes are called

communities, clusters or modules and are represented by different colors in the Internet network. Community detection in networks is challenging, and many algorithms have been proposed in the last few years to tackle this difficult problem. Most of these algorithms fall into three main categories; Null models, block models and flow models:

- Algorithms based on null models compare some measure of connectivity within groups of nodes to the expected value in a proper null model. Clusters are identified as the sets of nodes where the connectivity deviates the most from the null model. This is the procedure for the commonly used Louvain method [10].
- Algorithms based on block models identify blocks of nodes with common properties. Nodes that are assigned to the same block are statistically equivalent in terms of their connectivity to nodes within the block and to other blocks.
- Algorithms based on flow models use the fact that networks, through their edges, capture the flow between components of a real system. In these models, communities are defined as structures within the network where flow persists for a long time once entered. In this thesis, the map equation, which is a flow-based method, shall be used to find community structure in the constructed network.

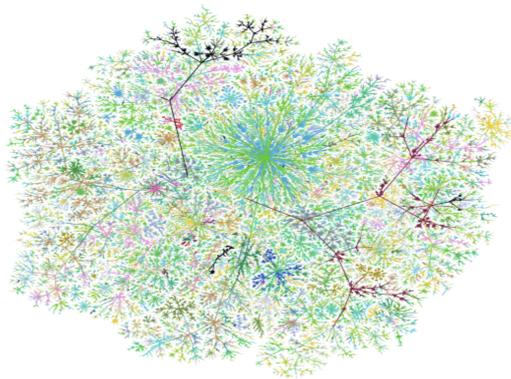


Figure 1: The Internet Network as found by Bell Labs during the Internet mapping project [8]. Colors represent a community.

4.5 The map equation

For most integrated systems, there is a flow of some form; passengers traveling among airports, money transferred between banks, signals transmitted in the brain, and this flow connects the components of a system and generates their independence. Therefore it is natural, especially when looking at information flow in a complex system, to understand the system's behavior by looking at the dynamics on the network. This leads to a flow-based method, such as the map equation, seeming like a good approach to capture the underlying communities that we wish to find. The map equation, first introduced in [33], takes advantage of the duality between finding community structure in networks and minimizing the description length of a random walker's movements on a network guided by the possibly weighted, directed links of the network. The sender wants to communicate to a receiver about the movement on the network, and for a given partition of communities, there is an associated information cost for describing the dynamics. The map equation is designed such that the description length of the random walker can be compressed if the network has regions where the random walker tends to stay for a long time [4]. Therefore, with a random walker as a proxy for real flow, the map equation is a direct measure of how well a given network partition captures modular regularities in the network. Regions in the network where the flow tends to stay for long periods of time can be thought of as communities and the map equation is able to capture the best partition of the network to find these regions of concentrated flow. For the interested reader, [32] provides a more mathematical description of the map equation and the intricate details of how it is used to find community structure in directed, weighted networks. Due to the map equation's natural relation to information flow and its previous success in finding communities in citation networks [32], the Infomap algorithm, which is based on the map equation framework, shall be used in finding communities in the directed, weighted cross-brain networks that we produce from PMIME.

5 Kolmogorov complexity and its approximation

This section discusses Kolmogorov complexity, a theoretical measure of the randomness in a signal. Then, Lempel-Ziv complexity is introduced and its practicality for approximating the theoretical Kolmogorov complexity in a time-series. An extension of Lempel-Ziv complexity, called permutation Lempel-Ziv is outlined, as well as why it may outperform its predecessor. The permutation Lempel-Ziv complexity will be used as a measure of the complexity at a local electrode in the EEG for the strict and let-go modes. This will provide a description of the local functional behavior of the brain during the musical performance and this could potentially be different in each mode.

5.1 Kolmogorov complexity

Kolmogorov complexity tries to answer the fundamental question: ‘what is a random signal?’. Consider the following two strings

2222222222

8435427952

Intuitively, the second string seems random and the first does not. However, from the perspective of probability, both strings have the same probability of being chosen when we take a string of 10 digits fully at random namely, each of them has probability 10^{-10} . So probability does not explain the intuitive notion of randomness. To explain the intuition, Kolmogorov stated that less random sequences can be described with fewer words than more random sequences [20]. For instance, the first string can be described by ‘repeat 2’ whereas the second string would need all ten characters to describe - the string cannot be compressed. To quantify this, the Kolmogorov complexity is the length of the shortest computer program (in a predetermined programming language) that produces the string as output. By the invariance theorem, the choice of programming language or the model is not crucial to the complexity of the string [24]. However, for almost all strings, it is not possible to compute the exact value of Kolmogorov complexity and so a variety of metrics have been proposed in order to efficiently approximate it [17].

5.2 Lempel-Ziv complexity

The metric of complexity proposed by Lempel and Ziv (LZ) has been extensively used to solve information theoretic problems and has been applied to areas such as coding [16], data compression [45] and for the generation of test signals [41]. Recently, LZ has been applied extensively in biomedical signal analysis as a metric to estimate the complexity of discrete-time physiological signals. Alike in this thesis, LZ has been used in the analysis of EEG signals. The LZ of the time-series of electrical activity at electrodes was found to be significantly higher in Schizophrenic than healthy patients [12], and lower than a healthy test group for patients suffering from Alzheimer's disease with (p-value < 0.01) [15]. LZ approximates the Kolmogorov complexity of a finite sequence containing a finite alphabet. The alphabet refers to the different numbers, letters or symbols that the sequence contains. LZ links the complexity of the specific sequence to the gradual build up of new patterns along the given sequence. It is computed by scanning the sequence S from left to right and then increasing a complexity counter $c(n)$ by one unit every time a new subsequence of consecutive characters is encountered. An algorithm for calculating LZ is as follows:

1. Let P and Q denote two subsequences of the sequence $S = s(1)s(2)...s(n)$ and PQ to the concatenation of P and Q . Additionally, let π denote the operation of deleting the last character so that $PQ\pi$ is equivalent to PQ with a deleted end character. Moreover, let $v(PQ\pi)$ denote the vocabulary of all different subsequences of $PQ\pi$. At the beginning $c(n) = 1, P = s(1), Q = s(2)$ and therefore $PQ\pi = s(1)$.
2. In general, $P = s(1)s(2)...s(r)$ and $Q = s(r+1)$, then $PQ\pi = s(1), s(2), \dots, s(r)$. If Q belongs to $v(PQ\pi)$ then Q is a subsequence of $PQ\pi$ and is not a new sequence.
3. Renew Q to be $s(r+1)s(r+2)$ and see if Q belongs to $v(PQ\pi)$ or not.
4. Repeat the previous step until Q does not belong to $v(PQ\pi)$. This implies that $Q = s(r+1)s(r+2)...s(r+i)$ is not a subsequence of $PQ\pi = s(1)s(2)...s(r+i-1)$. When this occurs, increase $c(n)$ by one.
5. P and Q are then renewed to be $P = s(1)s(2)...s(r+i)$ and $Q = s(r+i+1)$.

The above procedure is repeated until Q is the last character. Applying this algorithm to the sequence $S = 0001101001000101$ gives

$$0 \cdot 001 \cdot 10 \cdot 100 \cdot 1000 \cdot 101$$

where dot products indicate points where P and Q are both renewed and $c(n)$ is increased as a new sequence is found. In this case $c(n) = 6$. This complexity measure is dependent on the sequence length. To avoid this, $c(n)$ is normalized so that it describes the rate at which new sequences are found rather than the number of new sequences. If the length of the sequence is n and the number of different symbols in the alphabet is α , it has been proved that the upper bound of $c(n)$ is given by

$$c(n) < \frac{n}{(1 - \epsilon_n) \log_\alpha(n)} \quad (5.1)$$

where ϵ_n is a small quantity and goes to 0 as $n \rightarrow \infty$ [23]. Therefore, in general the upper bound of $c(n)$ is

$$b(n) = \frac{n}{\log_\alpha(n)}. \quad (5.2)$$

The normalized LZ complexity $C(n)$ can then be given by

$$C(n) = \frac{c(n)}{b(n)} \quad (5.3)$$

and reflects the rate of new patterns in the sequence. In order to practically implement the LZ measure on a discrete time-series signal, the temporal sequence must be converted to a sequence with a finite alphabet. This is most often done by converting the time-series to a 0-1 signal with the threshold determined by measures such as the mean, median or k-means. For instance, with the popular LZC_{mean} the mean value of the time series is selected as T_d . The binary sequence is then expressed as,

$$x_i = \begin{cases} 0 & \text{if } s(i) < T_d, \\ 1 & \text{otherwise} \end{cases} \quad (5.4)$$

Converting the time-series to a binary sequence then allows the LZ complexity measure to be applied. However, by converting the time-series to a binary sequence, a lot of information in the original signal is lost. In this thesis, the permutation LZ complexity will be adopted in order to reduce some of the information lost when converting the time-series to a binary sequence.

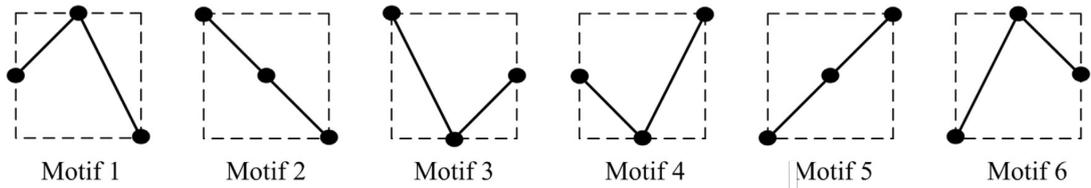


Figure 2: The 6 different possible motifs when the number of data points, $m = 3$.

5.3 Permutation Lempel-Ziv complexity

Permutation LZ uses order patterns to generate sequences with a finite alphabet. For instance, looking at a window with the first 3 numbers in the time-series and their ordering in terms of size. There are $3!$ motifs, i.e. ways of arranging the 3 numbers as shown in Figure 2. Therefore, the first three numbers and their order can be described by a number between 1 and $3! = 6$. The window is then shifted along the time-series by a parameter τ and the 3 numbers in that window are then described with an integer in the same range. In this procedure there are two parameters, m being the number of data points in each motif and τ determining the sample points spanned by each section of the motif. Using this procedure, the time-series can be converted into a sequence with a finite alphabet with integer symbols ranging from 1 to $m!$. This should provide a sequence that retains more information about the original signal than a conversion to a binary sequence would. Once this sequence is generated, then the LZ measure can be applied using the algorithm listed for the Lempel-Ziv complexity. Permutation LZ has been found to be more sensitive to dynamical changes in the underlying signal than standard LZ and has also been successful in detecting anesthesia states [1]. For this reason, the measure shall be used for our complexity analysis on the EEG data.

6 EEG analysis on music improvisation

6.1 Experimental setup

The music improvisation experiment was undertaken at the Data Science Institute, Imperial College London on the 21 March 2017. Synchronized EEG measurements were taken from three musicians; a flutist, pianist and a singer, as well as four audience members during a musical performance. The EEG recorders for each person, were set up according to the international 10-20 system and had 19 electrodes: Fp1, Fp2, F3, F4, F7, F8, C3, C4, T7, T8, P3, P4, P7, P8, O1, O2, Fz, Cz, Pz. The electrodes represent different large brain regions, F: frontal lobe (attention and executive control), C: central lobe (sensory and motor function), P: parietal lobe (perception, multi-sensory integration), O: occipital lobe (Visual processing). Odd numbers represent locations on the left brain, while even numbers stand for locations on the right. A reference electrode (CPz) was used, so that each EEG signal was mono polar referenced to this site and activity levels of the 19 sites could be compared relative to each other. The sampling frequency of the data acquisition was 250 Hz.

During the performance, each piece was performed twice: once in what the musicians described as a strict mode, corresponding to a prepared rendition and once in a let-go mode, corresponding to an improvised interpretation. To further clarify the difference between the modes, the musicians in the strict mode were mainly focused on controlling technical precision, timing co-ordination, accuracy of the score's details, avoiding risks, whilst also giving a convincing performance. In the let-go mode, the musicians played more freely, expressing themselves spontaneously and putting less focus on accuracy of note playing. The order of the strict and let-go modes for each piece was randomly varied by the musicians, with the audience unaware of which version was being played. Two of the audience members could both see and hear the performances, the other two could hear but not see them. Additionally, within each pair, one participant had a high degree of training in classical music, the other had a low degree.

6.2 Methods

Here, we aim to use PMIME to create a directed network of information flow between electrodes in multiple EEGs. This will give us a cross-brain network which can then be analyzed. A complexity analysis will also be used to quantify signals at an electrode scale.

6.2.1 Causality analysis

It was found that the Fp1, Fp2 and the Fz electrodes contained some anomalous data. Therefore, these electrodes were removed from the data set leaving time-series from 16 electrodes for each person. The EEG data from all seven different brains were then put together to form an 112 channel (16×7) augmented data file, one for the strict mode performance and one for the let-go performance. The file was edited such that all electrode recordings were synchronized, forming a 63332×112 file for the let-go mode and a 63129×112 file for the strict. These were then passed through a 2Hz high pass filter. The data files were then split by constant size moving time windows with $\Delta T = 4s$. Given that the sampling rate of the EEG was 250 Hz, this meant that both files now contained sixty-three 1000×112 blocks each corresponding to four seconds worth of synchronized EEG measurements for all people. Each of these blocks were then individually inputted to the PMIME software [21], which generated a 112×112 causality matrix for each block. These causality matrices can be thought of as a weighted, directed adjacency matrix with the value at position (i, j) corresponding to the strength of interaction from $i \rightarrow j$ and can therefore be viewed as a network. Consequently, both the strict and let-go modes will produce 63 cross-brain networks corresponding to interactions at 4 second intervals.

The PMIME software has various parameters which need to be set; the maximum delay to search for components for the mixed embedding vector (L_{max}), the number of time steps T in the vector of feature values, $\mathbf{x}_i^T = [x_{t+1}, \dots, x_{t+T}]$, the number of nearest neighbors for the density estimation of mutual information (nnei), the number of surrogates for a significance test (nsur) and the significance level to test for the termination criterion of the mixed embedding scheme (α). These were all set to their recommended values [21], namely $L_{max} = 5$, $T = 1$, nnei = 5, nsur = 100 and $\alpha = 0.05$.

6.2.2 Cross-brain network analysis

Once the 63 causality matrices were found for the strict and let-go modes, they were then turned into graphs using the networkx library in Python. We now have a set of weighted, directed networks showing the information flow every 4 seconds between all electrodes in the system for both modes. Some network back-boning was then done in order to reduce noise. This was achieved by removing any links between nodes that have a strength below a certain threshold. A histogram of the distribution of weights in all the strict networks and another in all the let-go networks was plotted as in Figure 3. Both these distributions are quite similar, although in the strict mode the distribution is slightly wider and flatter. The threshold was visually chosen to be 0.015 to remove links with a small causal strength that may have only been found from noise in the EEG. The network of the first 4 seconds in the strict-mode is shown in Figure 4 with the node color reflecting the community in which the node is.

Infomap, an algorithm for the map equation, was then used to find the communities in each of the directed, weighted networks. Plots of the number of communities over time for both the strict and the let-go mode can then be done from the results of Infomap. Additionally, as can be seen at the first time-step for the strict mode, most of the nodes fall into one community. Plots of how the size of the largest cluster varies over time for both the let-go and strict modes were done in order to get a sense of the connectedness of the network. A large maximum cluster size implies that a majority of the nodes are similar and well-connected, whereas a small one means that there are lots of significant regions on the network with restricted flow.

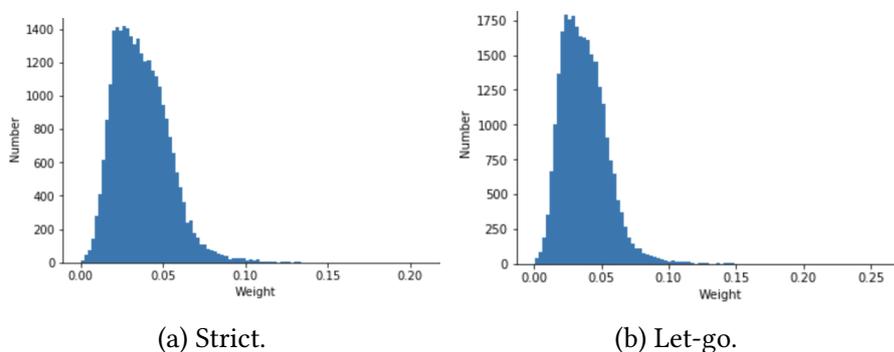


Figure 3: Histogram of the weights in all of the causal matrices for each mode. The number of bins is set at 100.

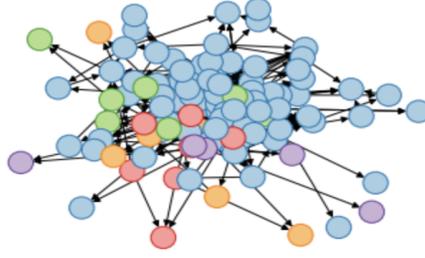


Figure 4: A directed network indicating information flow in the first 4 seconds of the strict mode. The color of the node represents the community to which it belongs

To further get a sense of the similarity between communities, the entropy of being in a particular community was calculated at each time step. This was done by finding the probability of being in community i (p_i) and calculating the entropy:

$$entropy = \sum_i -p_i \log(p_i) \quad (6.1)$$

This gives a more fine-grained picture than the size of the largest cluster. More communities with more nodes cause a high entropy, whereas a single large community implies a low entropy. The entropy was then plotted over time for both the strict and let-go modes.

An important characteristic of improvisation, is that the risk-taking and support are provided spontaneously by the musicians. Hence, one may expect the musicians to be more actively engaged with each other during the let-go mode. We hypothesis that in the let-go mode the neural networks may be more widely-distributed between musicians. Therefore, once we obtained the desired plots for the cross-brain network between all people, we then looked at just the cross-brain network between the musicians and repeated our analysis to see if our assumptions about differences in interactions between the modes can be quantified.

6.2.3 Complexity analysis

For the complexity analysis, differences in the let-go and strict modes were analysed for each person individually. The EEG measurements for a particular person in a particular mode, contain 16 time-series, one for each electrode minus the three with anomalous data, and were partitioned into blocks of length N . In our analysis, N was chosen to be 1000, so that

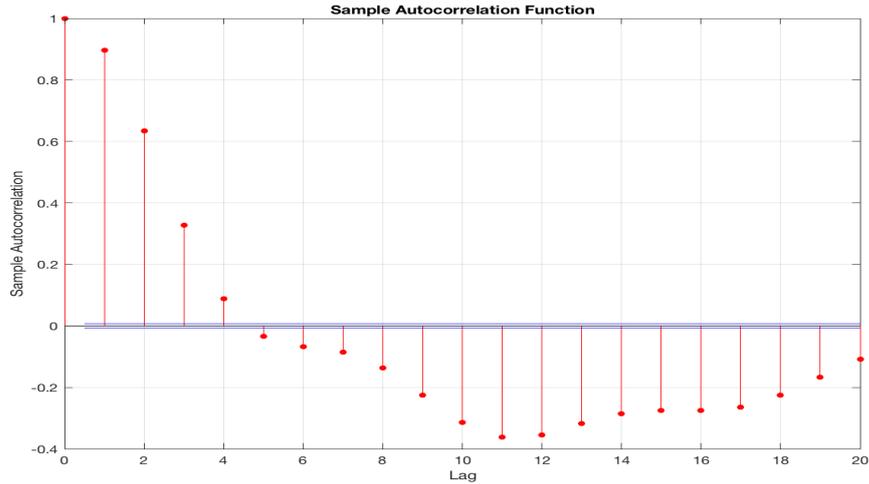


Figure 5: The normalized autocorrelation function for the electrical activity at the Flutist’s F3 electrode

we have blocks containing 16 time-series each of length 4s. The permutation Lempel-Ziv complexity was then calculated for each time-series within the block. This was then repeated for every block, giving a complexity value for each electrode every 4s. The complexity values for each electrode were then averaged at a particular time interval so that we get a time-series of complexity values representing the mean complexity across the EEG for that 4 second interval.

As mentioned, the permutation Lempel-Ziv complexity has two parameters; m being the number of data points in each motif and τ determining the sample points spanned by each section of the motif. The parameter τ was chosen using an autocorrelation function (ACF). ACF computes the cross-correlation of a signal with itself and when the ACF decays to e^{-1} of its peak value, the corresponding lag is found to be the optimal value of τ [28]. By plotting the normalized ACF of multiple time-series, one of which is shown in Figure 5, and finding the average lag when the ACF decays to e^{-1} of its peak value, we find that $\tau = 3$ is optimal. In the permutation process, m is usually recommended to be a value in the range 3-7 [2]. When $m < 3$, there would be too few possible patterns and the permutation would not make sense, while for $m > 7$, the computation would be too complex and expensive given $m!$ possible values will be scanned through at each window. Furthermore, to ensure that every possible ordinal motif occurs in the signals of length 1000, the condition $m! \leq 1000 - (m-1)\tau$ must hold. So, in order to be sensitive to the dynam-

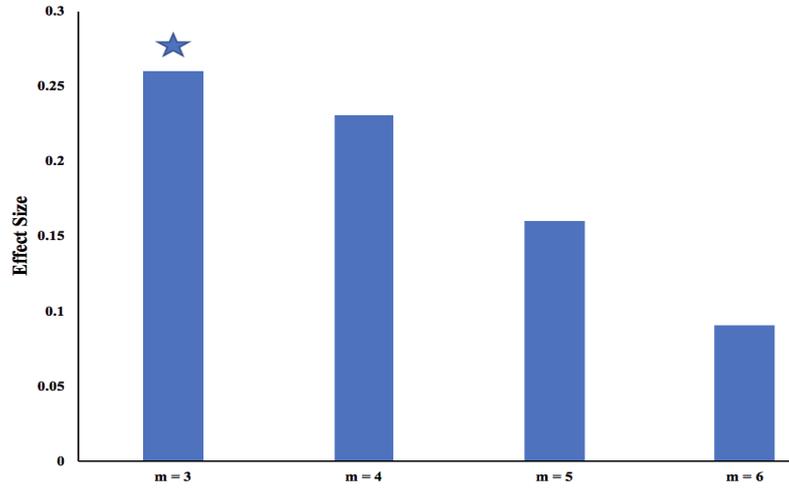


Figure 6: The effect size for different numbers of motifs, $m!$. Stars indicate the level of significance of the effect size from a permutation test.

ics of the system and to economize computational time, a large value of m should not be selected. For the flutist, values of m between 3 and 6 were all used with $\tau = 3$ for both the strict and the let-go mode. The effect size (Cohen's d) for T-test between the time-series of complexity values in the strict and let-go mode was calculated for each m . The Cohen's d between groups 1 and 2 with means M_1 and M_2 and standard deviations SD_1 and SD_2 is given by the equation

$$\text{Cohen's } d = \frac{M_2 - M_1}{SD_{pooled}} \quad (6.2)$$

where

$$SD_{pooled} = \sqrt{\frac{SD_1^2 + SD_2^2}{2}} \quad (6.3)$$

The effect size gives a standardized difference between the means of the complexity in the let-go and strict modes. In this case, a positive effect size implies that the mean of the complexity in the let-go mode is greater, and a negative effect size occurs when the mean is greater in the strict-mode. The p-value of the effect size was then found using a permutation test. The results are displayed in Figure 6. The m which showed the most significant change between the let-go and strict modes was selected, and this corresponded to $m = 3$ which has been previously found to be a good choice for $N = 1000$ [1]. Therefore, for the rest of the complexity analysis we used $m = 3$ and $\tau = 3$. The rest of the analysis was done in the same manner as with the flutist. The effect size between the complexity time-

series in the strict and the let-go modes was calculated, and then the p-value of this was computed with a permutation test.

6.3 Results

Here we present our results regarding the analysis of the cross-brain networks, from which we managed to identify neural differences between the two modes. Additionally, our complexity analysis shows differences in the local behavior at electrodes when improvisation occurs.

6.3.1 Causality and network analysis

To analyze the pattern of interaction between musicians and audience members, we have constructed a cross-brain network containing all the people in our analysis. Additionally, we have also constructed a network ignoring interactions with the audience to see if the musicians co-ordinate more during improvisation than in a strict mode as we hypothesized.

How the number of communities changes over time for both these cross-brain networks when the strict mode and let-go mode are being performed is shown in Figure 7. The x-axis represents time where the networks 1-63 correspond to 4 second intervals and the number of communities is found for each network using the Infomap algorithm.

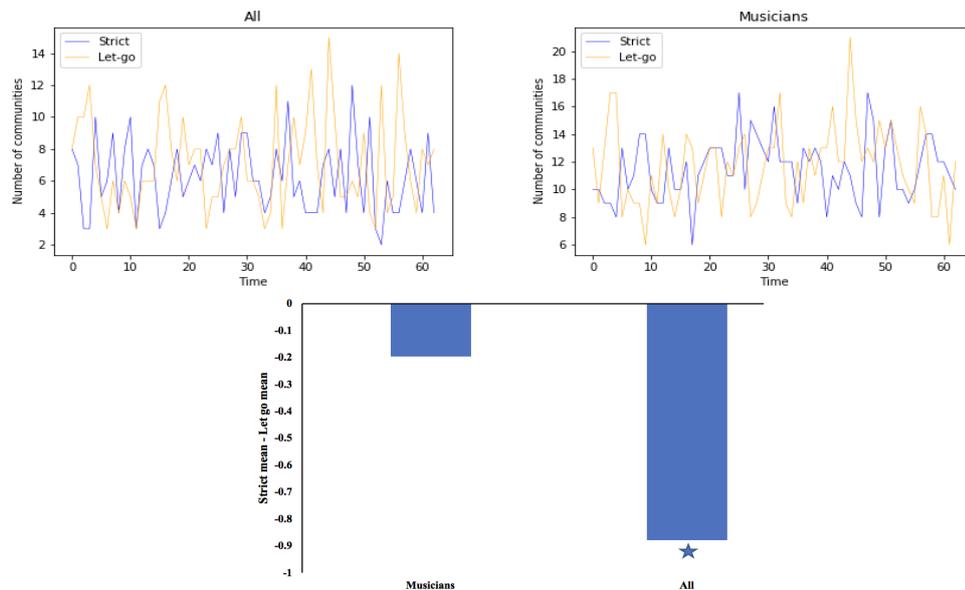


Figure 7: The number of communities as a function of time for the the two cross-brain networks. The differences in mean between the two modes is also shown.

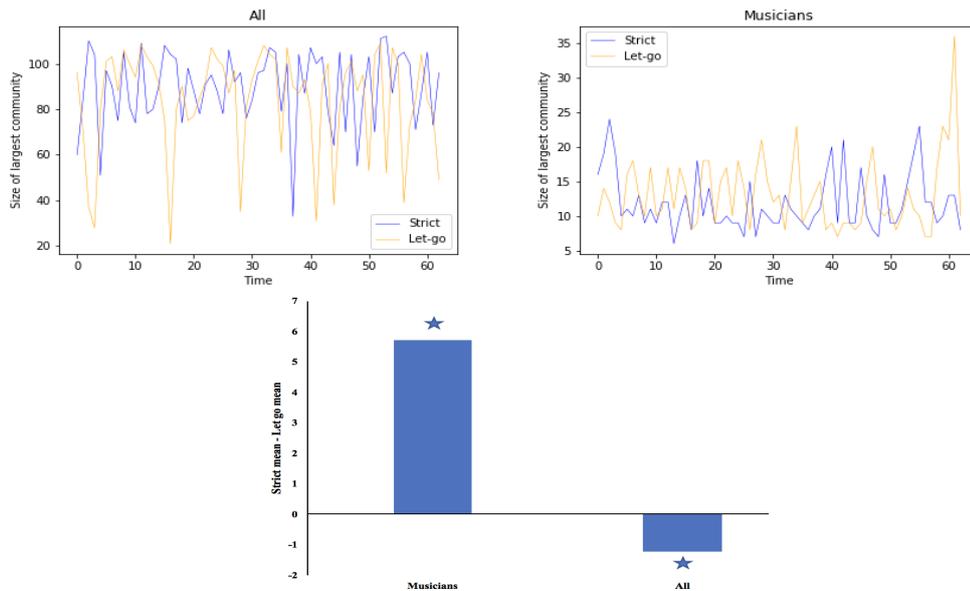


Figure 8: Plots of the size of the largest community over time for the the two cross-brain networks. A plot of the differences in mean between the two modes is also shown.

The figure also indicates the difference between the mean number of communities in each mode (Strict mean – Let-go mean) for the two different cross-brain networks as well as the level of significance of this difference calculated from a permutation test. In both the Musicians and All cross-brain networks, the mean number of communities is higher in the let-go mode. However, this is only significant when taking into account all people.

As previously discussed, the size of the largest cluster is a good measure of the connectedness of the network, i.e. how easily the flow spreads evenly throughout the network. Figure 8 shows how the size of the largest cluster varies over time for the cross-brain networks in both modes. The two different networks exhibit very different behavior. For the network considering everyone, the size of the largest cluster is statistically higher in the strict mode and there are lots of downward spikes in both modes where there is a sudden drop in the size of the largest cluster.

On the other hand, looking at just the musicians, the mean of the let go is significantly higher and the spikes are upwards where there is a sudden burst of connectivity between a large group of nodes.

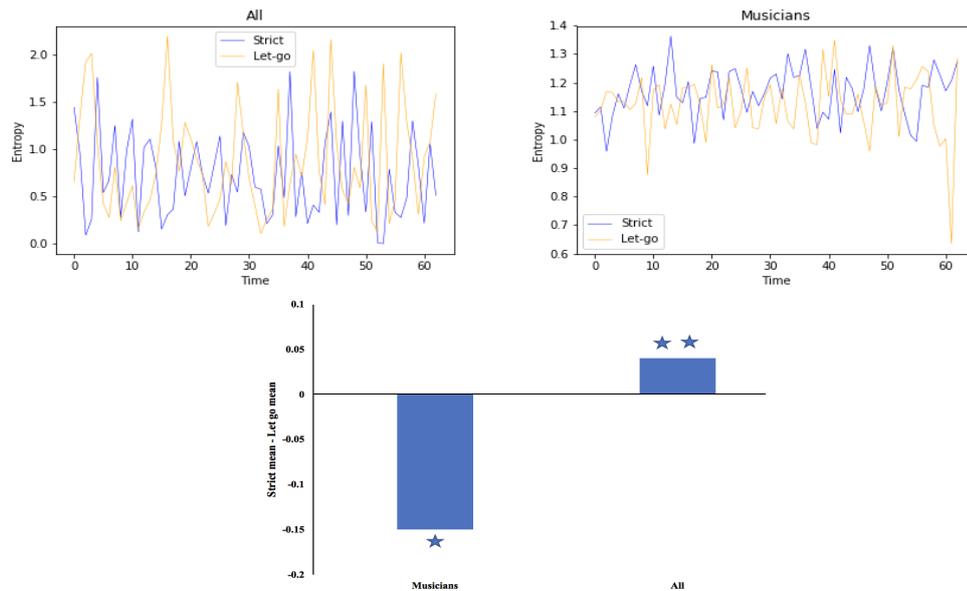


Figure 9: Plots of the entropy of the community structure over time for the the two cross-brain networks. A plot of the differences in mean between the two modes is also shown.

This analysis of community structure in both cross-brain networks detects a difference between the act of improvisation and that of playing a mechanical rendition.

To get a more fine-grained idea of the dynamics on the network and thus how information spreads between persons, the entropy of being in particular communities is plotted over time for the different cross-brain networks in both modes. The results are shown in Figure 9. Again, the inclusion of the audience into the cross-brain network seems to change the dynamics of the information flow to a large extent. The strict mode has a lower entropy than the let-go mode in the network with everyone and a higher entropy in the case of just the musicians and is statistically significant in both cases.

6.3.2 Complexity analysis

In the complexity analysis, the permutation Lempel-Ziv complexity is calculated for an individual at each electrode for a 4 second interval. The complexity is then averaged across all electrodes at a particular interval, giving a time-series of the average complexity value. An example of how this complexity value changes in time is shown for the pianist in both modes in Figure 10. The effect size between the mean of the time-series of complexity in the let-go and strict modes is then calculated for all the individuals with a positive effect size indicating a higher mean in the let-go performance.

Figure 11 shows the effect size found for each person and with a level of significance from a permutation test. As can be seen, everyone apart from Audience 3 shows an increase in complexity in the let-go mode. Additionally, a lot of the individuals have a statistically significant change between the modes and so the complexity measure seems to be able to distinguish neural differences when improvisation occurs.

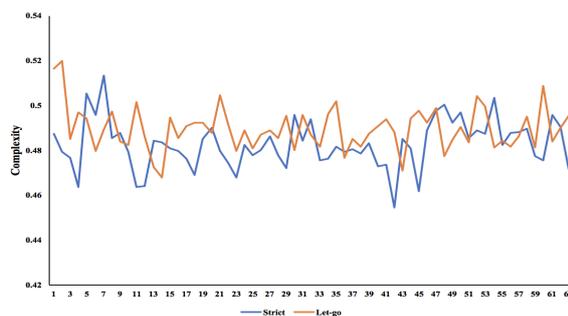


Figure 10: The average complexity over all electrodes as a function of time in both the strict and the let-go modes for the pianist

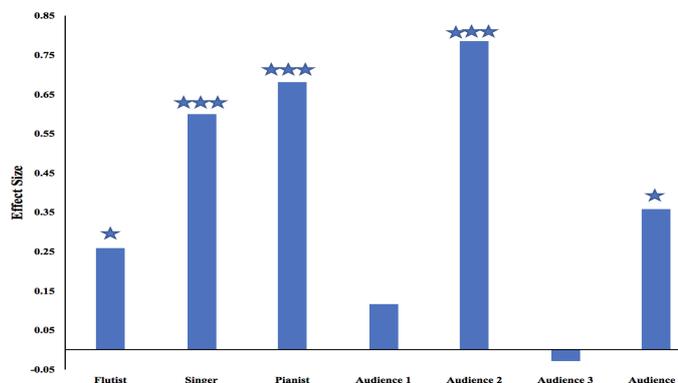


Figure 11: The effect size for the different people between the means of the Let-go mode – means in the Strict mode. Stars indicate the level of significance of the effect size from a permutation test.

6.4 Discussion

We have constructed a cross-brain network involving everyone's synchronized EEG measurements during the concert, by using the causality measure PMIME to specify interactions between pairs of electrodes in the whole system. Our results show that on average in the let-go mode, there is a larger number of communities, the size of the largest cluster is smaller and the entropy is higher than it is during the strict mode. These results are also all statistically significant with $p \leq 0.05$. Therefore, with our causality and network analysis, we are able to distinguish an improvising state through the interactions between musicians and audience members. A higher entropy in the let-go mode, suggests that the communities are less distinguishable in terms of their size when improvising (they are closer to a uniform distribution rather than having a single large cluster). This means that in the let-go mode there are more equally sized pockets of concentrated information flow and not a large connected cluster involving lots of functional parts in and between persons. This perhaps suggests that information flow between everyone is less diverse than in the strict mode and there are similar groups of functional parts doing their own tasks.

In the cross-brain network created from just interactions in the musicians, the size of the largest cluster is greater in the let-go mode ($p \leq 0.05$), and the entropy is higher in the strict mode ($P < 0.01$). This suggests that there is a high degree of connectedness for a large number of nodes in the network. We hypothesized that we would expect the musicians to be more actively engaged with each other in the let-go mode as a musician would need to be very aware of the other musicians, reacting and co-ordinating their performance. This seems to be confirmed with our analysis as there is a more diverse set of clusters in terms of size in the let-go mode. There tends to be a large cluster so in the let-go mode the information flow smoothly moves over a large part of the system showing a high degree of connectedness. Again we are able to differentiate between the modes of playing from our causal and network analysis of the interactions between the musicians. However, removing the audience members seems to give the opposite community structure in the network. It is unclear exactly why this is the case and further collaboration with Neuroscientists and musicians will have to be done to understand the cause.

In our complexity analysis we have calculated the average permutation Lempel-Ziv complexity over all intra-brain electrodes every 4 seconds. The results show that there is a statistically significant increase in the complexity for the flutist ($p \leq 0.05$), the singer ($p \leq 0.001$), the pianist ($p \leq 0.001$), audience 2 ($p \leq 0.001$) and audience 4 ($p \leq 0.05$). There is also an increase for audience 1 but not to a significant level. The only person involved in the concert where there is not an increase is for audience 3, where there is very little change in complexity when improvisation occurs. Lempel-Ziv complexity is associated with awareness and alertness, so it is unsurprising that the musicians all have an increase in this complexity when improvising. Additionally, you would expect this increase to be more significant than with the audience, who are not having to be aware and react in order to co-ordinate performance. This is the case with our results where only 50% of the audience show a significant increase in complexity in the let-go mode whilst 100% of the musicians show an increase. Interestingly, there is also an increase in complexity in general for the audience in the let-go mode. This may suggest that the state of mind during improvisation is communicable between the musicians and the audience and could lead to a heightened quality of shared experience.

Conclusions from the causality-network and the complexity analysis need to be made with caution. There are only 7 people in our analysis and differences in the let-go and strict mode are made for just one musical rendition which is approximately 5 minutes long. This small sample space may lead to over-fitting conclusions and so in future studies more participants could be employed and we could have both longer and more pieces of music. Additionally, although we have removed the Fp1, Fp2 and the Fz electrodes due to seeing anomalies, the EEG data could be further cleaned. This can be done by further filtering and using independent component analysis to remove artifacts such as blinking. This could prevent our conclusions being made on the basis of noise in the signal and future analysis should be done on thoroughly cleaned EEG.

6.5 Conclusion

A cross-brain network between the musicians as well as a network between everyone in a concert have been constructed using PMIME to assess the causal strength of all pairwise interactions between electrodes in the synchronized EEG data. The community structure of these networks has then be analyzed. Additionally, a complexity analysis has been done looking at the randomness of the signal at local electrodes. These analysis have been done when the musicians play in a let-go mode and in a strict mode. In both our complexity analysis and our causality-network analysis, we have found neuronal differences between the musicians improvising and when they are playing a mechanical rendition of music. The differences can be difficult to interpret, but there is a quantifiable difference in the information flow in the whole system of people, information flow between musicians and the local complexity of signals at electrodes. Further work in collaboration with Neuroscientists and musicians should be done to understand the found differences and why including the audience in the cross-brain network changes the community structure in the let-go mode from one with on average high connectedness and a few large communities to one with lots of equally sized communities. Additionally, repeat analysis should be done on a larger sample and with more performed pieces in order to confirm the findings. The method that we have undertaken in terms of constructing a network of information flow between time-series and then analyzing the network can be extended to other complex systems consisting of multivariate time-series such as climate and financial systems.

References

- [1] Yang Bai, Zhenhu Liang, and Xiaoli Li. A permutation lempel-ziv complexity measure for eeg analysis. *Biomedical Signal Processing and Control*, 19:102 – 114, 2015.
- [2] Christoph Bandt and Bernd Pompe. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.*, 88:174102, Apr 2002.
- [3] Roger E. Beaty. The neuroscience of musical improvisation. *Neuroscience and Biobehavioral Reviews*, 51:108 – 117, 2015.
- [4] Ludvig Bohlin, Daniel Edler, Andrea Lancichinetti, and Martin Rosvall. *Community Detection and Visualization of Networks with the Map Equation Framework*, pages 3–34. Springer International Publishing, Cham, 2014.
- [5] Edward Bullmore and Olaf Sporns. Complex brain networks: Graph theoretical analysis of structural and functional systems. 10:186–98, 03 2009.
- [6] Gyorgy Buzsáki. Rhythms of the brain, 01 2009.
- [7] Heather Chapin, Kelly Jantzen, J. A. Scott Kelso, Fred Steinberg, and Edward Large. Dynamic emotional and neural responses to music depend on performance expression and listener experience. *PLOS ONE*, 5(12):1–14, 12 2010.
- [8] Bill Cheswick, Hal Burch, and Steve Branigan. Mapping and visualizing the internet. In *Proceedings of the Annual Conference on USENIX Annual Technical Conference, ATEC '00*, pages 1–1, Berkeley, CA, USA, 2000. USENIX Association.
- [9] Dante R. Chialvo. Critical brain networks. *Physica A: Statistical Mechanics and its Applications*, 340(4):756 – 765, 2004. Complexity and Criticality: in memory of Per Bak (1947–2002).
- [10] P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti. Generalized Louvain Method for Community Detection in Large Networks. *ArXiv e-prints*, August 2011.

- [11] Chamila Dissanayaka, Eti Ben-Simon, Michal Gruberger, Adi Maron-Katz, Talma Hendler, Ramiro Chaparro-Vargas, and Dean Cvetkovic. Information flow and coherence of eeg during awake, meditation and drowsiness. *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5446–5449, 2014.
- [12] Alberto Fernández et al. Lempel-ziv complexity in schizophrenia: A meg study. *Clinical Neurophysiology*, 122(11):2227 – 2235, 2011.
- [13] Paul Expert, Renaud Lambiotte, Dante R. Chialvo, Kim Christensen, Henrik Jeldtoft Jensen, David J. Sharp, and Federico Turkheimer. Self-similar correlation function in brain resting-state functional magnetic resonance imaging. *Journal of The Royal Society Interface*, 8(57):472–479, 2011.
- [14] Nadine Gaab, Christian Gaser, Tino Zaehle, Lutz Jancke, and Gottfried Schlaug. Functional anatomy of pitch memory - an fmri study with sparse temporal sampling. *NeuroImage*, 19(4):1417 – 1426, 2003.
- [15] C. Gomez, R. Hornero, D. Abasolo, M. Lopez, and A. Fernandez. Decreased lempel-ziv complexity in alzheimer’s disease patients’ magnetoencephalograms. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pages 4514–4517, Jan 2005.
- [16] P. Grassberger. Estimating the information content of symbol sequences and efficient codes. *IEEE Transactions on Information Theory*, 35(3):669–675, May 1989.
- [17] P. D. Grunwald and P. M. B. Vitanyi. Algorithmic information theory. *ArXiv e-prints*, September 2008.
- [18] N. Hatsopoulos, Stuart Geman, Asohan Amarasingham, and Elie Bienenstock. At what time scale does the nervous system operate? *Neurocomputing*, 52-54:25–29, 2003.
- [19] S. Herzog, C. Tetzlaff, and F. Wörgötter. Transfer entropy-based feedback improves performance in artificial neural networks. *ArXiv e-prints*, June 2017.
- [20] A. Kolmogorov. Logical basis for information theory and probability theory. *IEEE Trans. Inf. Theor.*, 14(5):662–664, September 2006.

- [21] D. Kugiumtzis. Direct-coupling information measure from nonuniform embedding. *Physical Review E*, 87(6):062918, June 2013.
- [22] Edward W. Large, Jorge A. Herrera, and Marc J. Velasco. Neural networks for beat perception in musical rhythm. *Frontiers in Systems Neuroscience*, 9:159, 2015.
- [23] A. Lempel and J. Ziv. On the complexity of finite sequences. *IEEE Transactions on Information Theory*, 22(1):75–81, January 1976.
- [24] Ming Li and Paul M.B. Vitnyi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer Publishing Company, Incorporated, 3 edition, 2008.
- [25] I.G Meister and T Krings. Playing piano in the mind - an fmri study on music imagery and performance in pianists. *Cognitive Brain Research*, 19(3):219 – 228, 2004.
- [26] Robert A. Meyers. *Encyclopedia of Complexity and Systems Science - V. 1-10*. Springer Publishing Company, Incorporated, 1st edition, 2009.
- [27] Jukka-Pekka Onnela, Daniel J. Fenn, Stephen Reid, Mason A. Porter, Peter J. Mucha, Mark D. Fricker, and Nick S. Jones. Taxonomies of networks from community structure. *Phys. Rev. E*, 86:036104, Sep 2012.
- [28] Gaoxiang Ouyang, Chuangyin Dang, Douglas A. Richards, and Xiaoli Li. Ordinal pattern based similarity analysis for eeg recordings. *Clinical Neurophysiology*, 121(5):694 – 703, 2010.
- [29] Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, New York, NY, USA, 2nd edition, 2009.
- [30] Ana Luísa Pinho, Örjan de Manzano, Peter Fransson, Helene Eriksson, and Fredrik Ullén. Connecting to create: Expertise in musical improvisation is associated with increased functional connectivity between premotor and prefrontal areas. *Journal of Neuroscience*, 34(18):6156–6163, 2014.
- [31] Charles Rathkopf. Network representation and complex systems. *Synthese*, (1), 2018.

- [32] M. Rosvall, D. Axelsson, and C. T. Bergstrom. The map equation. *European Physical Journal Special Topics*, 178:13–23, November 2009.
- [33] Martin Rosvall and Carl T. Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123, 2008.
- [34] Jakob Runge. *Detecting and quantifying causality from time series of complex systems*. PhD thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät, 2014.
- [35] Thomas Schreiber. Measuring information transfer. *Phys. Rev. Lett.*, 85:461–464, Jul 2000.
- [36] C. E. Shannon. A mathematical theory of communication. *SIGMOBILE Mob. Comput. Commun. Rev.*, 5(1):3–55, January 2001.
- [37] Olaf Sporns. Networks analysis, complexity, and brain function. *Complex.*, 8(1):56–60, September 2002.
- [38] R. W. Thatcher, E. Palmero-Soler, D. M. North, and C. J. Biver. Intelligence and eeg measures of information flow: efficiency and homeostatic neuroplasticity. In *Scientific reports*, 2016.
- [39] A. A. Tsonis and P. J. Roebber. The architecture of the climate network. *Physica A Statistical Mechanics and its Applications*, 333:497–504, February 2004.
- [40] T. J. La Vaque. The history of the eeg and hans berger. *Journal of Neurotherapy*, 3(2):1–9, 1999.
- [41] K. Visweswariah, S. R. Kulkarni, and S. Verdu. Source codes as random number generators. *IEEE Transactions on Information Theory*, 44(2):462–471, March 1998.
- [42] Ioannis Vlachos and Dimitris Kugiumtzis. Nonuniform state-space reconstruction and coupling detection. *Phys. Rev. E*, 82:016207, Jul 2010.
- [43] Xiaogeng Wan, Björn Crüts, and Henrik Jeldtoft Jensen. The causal inference of cortical neural networks during music improvisations. *PLOS ONE*, 9(12):1–25, 12 2014.

- [44] Pengbo Yang, Pengjian Shang, and Aijing Lin. Financial time series analysis based on effective phase transfer entropy. *Physica A: Statistical Mechanics and its Applications*, 468:398 – 408, 2017.
- [45] J. Ziv and A. Lempel. Compression of individual sequences via variable-rate coding. *IEEE Transactions on Information Theory*, 24(5):530–536, September 1978.