# Determining Causality from Neurophysiologic Time Series Using Information Theory Based Measures

Arto Meriläinen

January 10, 2012

**Abstract**

Neurophysiologic time series are quasiperiodic: The time series consist of transient neural oscillations at varying frequencies. Although the oscillations operate on different frequency bands, the bands are functionally different. Therefore, directional metrics can be considered as a natural approach to research the activity on different frequency bands of the neurophysiologic data.

Synchronization of phase difference distribution has been used as an indicator of interaction between brain regions. Detecting the synchronization from EEG, LFP and MEG data can be done with several indices. However, these indices cannot reveal the direction of the information flow. In addition, these measures produce false-positive results in cases where the history of previous activity could reveal that the behavior should be considered natural.

This report extends the frameworks of predictive information and transfer entropy to directional metrics. The measures are tested using artificial simulated data in order to study their performance and sensitivity in different conditions. The simulation results show that the measures are able to distinguish the direction of information flow properly in noiseless conditions. Common artifacts have only minor effect to the results. Artificial volume conduction is noted to affect significantly to the results. However, with proper choice of parameters, transfer entropy can be considered reliable.

Surrogate data is often applied for determining the confidence intervals using the data itself. The results indicate that a commonly used surrogate method produces surrogate data which gives poor confidence interval estimates for phase locking value. This makes phase locking value based analyses vulnerable for false positive results. The results show that the surrogate method can be applied for predictive information and transfer entropy in order to determine the confidence intervals.

Common artifacts and artificial volume conduction had only minor effects to the sensitivities of predictive information and transfer entropy. The sensitivities of new methods are comparable to the sensitivity of phase locking value.

# Contents

# 1 Introduction

The active brain networks can be studied in a number of different approaches. Some approaches rely on studying the stability of the phase difference distribution between two brain regions of electrophysiological brain signals [1–4]. Providing that there is an interaction between the regions, the activation delay should be approximately constant. As networks usually are active relatively long periods, a network activation can be seen as the stability of the phase difference distribution. The brain signals are oscillatory - having both phase and amplitude for each time index - making the phase presentation natural. There is also a wide variety of methods that utilize the hemodynamics of the brain for finding brain networks [5].

Although phase difference synchronization provides a good basis for determining the interaction between brain regions, it has several drawbacks. Obviously, the approach cannot offer information about the direction of interaction. More problematic issue is the possibility to generate false positive results in cases, where two channels simply happens to produce similar phase difference distribution with a small time lag. The issue can be demonstrated with a simple watch example: If only the synchronization of phase difference distribution is considered, all watches are synchronized! In order to study the real interaction between the brain regions, the method for determining the interaction should observe also the history of both brain regions. Thus, the interaction should be studied through phase causality rather than phase difference synchronization.

In neuroinformatics, *The Granger causality test* [6] is a traditional method for determining the causal relations of time series. The test investigates whether information provided by one channel helps to predict data on another channel. Although the Granger causality test provides means for determining the causal relation, it is usually applied only in linear case, which makes analysis in nonlinear cases troublesome. This approach has been extended to directional metrics [7].

*Dynamic Causal Modeling* is often applied in hypothesis-led research where a hypothetical model is tested against the measured data [8]. The method is capable to show the incorrectness of the model, but incapable to show the correctness of the model. The concept of DCM is extended to explorative analyses, but it is demonstrated only using fMRI data [9].

Information theory [10] provides an interesting concept for studying the causality. *Predictive information* (PI) is defined as the *mutual information* (MI) between the past of the first time series and the present of the second time series [11]. In principle, predictive information studies the common features between the past of the first time series and the present of the second time series. Predictive information is not bound to linear analyses, but the measure detects also non-linear relationships in the data. Although the method can be useful, it does not observe the past of the target time series making the method vulnerable for false positive results. Predictive information is applied in causality analyses [12].

As in predictive information, also *transfer entropy* (TE) utilizes information theory in causality analyses [13]. Transfer entropy utilizes also the information about the past of target the time series and the method has been found to perform reliable for linear and non-linear data [14]. Simulations have shown that mixing do not affect the measure significantly in real valued case. Thereby, transfer entropy offers an interesting starting point for developing a causality measure in directional metrics.

Transfer entropy can be presented as *conditional mutual information* (CMI), which allows conditioning the variables in interest by other factors [15]. The basic concept is similar to mutual information: The measure allows studying the similarity between two data sets. In addition, CMI allows excluding effects of affecting variables. In the case of transfer entropy, the affecting variable is simply the past of the target time series. Similar method has been used in analysis of epileptogenic brain connectivity [16].

Although information theory provides a good basis for causality analyses, it has been applied only for real valued data. Brain signals are oscillatory, thus real valued data may not be the best approach for determining the causality. This report extends the concept of predictive information and transfer entropy into phase domain and performs several tests for the extended measures. These measures are compared to phase locking value [1], which is a common method in connectivity analyses. First, the report presents the basic concept of mutual information for time-lagged variables. Second, the concept is extended into the case where the mutual information can be given the information of other variables. Finally, The measure is demonstrated using artificial simulated data.

# 2 Methods

This chapter introduces the methods that are required to apply causality analyses in directional metrics. The analyses follow mainly the workflows shown in figure 1. The workflow can be divided into three parts: The first part (black route in the figure) shows the analysis of the interaction data. The data is filtered and the interaction is estimated using an interaction index. The second and third flows are related to the estimation of confidence intervals. The second flow (blue route in the figure) shows the surrogate data generation and confidence interval estimation. This route is used always when dealing with the real data.

The third route is available only in simulations. The third route estimates the confidence intervals directly from data that is known to have no interactions, which makes the estimates much more accurate compared to the estimates from the surrogate data. Despite this approach cannot be used for the real data, the approach is useful for determining the accuracy of confidence interval estimates.

This chapter is organized as follows: First, the simulation data generation and filtering steps are introduced. Second, the report introduces phase-locking value, which is applied often in connectivity analyses. Third, a histogram of the filtered data is created in order to refine the data. As the data is processed in directional metrics, this processing step is performed in an unusual manner. Fourth, the basic concepts of information theory, predictive information and transfer entropy are introduced. Finally, the report describes surrogate method, which is used to estimate the confidence intervals from the data.

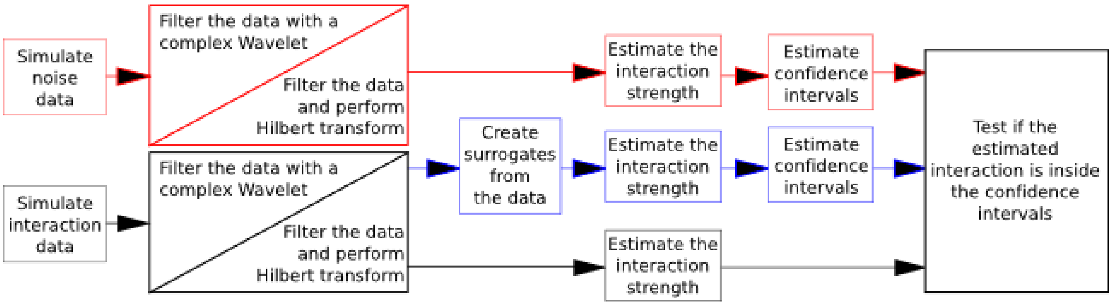



Figure 1: Workflow of analyses for simulated data. There are two different methods for estimating the confidence intervals: The red route uses simulated noise as the reference. The blue route creates the noise data using surrogate method and estimates the confidence intervals from the surrogate data.

## 2.1 Simulation Data Generation

Artificial simulated data is useful for several purposes. The simulated data can be used to verify the operation of the algorithm and the implementation. In addition, it can be used for comparing differences between different metrics.

The simulated data is generated by first creating artificial white noise for all channels in the simulation. The white noise on all channels is used as the initial data. The initial data is then mixed in order to create artificial links over some channels:

$$x^i_{interacted}(n) = x^i_{original}(n) + \sum_{j=1;j\neq i}^{K} \alpha_{j\to i} x^j(n - \tau_j), \tag{1}$$

where $i$ denotes the target channel, $j$ source channel, $K$ number of all channels and $\alpha_{j\to i}$ denotes the interaction strength and $\tau_j$ time lag.

The data is mixed in order to simulate the effect of volume conduction:

$$x^i_{vc}(n) = x^i_{interacted}(n) + \sum_{j=1;j\neq i}^{K} \alpha_j x^j(n), \tag{2}$$

where $\alpha_j$ denotes the constant mixing between the channels $i$ and $j$.

The physiological time series may have different contaminations (e.g. cardiac or muscular), which can be seen in all channels virtually simultaneously. This effect is simulated by adding same noise to all channels:

$$x^i(n) = x^i_{vc}(n) + \epsilon, \tag{3}$$

where $\epsilon$ denotes the additive noise.

## 2.2 Preprocessing

The data in each measurement channel is filtered with Morlet wavelet, which filters and transforms the data into time-complex -representation [17]. The complex representation can be further studied using directional metrics, where the signal consists of phase $\theta$ and amplitude $A$. However, information about the amplitude is not in interest, thus each sample in the data is normalized. The data is denoted simply as $\theta^{ch}(n)$, where $ch$ denotes the channel and $n$ the sample index of a single sample.

## 2.3 Phase Locking Value

Phase Locking Value (PLV) is a traditional measure for the stability of phase difference distribution. The measure is defined as:

$$PLV = \frac{|\sum_{n=0}^{N} e^{i(\theta^i(n) - \theta^j(n))}|}{N}. \tag{4}$$

where $\theta^i(n)$ denotes the phase of the filtered signal on channel $i$ at sample index $n$. Similarly, $\theta^j(n)$ denotes the phase of the filtered signal on channel $j$ at sample index $n$. [1]

The explanation of the definition is quite simple: If the channels are not phase locked, the phase difference distribution is random. Consequently, the length of the sum vector is small. If the channels are phase locked, the length of the sum vector will be near $N$ making the value of PLV near 1.

## 2.4 Histogram Estimation

The distribution of the directional data is estimated using histogram estimation. In histogram estimation the interval from data minimum to data maximum is divided into certain number of bins each of which represent simply a class of the data. Computationally this method is efficient and simple to implement. However, the number of bins affects much to the shape of probability distribution: the real distribution disappears if the number of bins is too small. On the other hand, if the number of bins is too large, each bin will be occupied by only few samples, which can lead to wrong conclusions.

In the directional case, the interval $[-\pi\ \pi]$ is divided into bins. Scott's choice [18, Chapter 3.2] is utilized for determining the length of a single bin:

$$h^{ch} = \frac{3.5\sigma^{ch}}{N^{\frac{1}{3}}}, \tag{5}$$

where $h^{ch}$ denotes the length of a bin for channel $ch$, $N$ the number of samples and $\sigma^{ch}$ the standard deviation of the samples. The standard deviation for a directional variable is defined as [19, Chapter 2.3]:

$$\sigma^{ch} = \sqrt{-2\ln(\overline{R^{ch}})} = \sqrt{-2\ln(|\frac{1}{N}\sum_{t=1}^{N} e^{i*\theta^{ch}(n)}|)}, \tag{6}$$

where $\overline{R^{ch}}$ denotes the sample mean resultant vector. When the bin length $h^{ch}$ is available, it is straightforward to determine the number of bins:

$$k^{ch} = \frac{2\pi}{h^{ch}}. \tag{7}$$

After the number of bins for each channel is determined, the data is simply categorized into different bins. The probability of group $i$ on channel $ch$ is denoted simply by $p(\theta_i^{ch})$. This categorization will be later used when the joint probability of simultaneous activation between channels is considered.

## 2.5 Information Theory in Directional Metrics

Information theory offers an interesting approach for determining (dis)similarity of variables and data sets. Throughout this report the information theory is utilized in phase domain. In addition, this report introduces the notation that is noted to be proper.

This chapter is organized as follows: First, the basic concepts in information theory are reviewed starting from the definition of information up to conditional mutual information. These concepts are then utilized in *predictive information* and *transfer entropy*.

### 2.5.1 Definition of Information

Previously is shown the categorization of the data using histogram estimation. Let us consider the information content of a single category. If samples from a category are present frequently, observing a sample belonging to the category does not carry much information: observing the sample is nothing special. On the other hand, if the category is present extremely rarely, samples belonging to the category carry much information. This gives the basis for the information content in the information theory. Formally, the information content is defined as [10]:

$$I(\theta_i^{ch}) = -\log(p(\theta_i^{ch})). \tag{8}$$

### 2.5.2 The Shannon Entropy

*The Shannon entropy* describes the expected information content of all categories [10]:

$$H(\theta^{ch}) = E(I(\theta^{ch})) = -\sum_{g=1}^{G} p(\theta_g^{ch}) \log p(\theta_g^{ch}), \tag{9}$$

where $G$ denotes the number of possible categories. Thus, the entropy explains the likeliness to observe an unlikely sample. It should be emphasized that the entropy is not a probability as itself: entropy is not bound to interval [0 1]. The entropy can be extended to bivariate case:

$$H(\theta^x, \theta^y) = E(I(\theta^x \cup \theta^y)) = -\sum_{g=1}^{G}\sum_{f=1}^{F} p(\theta_g^x, \theta_f^y) \log p(\theta_g^x, \theta_f^y), \tag{10}$$

where $p(\theta_g^x, \theta_f^y)$ denotes the joint probability over groups $g$ on channel $x$ and $f$ on channel $y$. In other words, $p(\theta_g^x, \theta_f^y)$ refers to the probability to observe data belonging to group $g$ at the same time as data belonging to group $f$ is observed on the second channel. The entropy can be extended to multivariate case in a similar fashion.

#### 2.5.3 Mutual Information

Let us now consider a case where the probability $p(\theta_f^y) = p(\theta_g^x)$ and $p(\theta_f^y|\theta_g^x)$ equals to 1. It is trivial to show that $H(\theta_g^x)$ and $H(\theta_f^y)$ are identical. In addition, it is trivial to show that $H(\theta^x, \theta^y)$ is also identical to the entropies. Thus, there is a clear connection between the probability distributions and the joint entropy. The connection is called *mutual information* (MI) and it is defined as the amount of information that is dependent between variables $\theta^x$ and $\theta^y$ [10]:

$$\begin{aligned} I(\theta^x \cap \theta^y) \quad &= E(\log(p(\theta^x, \theta^y))) - E(\log(p(\theta^x)p(\theta^y))) \\ &= E(\log(\frac{p(\theta^x, \theta^y))}{p(\theta^x)p(\theta^y)}). \end{aligned} \tag{11}$$

The mutual information can be represented by entropies thereby making the measure computationally convenient:

$$\begin{aligned} I(\theta^x \cap \theta^y) \quad &= E(\log(p(\theta^x, \theta^y))) - E(\log(p(\theta^x)p(\theta^y))) \\ &= E(\log(p(\theta^x, \theta^y))) - E(\log(p(\theta^x))) - E(\log(p(\theta^y))) \\ &= H(\theta^x) + H(\theta^y) - H(\theta^x, \theta^y). \end{aligned} \tag{12}$$

If the variables $\theta^x$ and $\theta^y$ are independent, the mutual information equals to 0. On the other hand, if there is a large dependency between the variables, the conditional entropy $H(\theta^x|\theta^y)$ is small and the mutual information $I(\theta^x \cap \theta^y)$ equals to $I(\theta^x)$. Thus, the mutual information can be used to measure the relationship between variables $\theta^x$ and $\theta^y$. The relation between entropy and mutual information is illustrated in figure 2a. [10]

#### 2.5.4 Conditional Mutual Information

*Conditional mutual information* is defined as the mutual information given the variable $\theta^z$ [15]:

$$\begin{aligned} I(\theta^x \cap \theta^y|\theta^z) \quad &= E(\log(p(\theta^x, \theta^y|\theta^z))) - E(\log(p(\theta^x|\theta^z)p(\theta^y|\ \theta^z))) \\ &= E(\log(\frac{p(\theta^x, \theta^y)|\theta^z)}{p(\theta^x|\theta^z)p(\theta^y|\theta^z)}). \end{aligned} \tag{13}$$

The conditional mutual information can be represented by entropies similarly as the mutual information:

$$\begin{aligned} I(\theta^x \cap \theta^y|\theta^z) \quad &= E(\log(p(\theta^x, \theta^y|\theta^z))) - E(\log(p(\theta^x|\theta^z)p(\theta^y|\ \theta^z))) \\ &= E(\log(p(\theta^x, \theta^y|\theta^z))) - E(\log(p(\theta^x|\theta^z))) - E(\log(p(\theta^y|\ \theta^z))) \\ &= E(\log(\frac{p(\theta^x, \theta^y, \theta^z))}{p(\theta^z)}) - E(\frac{\log(p(\theta^x, \theta^z))}{p(\theta^z)}) - E(\frac{\log(p(\theta^y,\ \theta^z))}{p(\theta^z)}) \\ &= E(\log(p(\theta^x, \theta^y, \theta^z))) - E(\log(p(\theta^x, \theta^z))) - E(\log(p(\theta^y,\ \theta^z))) \\ &+E(\log(p(\theta^z))) \\ &= H(\theta^x,\ \theta^z) + H(\theta^y,\ \theta^z) - H(\theta^z) - H(\theta^x, \theta^y,\ \theta^z) \end{aligned} \tag{14}$$

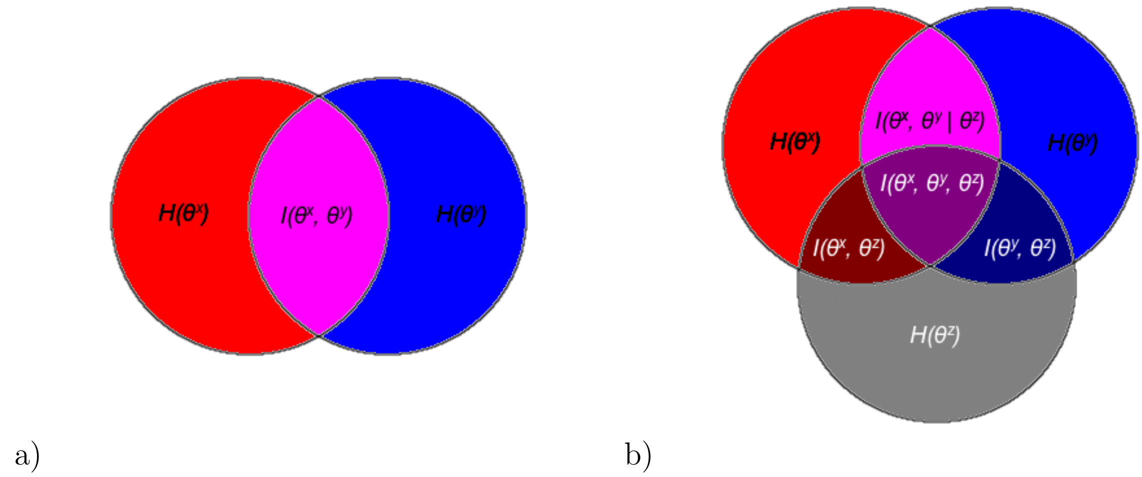



Figure 2: The relation between entropy, mutual information and conditional mutual information. a) shows the relation between entropy and mutual information. $I(\theta^x, \theta^y)$ can be considered as the similarity between variables $\theta^x$ and $\theta^y$. b) shows the relation between variables $\theta^x$ and $\theta^y$ given knowledge of $\theta^z$. $I(\theta^x, \theta^y | \ \theta^z)$ includes mutual information of variables $\theta^x$ and $\theta^y$ by given the variable $\theta^z$.

If the variable $\theta^x$ or $\theta^y$ is dependent on variable $\theta^z$, the effect of variable $\theta^z$ will not be considered in the conditional mutual information. Similarly, if the variables $\theta^x$ and $\theta^y$ are independent of $\theta^z$, the conditional mutual information will reduce into mutual information. The relation between entropy, mutual information and conditional mutual information is illustrated in figure 2b.

### 2.5.5 Predictive Information

This far we have discussed about measuring the dependencies between channels using mutual information. This idea can be extended to *predictive information* by calculating the mutual information between one time series and the past of another time series [11]. The predictive information is defined as measure $I(\theta^x \cap \theta^{y'})$, where the values on the channel $y$ are time lagged. This can be refined even more informative similarity index by utilizing predictive information to both directions:

$$SI_{PI} = I(\theta^x \cap \theta^{y'}) - I(\theta^y \cap \theta^{x'}). \tag{15}$$

The similarity index thereby will be either positive or negative depending on the direction of information flow.

Although the predictive information is able to find correlations between two time series, the information about the past of the first time series is not utilized. Thus, the positive result does not guarantee the causality.

### 2.5.6 Transfer Entropy

*Transfer entropy* utilizes the past of both time series and the present of the target time series in order to determine the causality between the two time series. The measure is defined as [13]:

$$I_{Y \to X} = E(\log(\frac{p(\theta^x | \theta^{x'}, \theta^{y'})}{p(\theta^x | \theta^{x'})})). \tag{16}$$

Clearly, the transfer entropy describes the relative amount of information that the past of channel $y$ "adds" to the prediction. As an interesting feature, transfer entropy can be represented as a conditional mutual information:

$$\begin{aligned} I_{Y \to X} &= E(\log(\frac{p(\theta^x | \theta^{x'} \theta^{y'})}{p(\theta^x | \theta^{x'})})) \\ &= E(\log(\frac{p(\theta^x, \theta^{x'}, \theta^{y'}) p(\theta^{x'})}{p(\theta^x, \theta^{x'}) p(\theta^{x'}, \theta^{y'})})) \\ &= E(\log(\frac{p(\theta^x, \theta^{x'}, \theta^{y'}) p(\theta^{x'})}{p(\theta^x, \theta^{x'}) p(\theta^{y'}, \theta^{x'})})) \\ &= I(\theta^x \cap \theta^{y'} | \theta^{x'}), \end{aligned} \tag{17}$$

where values of $y'$ and $x'$ are time lagged values of $x$ and $y$.

In other words, transfer entropy describes the mutual information between the past of the first time series and the present of the second time series given the past of the second time series. The similarity index for transfer entropy can be defined as:

$$SI_{TE} = I(\theta^x \cap \theta^{y'} | \theta^{x'}) - I(\theta^y \cap \theta^{x'} | \theta^{y'}). \tag{18}$$

The notation is practical as the measure can be extended by adding condition variables. This feature becomes an useful property when dealing with physiological time series: The effect of disturbing factors (e.g. volume conduction, eye blinks) can be suppressed by adding a conditioning variable representing the disturbing factor. This approach has been applied for real valued data in [11] and [16].

## 2.6 Estimating Confidence Intervals with Surrogate Data

The surrogate data is useful for determining confidence intervals from the data. Each channel in the surrogate data has the same autocorrelation function as the corresponding real data, but the correlations over channels are removed. In other words, the surrogate data can be considered as "good noise" in connectivity analyses. Everything that a measure gives for surrogate data can be considered natural. Thus, the surrogate data can be used to determine the confidence intervals for the noise. If the method produces value that is outside the confidence intervals, the result can be considered statistically significant. [20]

The surrogate data can be generated by choosing random data fragments from each channel [1, 20]. This operation removes correlations between channels while preserving the autocorrelation within each channel. However, the method do not remove only the real interactions between the channels, but also the effect of volume conduction and common artifacts. Therefore, the surrogate data is free of volume conduction and common artifacts, which affects to the confidence interval estimates.

### 2.6.1 Specificity of Surrogate Method

Although the surrogate method may give wrong estimates for the confidence intervals, the goodness of the surrogate method for the measure in interest can be estimated with simulations. Let us consider simulated data with artificial volume conduction and common artifacts, but with no interactions. This data can be used to determine the real confidence intervals for the data. The surrogate method can be applied for the noise data in order to estimate the confidence intervals that the surrogate data would produce.

The goodness of surrogate method for the measure in interest can be computed by creating several surrogates of the noise data. The surrogates are used to estimate of confidence intervals for the data. The original data should always belong to the confidence interval, because no interactions are present in the data. The *specificity* is defined as the percentage of the measure values that belonged to the estimated confidence interval. If the specificity is low, the surrogate method produces data, which gives poor confidence interval estimates.

### 2.6.2 Estimating Sensitivity with Surrogate Data

Simulations and surrogate data can be utilized for estimating the sensitivity of the measure in interest. However, it is equally important to validate that the method can produce statistically significant results.

Let us consider simulated data with interactions. The confidence intervals can be estimated from the data using surrogate method. Because we already know that the simulated data has interactions, the measure should always differ significantly from the surrogate data. The sensitivity is defined as the percentage of detected interactions.

# 3 Results

The data analysis was performed using both predictive information and transfer entropy. Simulations and analyses were performed in different conditions each of which had different parameter values for interaction lag and mixing strength. Each simulation consists of 5000 samples. The results presented at this chapter have been averaged over 100 trials. The simulated data was Morlet filtered ($f_0 = 10Hz$, $m = 5$, $f_s = 100Hz$).

## 3.1 Determining the Causality

In order to determine the significance of a result, the method was applied for simulated noise data having the same noise and mixing, but no real interactions between channels. The 95% confidence interval from the simulated noise data is used for determining if a result is statistically significant. It should be noted that this method is not identical to surrogate data [1, 20] as the volume conduction is not present in the surrogate data. Thus, simulated white noise should give better estimate for the significance level.

Figure 3 shows the measure response as a function of coupling/mixing strength for predictive information (figure 3a), transfer entropy (3b) and phase-locking value (3c) when interaction and analysis lags have value 8. The negative values denote interaction $X \rightarrow Y$ and positive values interaction $Y \rightarrow X$. Both analyses were performed without artificial volume conduction or noise. The red horizontal bars show the 95% significance limits. The yellow vertical bars show points where the measure crosses the 95% significance. The size of the region between the lower and upper boundary is used as the goodness measure in other figures; the smaller the region is, the better the measure is to distinguish the direction.

The effects of using different analysis and interaction lags to the size of the insignificant region are shown in figures 4 and 5. Figures 4a and 4b show the performance of transfer entropy and predictive information in noiseless conditions. The response of phase locking value for different interaction lags in noiseless conditions is shown in figure 5. The figures show that phase locking value, predictive information and transfer entropy perform adequately in noiseless conditions.

## 3.2 Influence of Additive Noise and Mixing

The effect of applying additive noise of 0.05 is shown in figures 4c (predictive information), 4d (transfer entropy) and 5b (phase locking value). The figures show the response as a function of interaction lag and analysis lag. The results show that additive noise has only minor effect to the measures.

Figures 4e (predictive information), 4f (transfer entropy) and 5c (phase locking value) show the effect of volume conduction to the measures when the mixing coefficient has value 0.3. Clearly, the mixing affects all measures: If the interaction lag is about 15 samples and the analysis lag is small, the measures cannot reveal the true interaction. It should be emphasized that these observed values did not include expected significant results.

Figure 6 shows the size of the insignificant region as a function of volume conduction while using two different interaction lag values (10 and 15). The analysis lag had value of 10 in all simulations. The figure shows that the effect of volume conduction is highly dependent on the interaction and analysis lags. If interaction lag is 10, the measures tolerate volume conduction adequately. However, if the interaction lag is 15, all tested measures perform poorly.

Predictive information and causal information perform in a similar fashion in almost all conditions. Strong interactions from $X \rightarrow Y$ are seen as negative responses in similarity index and strong interactions $Y \rightarrow X$ as high positive responses. If the difference between analysis lag and interaction lag is small, the response is stronger compared to situation where the analysis lag differs much from the interaction lag. Volume conduction lowers the response that is seen in similarity index.

## 3.3 Specificity analyses

This far the analyses are performed using known confidence intervals. Confidence intervals are never available for the real data, thus the intervals must be estimated from the data itself. However, the estimates may be incorrect making the measure in interest vulnerable for false positive results. Therefore, the accuracy of surrogate method must be validated with simulated data.

Specificity of each measure was computed in different conditions. The estimated 95% confidence intervals were computed using 100 surrogates from the noise data. The analysis lag had value of 8 whereas no interactions were present during specificity tests. Figure 7 shows the specificity for predictive information, transfer entropy and phase locking value. Figures show that the confidence interval is estimated correctly for phase locking value when additive noise is present, but the confidence intervals are estimated incorrectly when even little mixing is present. Predictive information shows good specificity when additive noise is present and the specificity is not affected significantly by volume conduction. Transfer entropy has good specificity in all conditions: additive noise and mixing affect only little to the measure.

### 3.4 Sensitivity analyses

Specificity shows the goodness to estimate valid confidence intervals from the data and good specificity is a prerequisite for successful analysis. It is equally important to validate that each measure actually can detect an interaction from the data. The sensitivity to interactions were researched by running several simulations. The analyses related to sensitivity were run using analysis lag 8 and interaction lag 8. The 95% confidence intervals were estimated using 100 surrogates from the interaction data.

Figures 8a and 8b show the sensitivity in noiseless conditions for predictive information and transfer entropy, respectively. Figure 9a shows the sensitivity of phase locking value when neither additive noise nor mixing is present. The figures show the sensitivity at different couplings between the channels. Ideally the sensitivity should be 1 with all coupling strengths, which would imply 100% detection given the surrogate that is generated from the interaction data. However, the specificity usually drops when the interaction strength between the channels is near zero.

The effect of additive noise is shown in figures 8c (predictive information), 8d (transfer entropy) and 9b (phase locking value). The additive noise has only little effect to all measures. Figures 8e (predictive information), 8f (transfer entropy) and 9c (phase locking value) show the effect of mixing. Interestingly, mixing seem to even improve the sensitivity of transfer entropy. Mixing does not have large effect to transfer entropy. Phase locking value is virtually 100% sensitive, but it should be noted that the specificity of confidence interval is extremely low for phase locking value at 0.3 mixing.

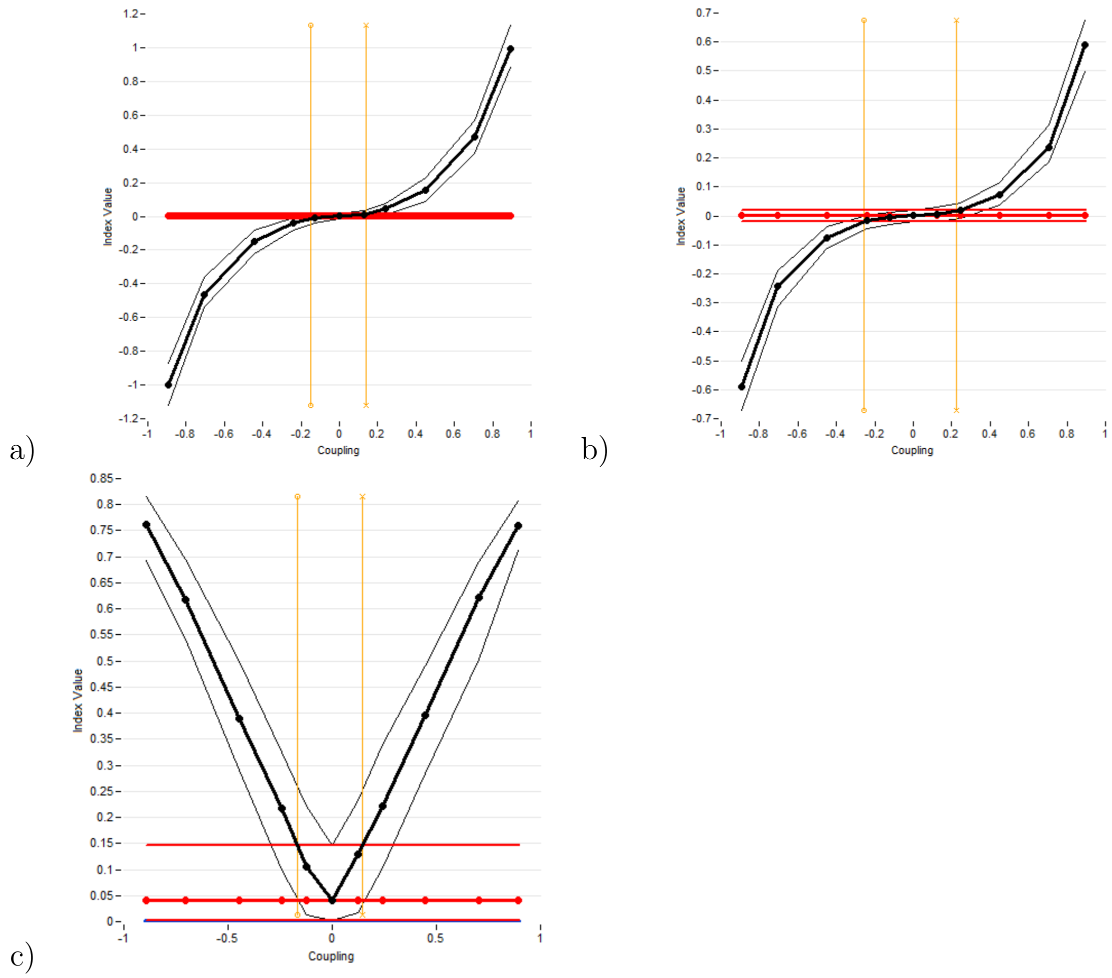



Figure 3: The response of a) predictive information b) transfer entropy c) phase locking value as a function of interaction strength. Negative strength refers to interaction $X \rightarrow Y$ and positive strength to interaction $Y \rightarrow X$. The analyses were performed using interaction lag 8 and analysis lag 8. Noise or mixing was not present in the simulations.

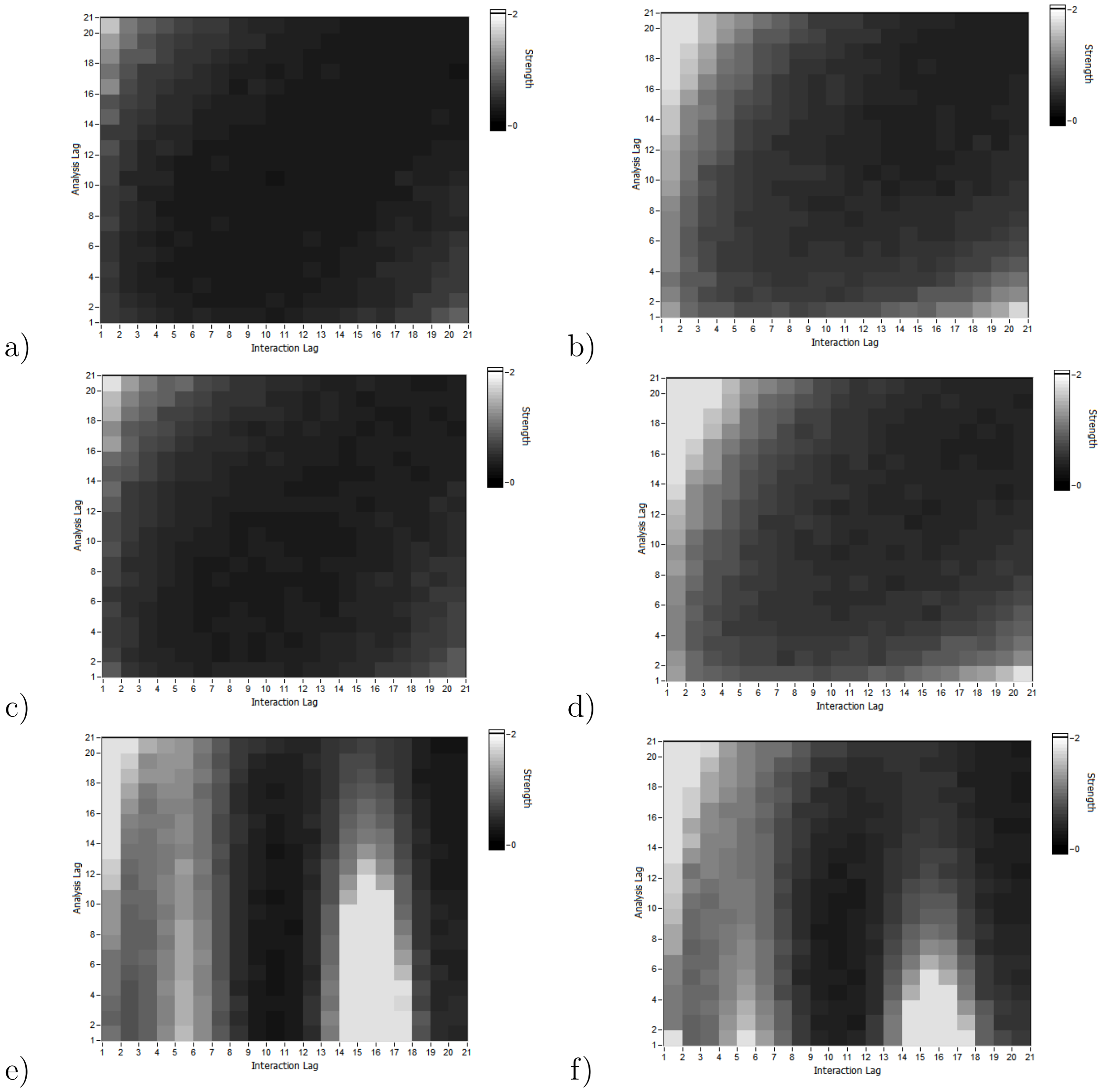
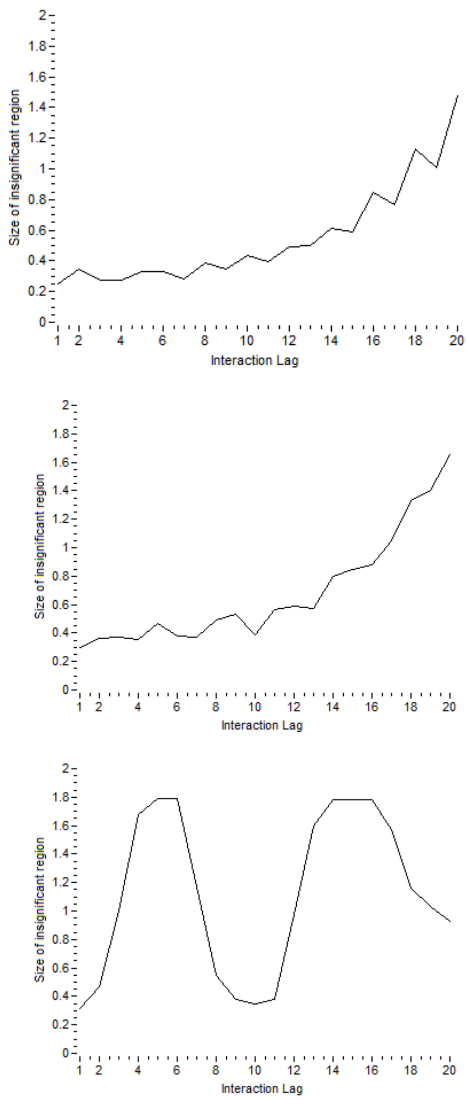



Figure 4: The goodness of predictive information and transfer entropy measures as functions of interaction and analysis lags. The goodness is measured as the difference in thte coupling that is required for producing a significant result. The lower (darker) result is better. The simulation result without volume conduction or noise for predictive information is presented in a) and for transfer entropy in b). Figures c) and d) show the effect of additive noise (amplitude 0.05) for predictive information and transfer entropy, respectively. Figures e) and f) show the effect of mixing to predictive information and transfer entropy, respectively. Mixing coefficient was set to 0.3.




a)

b)

c)

Figure 5: The goodness of phase locking value as a function of interaction lag. The goodness is measured as the difference in the coupling that is required for producing a significant result. The lower result is better. The simulation result without volume conduction or noise for is presented in a). Figure b) shows the same diagram when 0.05 additive noise is present. Figure c) shows the effect of mixing to phase locking value. Mixing coefficient was set to 0.3.

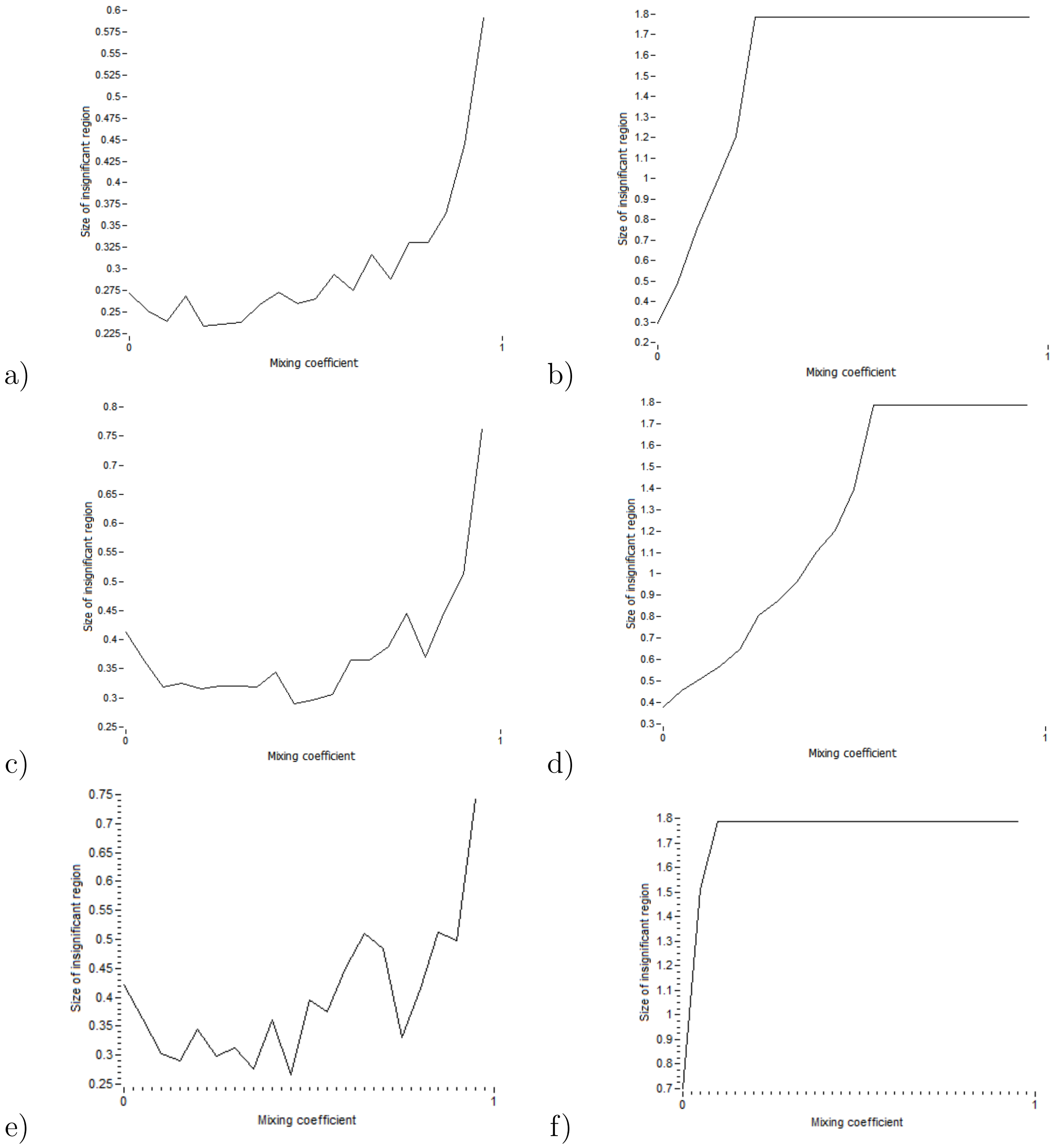



Figure 6: The size of the insignificant region as a function of mixing coefficient. a) and b) show the behavior of predictive information when the analysis lag is 10 and the interaction lag is 10 (a) or 15 (b), respectively. c, d, e and f show the behavior of transfer entropy (c, d) and phase locking value (e, f) in a similar fashion. Each figure shows the average over 100 simulations.

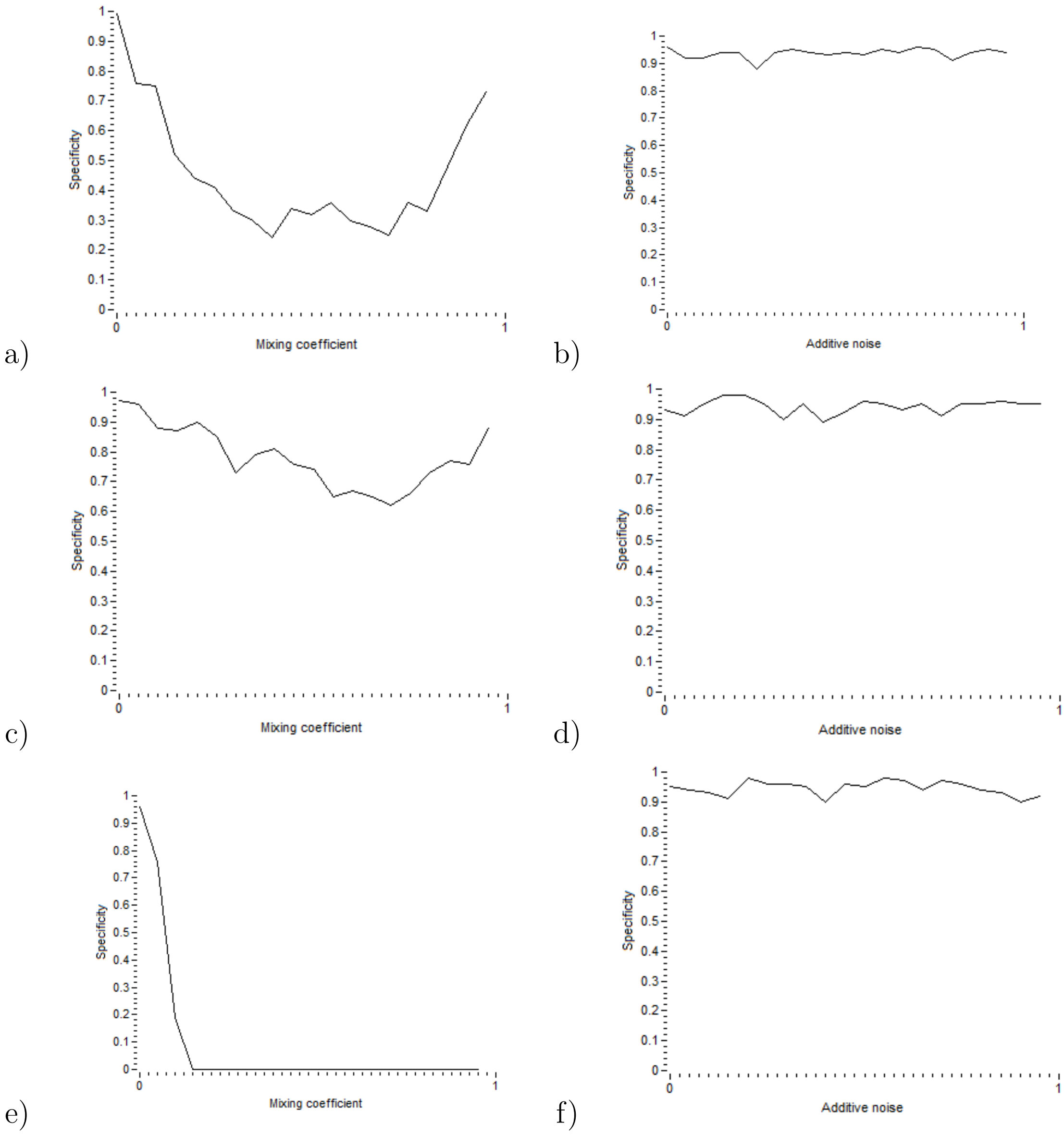



Figure 7: The specificity of the measures. Figures a) and b) show the effect of artificial volume conduction (a) and additive noise (b) to predictive information. Figures c, d, e, f show the same tests for transfer entropy (c, d) and phase locking value (e, f).

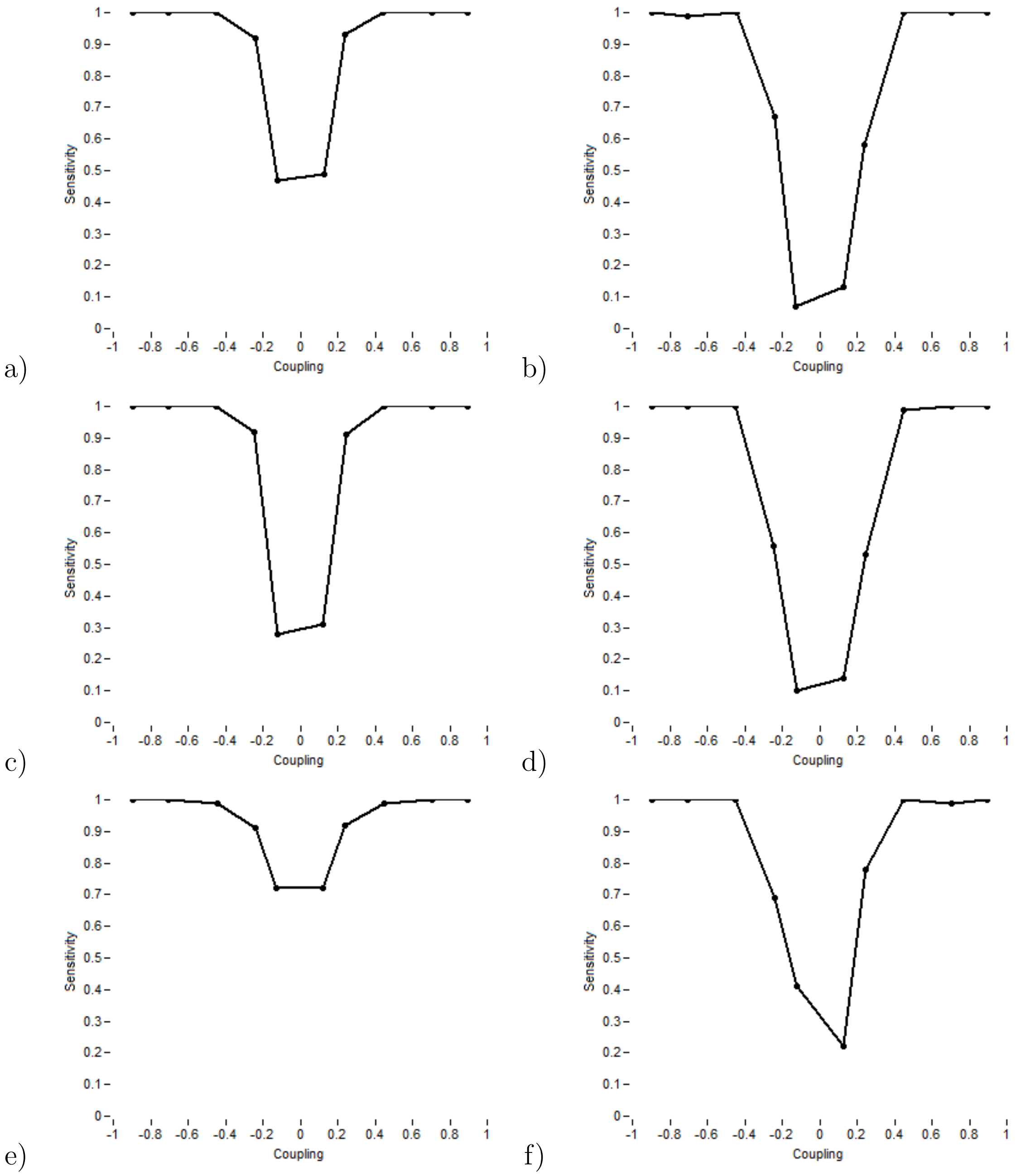



Figure 8: Sensitivity of predictive information and transfer entropy in different conditions. Figures a, c, e show behavior of predictive information when no noise (a), additive noise (b) and mixing between channels (c) is present. Mixing coefficient was set to 0.3 and additive noise had amplitude 0.05. Interaction and analysis lags were set to 8.

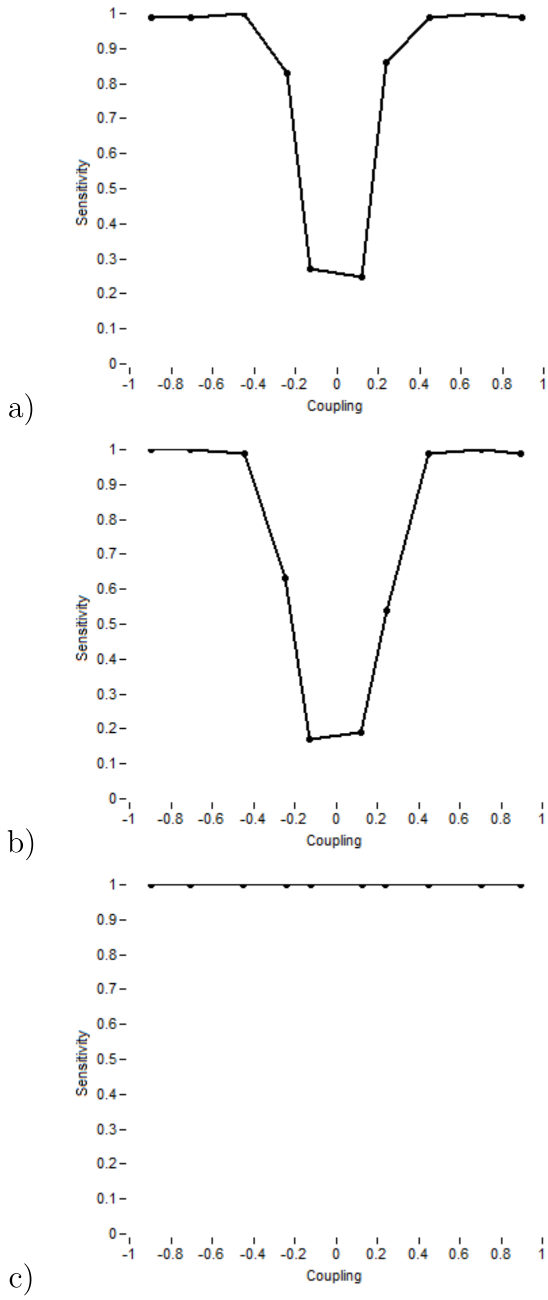



Figure 9: Sensitivity of phase locking value. Figure a shows the sensitivity when no noise is present. The effect of additive noise to the sensitivity is shown in figure b. Figure c shows the effect of additive noise. Mixing coefficient was set to 0.3 and additive noise had amplitude 0.05. Interaction and analysis lags were set to 8.

# 4 Discussion

Active brain networks can be discovered with various different methods. This report described directional metrics extensions for two information theory based measures and tested their performance using simulated data. In addition, the same performance analyses were performed for phase locking value, which is commonly applied while analyzing the connectivity.

## 4.1 Detecting the Interaction

In noiseless conditions predictive information, transfer entropy and phase locking value perform adequately: The measures were able to distinguish the direction or connectivity in all cases. In addition, all measures showed good resistance against additive noise. According to the results the sensitivity of predictive information is better than the sensitivity of transfer entropy. Interestingly, both measures performed adequately even when analysis lag differed much from interaction lag. Phase locking value performance was excellent when neither noise nor volume conduction was present.

Despite the predictive information and phase locking value performed reliable in the noiseless simulations, the measures cannot be considered reliable. The measures do not utilize the history information of the target channel. Thereby, both phase locking value and predictive information are vulnerable to wrong positive results. Thus, transfer entropy should be considered superior to predictive information and phase locking value and usage of transfer entropy should be preferred in real analysis.

## 4.2 Behavior in Noisy Conditions

Although the results show that all measures performed correctly in noiseless conditions, relatively small mixing, 0.3, affected significantly to the test results. All measures showed issues when the analysis lag was small and the interaction lag was about 15 samples. Phase locking value did not produce significantly high values and both directional interaction measures failed to find expected direction from the observed interaction strength values. Thus, the directional measures should be considered unpredictable and unreliable when using small interaction lag. The literature show that transfer entropy for real valued data has good tolerance against linear mixing [14], thus the issue may be related to data preprocessing.

Although the volume conduction issue definitely is notable, it can be explained by the Morlet filter: The issue is present when the interaction lag is half of the Morlet filter $f_0$. At that particular point, the signaling sample has 180-degree difference to the target sample. As these samples are summed in simulation data generation, the phase difference is reduced. Therefore, the issue can be explained as the property of the filter.

## 4.3 Estimating the Confidence Intervals

The previous tests were performed using artificial simulated data with known confidence intervals. However, the intervals are unknown when dealing with the real data and the confidence intervals must be estimated from the data itself. Usually the confidence intervals are estimated using surrogate data [1, 20], which can be considered as a good noise in connectivity analyses. The results show that the confidence intervals can be estimated from surrogate data for all measures when no mixing or additive noise is present. In addition, the additive noise had only minor effect to the specificity of confidence intervals. However, confidence intervals were estimated incorrectly for phase locking value when even small mixing was present. Predictive information had slightly better specificity compared to phase locking value. The specificity was always higher for transfer entropy compared to predictive information and phase locking value.

## 4.4 Sensitivity of the Measures

Sensitivity of investigated measures are adequate. Phase locking value performs adequately in noiseless conditions and when additive noise is present. However, the phase locking value has virtually 100% sensitivity when 0.3 mixing is present, which can be explained by the bad specificity of the confidence intervals of phase locking value. Compared to phase locking value, the sensitivity of transfer entropy is a little lower in all conditions. Predictive information has a good sensitivity in all conditions.

## 4.5 Future Development

The future development of the measure consists of several different parts. First, the entropy should be estimated by utilizing more reliable methods. Currently, the entropy is estimated with a simple histogram estimate, which is very sensitive to the selected number of bins. Thus, the histogram estimate is not accurate, but indicative. This issue could be overcome by estimating parameters for some circular distribution, which has a closed form solution for entropy. Generalized von Mises distribution [21] gives very promising starting point. Another solution could be estimating the entropy using some another entropy estimate.

Second, the filtering step causes inconveniences to the structure of the data. The issue could be solved by using some other filter, which would provide the phase information. Another possibility would be usage of Hilbert transform.

Third, the full potential of the conditional mutual information was not utilized. As noted before, this feature can be utilized by adding other condition variables. In theory, the effect of volume conduction could be reduced by conditioning the mutual information by the present of the source time series. This idea has been applied for real valued data [11,16].

Fourth, the indirect influence was not considered in the report. In the real data, there are hundreds of channels instead of two. If causal relations $X \rightarrow Y$ and $Y \rightarrow Z$ exist, transfer entropy will not detect only causal relations $X \rightarrow Y$ and $Y \rightarrow Z$, but also an indirect relation $X \rightarrow Z$! In theory, this issue can be solved by conditioning the mutual information by the history of some third channel. Computationally it is not feasible to condition the mutual information by all other channels.

Finally, the study of network hierarchies was excluded from this work. The developed index can be used to create a distance matrix between brain regions, which is valuable information alone. However, in order to analyze the networks in the large scale, different network algorithms could be used to find hubs and other interesting features from the distance data.

# 5 Conclusions

This report extended predictive information and transfer entropy to directional metrics. The results indicate that the measures perform adequately for simulated data. The measures were able to find causal relationships in noiseless conditions. The presence of linear mixing and additive noise affect the measures, but the simulations show that the measures tolerated the noise much better compared to phase locking value. The results show that the confidence intervals can be estimated using surrogate data for predictive information and transfer entropy.

Although both information theory based measures performed adequately, the usage of transfer entropy should be prefered due to theoretical limitations of predictive information. The predictive information does not utilize the information about the past of both time series making it vulnerable for false positive results in situations where the phase difference exists without phase causality.

Simulation results indicate that surrogate method should not be used for estimating confidence intervals in phase locking value based analyses. Even small mixing affects to the specificity of confidence interval estimates, which makes false positive results likely.

# References


[1] Jean-Philippe Lachaux Eugenio, Eugenio Rodriguez, Jacques Martinerie, and Francisco J. Varela. Measuring phase synchrony in brain signals. *Hum. Brain Mapp*, 8:194–208, 1999.

[2] Guido Nolte, Ou Bai, Lewis Wheaton, Zoltan Mari, Sherry Vorbach, and Mark Hallett. Identifying true brain interaction from eeg data using the imaginary part of coherency. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 115(10):2292–2307, 10 2004.

[3] Cornelis J. Stam, Guido Nolte, and Andreas Daffertshofer. Phase lag index: Assessment of functional connectivity from multi channel eeg and meg with diminished bias from common sources. *Human Brain Mapping*, 28(11):1178–1193, 2007.

[4] Vinck M, Oostenveld R, van Wingerden M, Battaglia F, and Pennartz CM. An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *Neuroimage*, 01 2011.

[5] Stephen M. Smith, Karla L. Miller, Gholamreza Salimi-Khorshidi, Matthew Webster, Christian F. Beckmann, Thomas E. Nichols, Joseph D. Ramsey, and Mark W. Woolrich. Network modelling methods for fmri. *NeuroImage*, 54(2):875 – 891, 2011.

[6] C. W. J. Granger. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3):424–438, 1969.

[7] Mukeshwar Dhamala, Govindan Rangarajan, and Mingzhou Ding. Analyzing information flow in brain networks with nonparametric granger causality. *NeuroImage*, 41(2):354 – 362, 2008.

[8] K.J. Friston, L. Harrison, and W. Penny. Dynamic causal modelling. *NeuroImage*, 19(4):1273 – 1302, 2003.

[9] Karl J. Friston, Baojuan Li, Jean Daunizeau, and Klaas E. Stephan. Network discovery with DCM. *NeuroImage*, December 2010.

[10] C. E. Shannon. A mathematical theory of communication. *Bell system technical journal*, 27, 1948.

[11] M.E. Pflieger and R. E. Greenblatt. Using conditional mutual information to approximate causality for multivariate physiological time. *Int J Bioelectromagnetism*, 7(2):285–288, 2005.

[12] Hermann Hinrichs, Toemme Noesselt, and Hans-Jochen Heinze. Directed information flow: a model free measure to analyze causal interactions in event related eeg-meg-experiments. *Human Brain Mapping*, 29(2):193–206, 2008.

[13] Thomas Schreiber. Measuring information transfer. *Phys. Rev. Lett.*, 85:461–464, Jul 2000.

[14] Raul Vicente, Michael Wibral, Michael Lindner, and Gordon Pipa. Transfer entropy–a model-free measure of effective connectivity for the neurosciences. *J. Comput. Neurosci.*, 30:45–67, February 2011.

[15] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory 2nd Edition (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, July 2006.

[16] M.E. Pflieger and A. A. Bassam. A noninvasive method for analysis of epileptogenic brain connectivity. *Epilepsia*, 45(Suppl: 7):70–71, 2004.

[17] P. Goupillaud, A. Grossmann, and J. Morlet. Cycle-octave and related transforms in seismic signal analysis. *Geoexploration*, 23(1):85 – 102, 1984. Seismic Signal Analysis and Discrimination III.

[18] David W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization (Wiley Series in Probability and Statistics)*. Wiley, September 1992.

[19] N. I. Fisher. *Statistical Analysis of Circular Data*. Cambridge University Press, January 1996.

[20] J Theiler, S Eubank, A Longtin, B Galdrikian, and J Doynefarmer. Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena*, 58(1-4):77–94, 1992.

[21] R Gatto and S Jammalamadaka. The generalized von mises distribution. *Statistical Methodology*, 4(3):341–353, 2007.