ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES

MSc THESIS

Erçin ÖZCAN

A COMPARATIVE STUDY OF DISTANCE/DISSIMILARITY MEASURES USED FOR CLASSIFICATION OF ELECTROENCEPHALOGRAM SIGNALS

DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

ADANA, 2015

ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES

A COMPARATIVE STUDY OF DISTANCE/DISSIMILARITY MEASURES USED FOR CLASSIFICATION OF ELECTROENCEPHALOGRAM SIGNALS

Erçin ÖZCAN

MSc THESIS

DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

We certify that the thesis titled above was reviewed and approved for the award of degree of the Master of Science by the board of jury on …./…./201….

……………………………	…………………………..…..	..………………………………
Assoc. Prof. Dr. Sami ARICA	Assoc. Prof. Dr. Turgay İBRİKÇİ	Assis. Prof. Dr. Melik KOYUNCU
SUPERVISOR	MEMBER	MEMBER

This MSc Thesis is written at the Department of Institute of Natural And Applied Sciences of Çukurova University.
Registration Number:

Prof. Dr. Mustafa GÖK
Director
Institute of Natural and Applied Sciences

ABSTRACT

MSc THESIS

A COMPARATIVE STUDY OF DISTANCE/DISSIMILARITY MEASURES USED FOR CLASSIFICATION OF ELECTROENCEPHALOGRAM SIGNALS

Erçin ÖZCAN

ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

Supervisor : Assoc. Prof. Dr. Sami ARICA
Year : 2015, Pages 66
Jury : Assoc. Prof. Dr. Sami ARICA
: Assoc. Prof. Dr. Turgay İBRİKÇİ
: Asst. Prof. Dr. Melik KOYUNCU

Biological signal is a general term that refers to the signal measured from a biological system. Some representative examples include the electrocardiogram, the blood pressure waveform, the cellular action potential, etc. Many biological signals show distinctive waveform morphology which reflects the dynamics of the biological systems.

The electroencephalogram (EEG) signal is a measure of the summed activity of approximately 1–100 million neurons lying in the vicinity of the recording electrode, and may provide insight into the functional structure and dynamics of the brain. Therefore, the exploration of hidden dynamical structures within EEG signals is of both basic and clinical interests.

In clinical practices EEG is used to diagnose or monitor the following health conditions. In a computer aided diagnosis abnormal activities are detected and normal and abnormal activities are distinguished.

An electroencephalograph (EEG)-based communication system, also known as brain–computer interface (BCI), utilizes the information in EEG and provide a new communication channel for patients with several motor disabilities, such as brain stem infarct or amyotrophic lateral sclerosis. The BCI requires classification or distinction of the information in EEG or state of EEG.

All these applications necessitate distinction of state of the EEG. The distance or dissimilarity measures separability of the states of the EEG and provides that two or more EEG patters are different and each of them corresponds to a distinct state.

In this study distance/dissimilarity measures for classifying EEG will be investigated and distinct features they recognize will be determined.

Key Words: EEG, ECG, Distance/Dissimilarity Measures

ÖZ

YÜKSEK LİSANS TEZİ

ELEKTROENSEFALOGRAM SİNYALLERİNİN SINIFLANDIRILMASINDA KULLANILAN BENZERLİK/BENZERSİZLİK ÖLÇÜMLERİNİN KARŞILAŞTIRILMASI

Erçin ÖZCAN

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

Danışman : Doç. Dr. Sami ARICA
Yıl: 2015, Sayfa 66
Jüri : Doç. Dr. Sami ARICA
: Doç. Dr. Turgay İBRİKÇİ
: Yrd. Doç. Dr. Melik KOYUNCU

Genel bir terim olan biyolojik sinyaller, biyolojik bir sistemden kaydedilen sinyal türüdür. Örnek vermek gerekirse; elektrokardiyogram, kan basıncı dalga formu, hücresel potansiyel vb. Birçok sinyal, ilgili biyolojik sistemin ayırt edici özelliğini bünyesinde barındırır.

Bu sinyallerden birisi olan elektroensefalogram (EEG) yaklaşık olarak 1 ile 100 milyon nöron hücresinin aktivitelerinin sonuçları oluşan sinyal grubudur. Bu sinyal grubu ise beynin fonksiyonel yapısı ve hareketliliği üzerine çeşitli bilgiler sunmaktadır. Böylelikle, bu sinyaller içerisinde gizli olan yapıların keşfi beyin ile alakalı olan birçok bilimsel alanda önem arz etmektedir.

Klinik uygulamalarda EEG sinyalleri teşhis ve sağlık durumun gözlenmesi amacı ile kullanılmaktadır. Bilgisayar destekli uygulamalarda ise normal ve anormal durumların gözlemlemesinde kullanılmaktadır. Elektroensefalogram sinyalleri üzerine kurulu olan beyin-bilgisayar ara yüzü, EEG sinyallerini kullanarak motor engeli bulunan hastalara destek amacıyla kullanılabilmektedir. Bu engellere örnek verilecek olursa; Beyin Kökü Enfarktı veya Amiyotropik Laterel Sklerosis verilebilir.

Bütün bu uygulamalar EEG sinyallerinin sınıflandırılması gerektirmektedir. Bu bağlamda uzaklık veya benzerlik-benzersizlik ölçütleri bir veya birden fazla EEG sinyal örneklerinin karşılaştırılmasında ve hangi duruma ait olduklarının bulunmasında kullanılır.

Bu çalışmada, EEG sinyalleri sınıflandırılmasında kullanılan çeşitli benzerlik-benzersizlik ve uzaklık ölçütleri incelenecek ve bunların ayırt edici özellikleri belirlenecektir.

Anahtar Kelimeler: EEG, ECG, Benzerlik/Benzersizlik Ölçütleri

ACKNOWLEDGEMENTS

The subject of this thesis was suggested by my supervisor, Assoc. Prof. Dr. Sami ARICA to whom I would like to express my heartfelt thanks for his supervision, guidance, encouragements, extremely useful suggestions and continuous confidence in me throughout this thesis. It has been a great honor to have Assoc. Prof. Dr. Sami ARICA as supervisor.

Finally, I wish to express my deepest gratitude to my wife Gözde ÖZCAN for her endless support, encouragement and patience.

Erçin ÖZCAN

CONTENTS **PAGE**

LIST OF TABLES **PAGE**

LIST OF FIGURES **PAGE**

LIST OF SYMBOLS

S^w	: The Morlet Kernel
$s(t)$	: Signal Under Inspection
t	: Time
f	: Frequency
Δt	: Time Difference
π	: Pi
e	: Euler number
B	: MEG Measurement For a Given Instant
V	: EEG Measurement For a Given Instant
J	: The Vector of Distributed Model’s Magnitude
G_B	: Gain MatriX Respectively for MEG measurements
G_V	: Gain MatriX Respectively for EEG measurements
E	: Conditional Entropy
b_B	: Additive Noise
b_V	: Additive Noise
K	: Number of Conditions
I_d	: Direct Current
p	: Weighting Coefficient
r	: Weighting Coefficient
I	: Dimension of X
K	: Dimension of X
T	: Dimension of X
X	: Source Current
G	: Source Current
M	: Source Current
a	: Given Vector
b	: Given Vector
c	: Given Vector
q	: Given Scalar

$\|\boldsymbol{x}\|_p$	: Vector Norm
N	: Number of Elements
p	: Vector Norm
$H(x, y)$	: Entropy

LIST OF ABBREVATIONS

EEG	:	Electroencephalogram
BCI	:	Brain-Computer Interface
ECG	:	Electrocardiogram
EMG	:	Electromyogram
MMG	:	Mechanomyogram
EOG	:	Electrooculography
GSR	:	Galvanic Skin Response
MEG	:	Magnetoencephalogram
MRI	:	Magnetic Response Imaging
AgCl	:	Silver Chloride
ADC	:	Analog To Digital Convertor
DSP	:	Digital Signal Processor
DPSS	:	Discrete Prolate Spheroidal Sequences
MI	:	Mutual Information
CE	:	Cross Entropy
JE	:	Joint Entropy
RE	:	Relative Entropy
ICA	:	Independent Component Analysis
SNR	:	Signal To Noise Ratio
SAC	:	Similarity Assessment Criteria
DC	:	Dissimilarity Cost
RMSE	:	Root Mean Square Error
MAE	:	Mean Absolute Error
AR	:	Auto Regressive
2D	:	Two Dimensional

1. INTRODUCTION

1.1. Bioelectrical Signals

The human body can be associated with a machine in which every process causes some sort of changes in the inner environment of it. These changes, whether they are chemical, acoustic or electrical, generates some sort of signals whereby we can infer information regarding the relevant processes. Almost most of such signals do not directly convey the desired information, but we can extract or, loosely speaking, decode it by considering structure of the signals in accordance with the associated processes. Since the 'extracted or decoded' signals exhibit properties relevant to the processes in the body, any structural failures or flaws sign, by and large, presence of some abnormalities. Therefore, such extraction or decoding may be very helpful in detecting and explaining pathological cases.

The most eminent signals used in biomedical applications are Electrocardiogram (ECG) Electromyogram (EMG), Mechanomyogram (MMG), Electrooculography (EOG), Galvanic skin response (GSR), Magnetoencephalogram (MEG) and Electroencephalogram (EEG).

1.1.1. Electrocardiography (ECG) Signals

Electrocardiogram is a type of bioelectrical signal processing approach in which electrical activities of human heart is taken under consideration. As we all know, during its normal activity the heart makes a regular movement in which it contracts or swells. Therefore, it generates some sort of electrical waves and this waves also induce electrical signals on the skin. These signals are then captured and processed by an ECG signal processing system and intrinsic properties of the heart under consideration are deduced. In case of any abnormality in the activity, these signals diverge from the usual pattern and signify the presence of a problem or problems of the heart. Therefore, ECG signals can be used in diagnostic approaches and be fruitful in the detection of cardiac failures.

1.1.2. Electromyography (EMG)

As we all know, the skeletal muscles are controlled by the neuromuscular mechanism and any abnormality in this mechanism in most cases causes pain, movement problems and so forth. The electromyography is utilized to examine behavior of the neuromuscular system by capturing the electrical signals generated during the activities of muscles. Such signals are deployed to find the properties of the neuromuscular system for a patient and unusual signals in most cases are signs of a problem. Therefore, this approach is used in diagnostic purposes where the neurological or muscular structures are under consideration.

1.1.3. Mechanomyography (MMG)

Mechanomyogram is an approach in which mechanical properties of the muscles are examined. When they are contracted, the muscles produce acoustic vibrations and these vibrations are used to extract information respecting the properties of the muscles. The vibrations are converted to electrical signals by sensitive accelerometers or microphones attached to the skin and in the presence of any problem, large peaks in the signal are observed.

1.1.4. Electrooculography (EOG)

Electrooculography is used to measure eye movement. In this method, two electrodes are inserted to the eye under consideration one of which detects positive movement and the other negative. Positive movement is the case when the eye moves towards the positive electrode and the negative is vice-versa.

1.1.5. Galvanic Skin Response (GSR)

It is a well-known fact that the presence of sweat over the skin alters electrical conductive property of it. Since mechanism of sweating is managed by the nervous system, skin conduction is also affected by the nervous activities. Thus, electrical conductance of skin can give some clues regarding the psychological or physiological conditions.

1.1.6. Magnetoencephalography (MEG)

As known, neurological activities in the brain cause electrical currents in the tissue and these currents also generates changes in the magnetic field around the brain. Magnetoencephalography is exploited to map brain activities in accordance with the alteration of magnetic field. Since the changes are in the femtotesla scale, use of very sensitive magnetic sensors is a requisite. This technique requires magnetic isolation to suppress noise occurring due to the Earth’s magnetic field.

1.2. Electroencephalography (EEG)

Activities in the brain causes electrical ion flows in the brain neurons thus induce electrical currents on the skull skin. These currents are recorded by an EEG system to analyze temporal and spatial brain activities. Since the activities are done in accordance with the neuronal oscillations, spectral properties of the signals are the most important and thus are mostly probed feature of the signals in the diagnostic approaches.

There are mainly six frequency bands in the EEG signals which occur in different activities of the brain. Namely, Delta 0-4Hz, Theta 4-7Hz, Alpha 8-15Hz, Mu 8-12, Beta 16-31Hz and Gamma >31.

Delta band signals come into existence mostly during sleep and have high amplitude. Theta and alpha band signals are present while the brain is in an active task. Beta band signals occur when the active thinking is being done and gamma band signals happen during some memorial activities.

1.2.1. Use of EEG

For the research purposes, the EEG is used extensively in neuroscience, psychology, cognitive science, neurolinguistics, and so forth. Its experimental setup is considerably cheaper than those of other methods such as MRI, CT, etc. and provides much temporal resolution. Additionally, it is totally non-invasive and does not require an isolated room feature.

With regard to its diagnostic use, it is used in order for abnormal activities or presence of abnormalities in the brain to be detected. For example, epileptic activities of a brain cause manifestly abnormal EEG results and therefore the method is widely used to diagnose epilepsy non-invasively. The EEG is also exploited to detect brain death, coma and diagnose sleep disorders. Further, albeit with less obviousness and resolution, it can also be utilized to examine the presence of tumors or structural abnormalities.

1.2.2. Signal Acquisition and Processing of EEG Signals

The EEG signals are within the microvolt scale (Teplan 2002) and requires amplification ratio about 1000-1000000. Since the amplitude of the signals is very weak, the acquisition process is very sensitive to the presence of noise and thus there occurs need for high quality and fidelity acquisition systems. There are four main parts in the EEG signal acquisition and processing system (Teplan 2002). The first part involves electrodes and conductive components attached to head skin to convey electrical signals to amplifiers. To attach electrodes to the right places, usually an electrode cap is used. Such a cap encompasses AG-AgCl discs with flexible cables conveying the signals to amplifiers (Bronzino 1995, Picton 2000). For the attachment points standardized by the "International Federation in Electroencephalography and Clinical Neuurophysiology" (Jasper 1958) and their names on the head surface see figure 1.1.

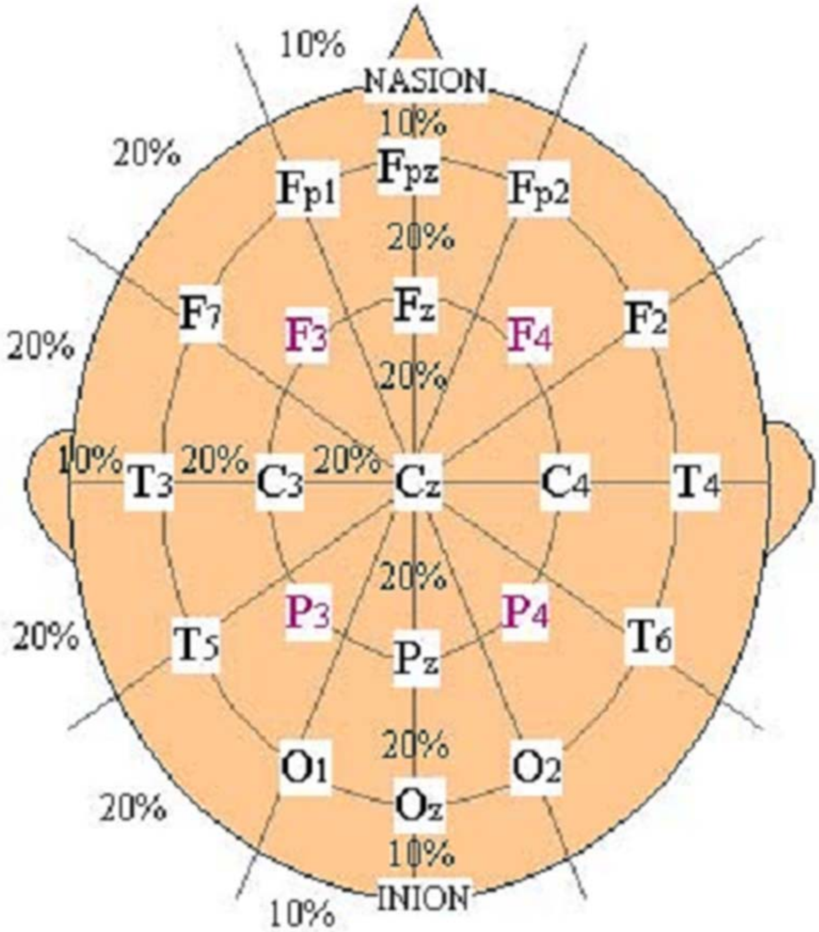

Figure 1.1. Electrode attachment points on the head surface with their names in literature. (Teplan 2002)

The second one consists of amplifiers which amplify microvolt scale signals to the level appropriate for the sampling and quantization process done by analog to digital converters (ADC). Since the signals have very low level in amplitude, the amplifiers should have high precision and fidelity in order not to distort them. Additionally, the signals are sampled at a fixed frequency and thus to satisfy Nyquist requirement for the sampling, this stage also involves anti-aliasing filter. The third part is composed of ADC group whose duty is to convert analog signals to digital signals to be processed by a computer or digital signal processor (DSP). After the signals are sampled, quantized and converted to digital form, a computer system or DSP processes the digitalized signals. This process includes band pass filtering, clustering, pattern recognition, and so on.

In most cases, the main objective of the signal processing step is to develop brain computer interfaces (BCI) and distinguish different mental tasks. The most significant goal of the BCI applications is to make assistance for the people having motor disabilities, establish a communication with computers via EEG signals. Performance of the BCI systems has a concrete relationship with processing algorithms and signal distortion or noise. The most important contributors of the noise are eye movements/blinkings, power grid and hearth beats. Such distortions are removed by suitable approaches specific to the type of artefact. For instance, noise stemming from the power line is eliminated by using a Notch filter.

With regard to discrimination of different mental tasks, after the removal of distortions, there remains the most significant part of the EEG signal processing in which any neuronal activity are distinguished and associated with pre-defined or determined activities. For example, if a captured signal has features resembling to that occurring when the right hand is moved, it can be inferred by the EEG system that there is a motion in relationship with the right hand. Therefore, a response by the system can be made to satisfy the determined need for the right hand movement. Such a system can be useful for the people having arm or leg prosthesis.

Association of a sampled signal train with a defined signal cluster is of primary importance in many signal processing problems. Presently, there are many studies, approaches and theoretical frames available for the task and the same is also true for the EEG signals. The main idea behind the process is to assess resemblance of the signal under consideration to the defined signal clusters and find the most appropriate choice by which the highest resemblance is achieved for a chosen cluster.

There are mainly two types of signal comparison methods which evaluate similarities of the signals given. The first one is intensity based approach and the second one is based on the statistical frames of the signals under consideration. Note that these two are still available for transform domain parameters such as Fourier transform and wavelet transform parameters etc. and difference between direct means and transform domain is not stressed in this study. The only attention is paid to the similarity criteria.

In the intensity based approaches, vectoral distance between the signals for a given vector space norm is found and this distance constitutes the measure of similarity of the signals. The more the distance, the less the resemblance is expected. In this approach, the final decision is made in favor of the lest distance between the signals.

In the statistical approaches, statistical features of the signals are used to evaluate their similarities to each other. Joint entropy and mutual information measures for the signals can be given as examples in this respect.

1.3. Aim of the Study

There are many similarity criteria in the literature which have different discrimination performance. Differences between their performances stem from the fact that each approach handles the problem with different attitudes, in different norms etc.

The aim of this study is to compare different signal similarity criteria, find their relative performances and present their superior and inferior sides. In the comparisons, simulated EEG signals are used as a base and there is pre-knowledge about similarities of them which is helpful in the assessments. We, in the experiments, exploit MATLAB as a simulation and computation medium and the codes will be supplemented at the appendices section.

The rest of the study is as follows. In the next chapter, the previous works will be addressed and in the following chapter, materials and methods used in the study are discussed. The results are presented in the penultimate chapter and finally, the study will be concluded with the conclusion and discussion chapter.

2. PREVIOUS WORKS

2.1. Previous Studies

In this chapter, the previous works are discussed to give an insight to the reader.

In their study, Vugt et al. compares three similarity search algorithms using spectral analyses of the signals (Vugt et al. 2007). These are wavelets, multitapers and $P_{episode}$ approaches. They use simulation methods to examine characters of each method in detail. It is shown that $P_{episode}$ is more sensitive to the distinctions in length, wavelet approach has high sensitivity to lower signal amplitudes and multitapers method gives better results considering frequency. It is also recommended by the group that to benefit from the all good aspects of each method, one should use more than one method. With regard to the wavelet approach, since it is common in EEG analyses, the Morlet wavelet kernel is used. The Morlet kernel is defined as follows;

$$S^w = s(t) * \frac{1}{(\sigma_t\sqrt{\pi})^{\frac{1}{2}}} e^{-(t^2/2\sigma_t{}^2)} e^{i2\pi ft}. \qquad (2.1)$$

Where s(t) is the signal under inspection, S^w is wavelet-transform of it, t and f represent time and frequency and * denotes convolution.

In the multitaper method, the data are, by and large, multiplied by Discrete Prolate Spheroidal Sequences(DPSS) window and after than a Fourier transform is done.

$$S^M = \frac{1}{K}\sum_{k=1}^{K} \Delta t| \sum_{k=1}^{K} DPSS(s(t))e^{-i2\pi ft\Delta t} |^2. \qquad (2.2)$$

Where DPSS is function window, s(t) is the signal and the others denote Fourier transform.

$P_{episode}$ is computed by using background spectrum found through mean wavelet power of all experimental sets.

Baillet et al. proposed a cooperative processing method for EEG and MEG signals to improve localization of active sources (Baillet et al. 1999). In their proposal, they deployed statistical approaches mutual information (MI) and conditional entropy (CE) as similarity assessment criteria. The main idea behind the proposal is to reduce redundancy of information between the MEG and EEG signals. This aim is achieved by diminishing MI and CE between transfer functions of the signals. Mathematical description of the proposal is as follows.

$$\boldsymbol{B} = \boldsymbol{G}_B \boldsymbol{j} + \boldsymbol{b_B}$$
$$\boldsymbol{V} = \boldsymbol{G}_V \boldsymbol{j} + \boldsymbol{b_v} \tag{2.3}$$

Where $\boldsymbol{j}$ is the vector of distributed model's magnitude, $\boldsymbol{B}$ and $\boldsymbol{V}$ MEG and EEG measurement for a given instant, and $\boldsymbol{G}_B$ and $\boldsymbol{G}_V$ stand for gain matrices respectively for MEG and EEG measurements. Following that, conditional entropies $E(\boldsymbol{B}_j|\boldsymbol{V}_j)$ and $E(\boldsymbol{V}_j|\boldsymbol{B}_j)$ are found. The conditional entropies are measures of uncertainty of the distributions for the given conditions. Using all, mutual information for the given EEG and MEG signal arrays are found and minimized for the best alignment.

In their study, Gramforth and Kowalski develop an improvement for MEG or EEG signal source localization. The similarity measures used in their study are based on intensity approach and the main attention is paid to absolute, Euclidian and mixed norm distances considering ℓ_1 and ℓ_2 norms. The mixed norm in their study is defined as follows;

$$\|\boldsymbol{X}\|_{w;pqr} = (\sum_{i=1}^{I}(\sum_{k=1}^{K}(\sum_{t=1}^{T} w_{i,k,t}\left|x_{i,k,t}\right|^{p})^{q/p})^{r/q})^{1/r} \tag{2.4}$$

Where is weighting coefficient, p, r and are norm degree values, I, K and T are dimension of X and I, k and t are their index.

The problem is defined in the study as follows;

$$X^{*} = arg_{X}^{min}\|\boldsymbol{M} - \boldsymbol{GX}\|_{F}^{2} + \|\boldsymbol{X}\|_{w;212}^{2} \tag{2.5}$$

Where X, G and M denote current sources, forward operation matrix and measurement data, respectively.

Xu et al., in their study, propose an EEG source imaging algorithm relying on spatial sparse assumption of brain activities. For this purpose, Lp norm for $p \leq 1$ is used to establish sparse solution. The main attention is paid to the sparsity of the signals under consideration and regularized cost function of the estimation is constructed in this respect. The estimation for the activity location is done in accordance with the algorithm given below;

$$\hat{X} = arg_{X}^{min}\|\boldsymbol{Y} - \boldsymbol{AX}\|_{F}^{2} + \|\boldsymbol{X}\|_{p} \tag{2.6}$$

Such a function is hard to be minimized and some special optimization approaches must be adopted. In the study, a non-convex optimization algorithm named as BFGS (Li and Fukushima 2001) is used.

A feature selection method based on independent component analysis (ICA) and mutual information maximization algorithms was developed by Lan et al. The method is as follows. First, signals are preprocessed by an ICA algorithm to extract independent features. Then these features are sorted by using mutual information based similarity measurement algorithms. Visualization of the algorithm is given by the figure 2.1.

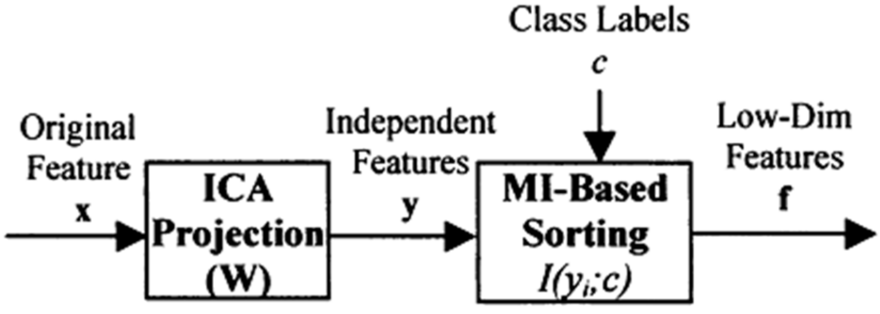

Figure 2.1. Visualization of them

In their study, Wendling et al. (Wendling et al. 2009) review three assessment families are taken into consideration. Namely, regression (linear and nonlinear), phase synchronization and generalized synchronization. Evaluation is done considering model-based methodology and a priori knowledge for the problem regarding the relationship between signal generating systems is present. The conclusion they do in their study is;

1. The methods in comparison are not affected by the coupling parameter which defines interrelationships between signals.
2. There is no state of the art method handlind signal coupling.
3. Result are dependent of signal properties.

In the paper (Arnhold et al. 1999), a new data statistics based SAC is introduced and it is compared with other statistics based SACs. The main aim of the work is to detect weak interdepence of signal clusters and it is shown that it outperforms others in this regard.

A novel method relying on conditional mutual information and permutation analysis for estimation of a directionality index between two neuronal populations is proposed (Li and Ouyang 2010). In the study, the method is compared considering noise, signal length, coupling strength and time dependent coupling with several methods and the conclusion is as follows;

1. Coupling a sample length greater than 5 s is required to estimate the directionality index with bidirectional coupling.
2. The value of directionality index for unidirectional coupling is affected considerably by the coupling strength.
3. There is also significant directionality index is dependant of the SNR level.
4. The proposed method outperforms the conditional mutual information method and Granger causality method in detecting the coupling direction.

Ouyang et al., in their paper, propose a SAC for EEG signals based on permutation entropy of the signals (Ouyang et al. 2010). Initially, EEG signal sequence fragments are mapped into a phase space, ordinal sequences based on permutation entropy of signals and the distribution of these ordinal patterns are calculated. It is shown that dissimilarity measure for different brain states are hgiher than those of intra-states and the method proposed performed well. It was also stated that the method can be utilized in classification of dynamics of EEG signals.

3. SIMILARITY ASSESSMENT METHODS

In this section, we discuss the similarity/resemblance assessment criteria (SAC) mentioned briefly afore in detail by giving their intrinsic properties so as to invoke some intuitive or direct understanding in the readers' minds. We handle two types of SAC in this study; intensity based and statistics based.

3.1. SACs Based on Intensities of Data Vectors

Intensity based SACs essentially find the distance between two data vectors in a given vector space norm. In such approaches, it can easily be said that the more the distance between the two data vectors, the less the two resemble to each other. Before giving details of the norms considered in this study, some definition for the vector space may be elaborative. In this respect, it is suitable to commence with vector spaces.

3.2. Vector Spaces

In definition, a vector space V is a non-empty set of vectors and there are two operations and ten axioms on the vectors in this space (Aubin 1999). The two operations are addition of vectors and multiplication of vectors by scalars. The ten axioms are given below;
For the given vectors **a, b** and **c**

1. **a**+**b** is in V.
2. **a**+**b**=**b**+**a**.
3. **(a+b)+c=a+(b+c)**.
4. There is a vector **0** in V such that **a+0=a,** which is called as zero vector.
5. V also includes negative of **a** which satisfies **a+(-a)=0**
6. For a given scalar q, q**a** is also in V.
7. q(**a**+**b**)=q**a**+q**b**.

8. (q+p)**a**=q**a**+p**a**.
9. (qp)**a**=q(p**a**).
10. 1**a**=**a**.

Further, each vector in the V has a magnitude varying according to the norm selection and vector norm also satisfies these conditions;

1. $\|\boldsymbol{x}\| > 0$ when $\boldsymbol{x} \neq 0$ and $\|\boldsymbol{x}\| = 0$ if and only if $\boldsymbol{x} = \boldsymbol{0}$.
2. For a scalar k $\|k\boldsymbol{x}\| = |k|\|\boldsymbol{x}\|$.
3. $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$

Further, a vector norm is found as follows;

$$\|\boldsymbol{x}\|_p = \left(\sum |x_i|^p\right)^{1/p} \tag{3.1}$$

The most popular one is Euclidian norm for p=2;

$$\|\boldsymbol{x}\|_2 = \sqrt{x_1^2 + x_2^2 \ldots + x_n^2} \tag{3.2}$$

To visualize behavior of the norm with respect to different values of p, plots of the unit circle in norms p=0.5, p=1, p=1.5 and p=2 are given in figures 3.1, 3.2, 3.3 and 3.4.

For each value of p the figures are acquired using this formula;

$$\|\boldsymbol{x}\|_p = \left(\sum |x_i|^p\right)^{1/p} = 1 \tag{3.3}$$

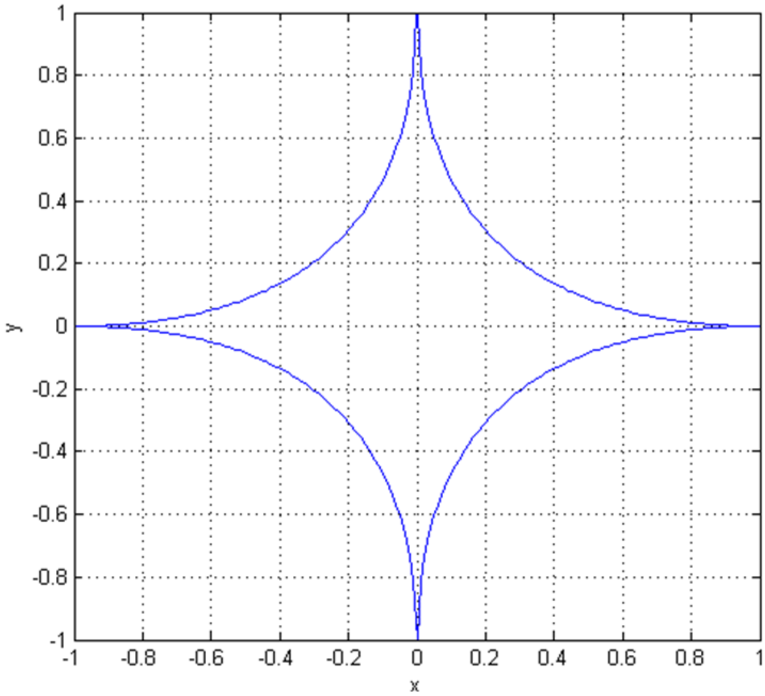

Figure 3.1. Unit circle in p=0.5 norm.

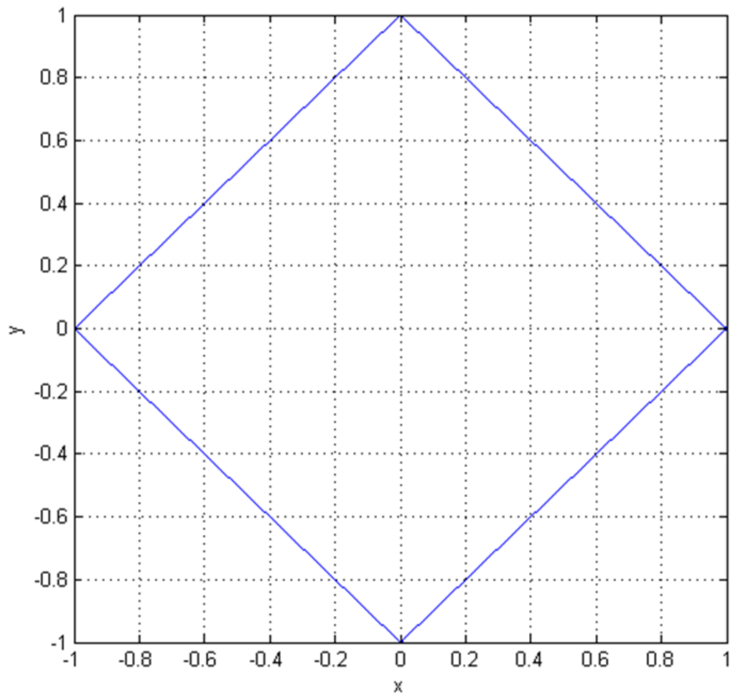

Figure 3.2. Unit circle in p=1 norm.

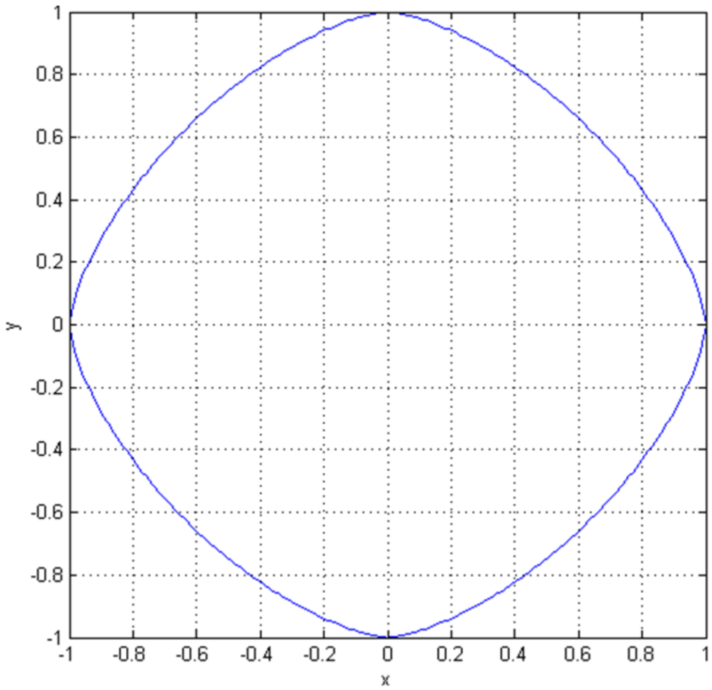

Figure 3.3. Unit circle in p=1.5 norm.

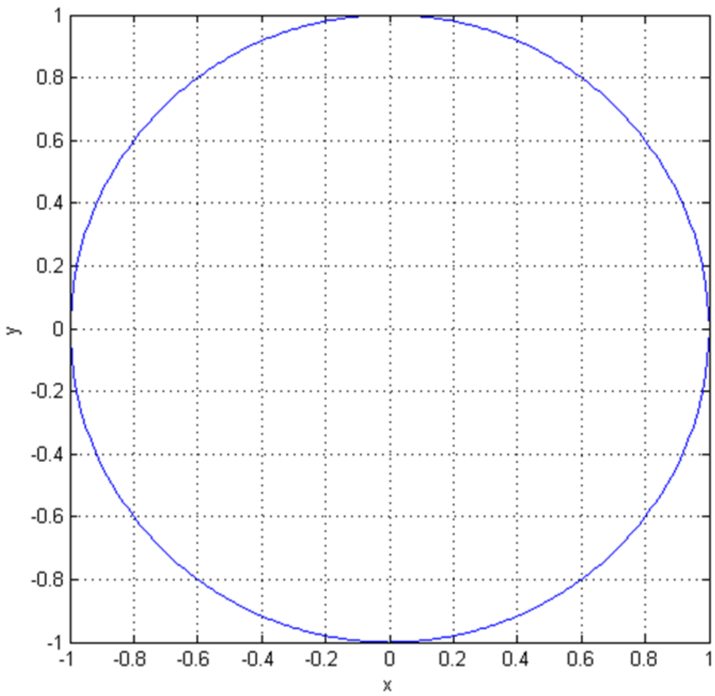

Figure 3.4. Unit circle in p=2 norm.

However, it should be noted that any norm for the values of p below 1 called as semi-norm. This is because for p<1, the third axiom is not satisfied since $\frac{d^2x^p}{d^2x} < 0$ for x>0.

For the p values greater than 1, magnitude of a vector norm is dominated by the biggest elements of the vector. However, when the p is below 1, contribution of elements having small values gets significance. This issue has a direct relationship with sparsity of vectors or vector transforms values. To visualize this, equation and figure 3.5 is given below.

$$y = |x|^p \tag{3.4}$$

In figure 3.5, it is obvious that for a given threshold between 0 and 1, the norm p=0.5 includes more points than the p=2 norm. For instance, if the threshold is chosen as y=0.5, the span is between 1 and 0.25. However, for p=2 this span is between 1 and ~0.7. Therefore, it can be said that contribution of small values has an inverse proportion to the p.

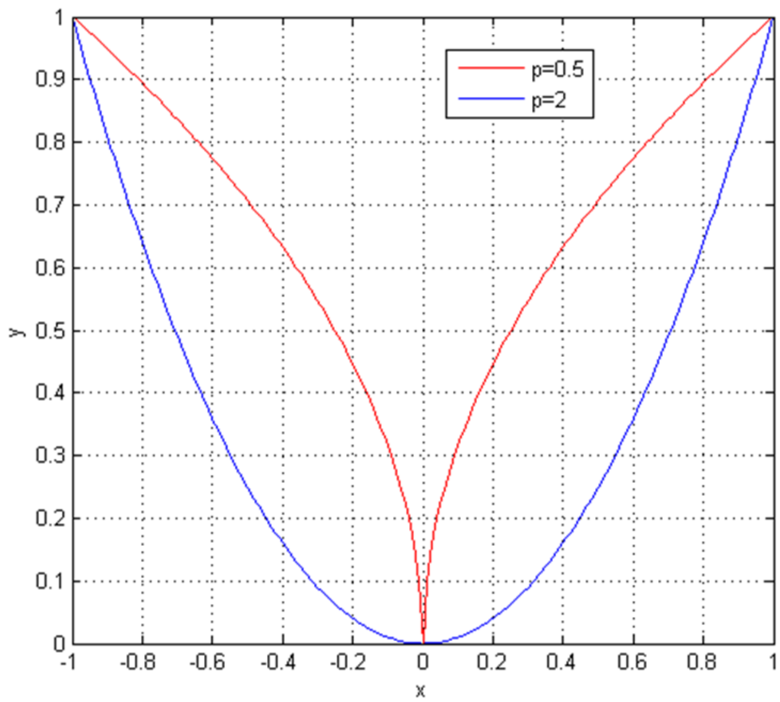

Figure 3.5. Plot of y with respect to x considering (3.4).

3.3. Different Intensity Based SACs

As we saw previously, behavior of a norm changes considerably for different values of p. For this reason, we tried to examine several norms for different SACs in this study. The examined values of p are p=0.5, p=1, p=1.5 and p=2.

For the given two data vectors, say **x** and **y**, inverse SAC or dissimilarity cost (DC) is defined as follows;

$$DC = \|\boldsymbol{x} - \boldsymbol{y}\|_p^p \tag{3.5}$$

For p=2, DC is named as square error and minimization of the DC_2 essentially becomes a least squares problem (Strang 2009). Moreover, the so-called term root mean square error (RMSE) has a direct relationship with 3.5 and is defined as follows;

$$RMSE = \sqrt{\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 / N}. \tag{3.6}$$

Where N is element number of the vectors **x** and **y**.

The other common norm is p=1 norm and DC for this norm has a name; absolute error (AE).

$$DC_1 = \sum_{i=1}^{N} |x_i - y_i| \tag{3.7}$$

Similar to the relationship between DC_2 and RMSE, DC_1 has a strong relationship with mean absolute error (MAE) which is another golden-standard error measure in signal processing.

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |x_i - y_i| \tag{3.8}$$

It should be noted that for the increasing values of p, the most significant contributors are those having the biggest values and small elements are ignored when the problem is solved. Therefore, any measure based on this norm falls vulnerable to noise with high amplitude even for small number of elements. However, for the values of $0<p<1$, small elements also make considerable contributions and this condition increases robustness of the assessment for the high outliers in distortion.

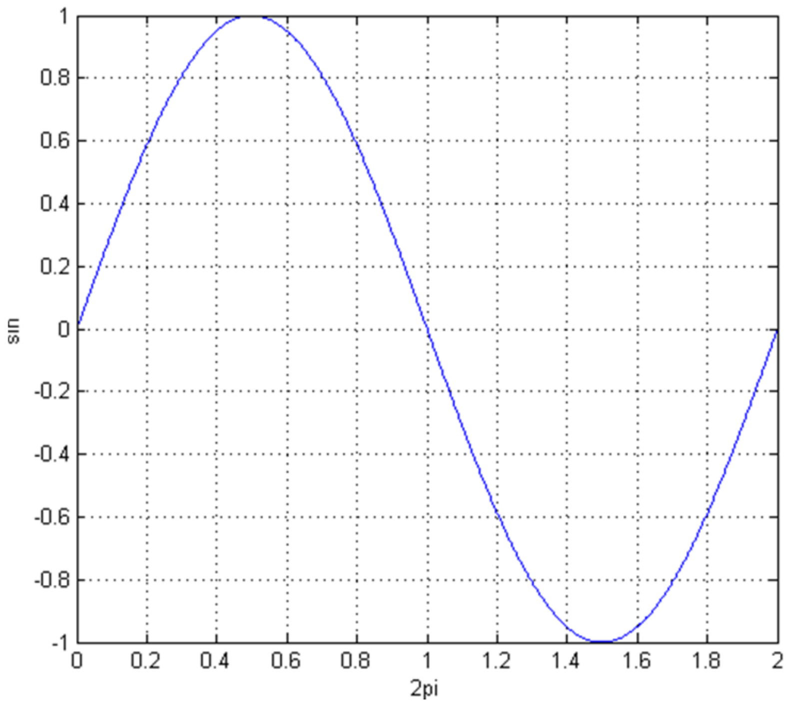

Figure 3.6. Sine function.

This pattern can be seen by taking the signals in figures 3.6, 3.7 and 3.8, and errors for different p values into consideration. In figure 3.6, we see plot of sine function in 0-2pi domain. To assess performances of SAC for different p values, we deliberately add two high spikes with amplitude 2.82 to the lower and upper hills and get the plot in figure 3.7. We also add white Gaussian noise to the sine with SNR=15dB and see the result in figure 3.8.

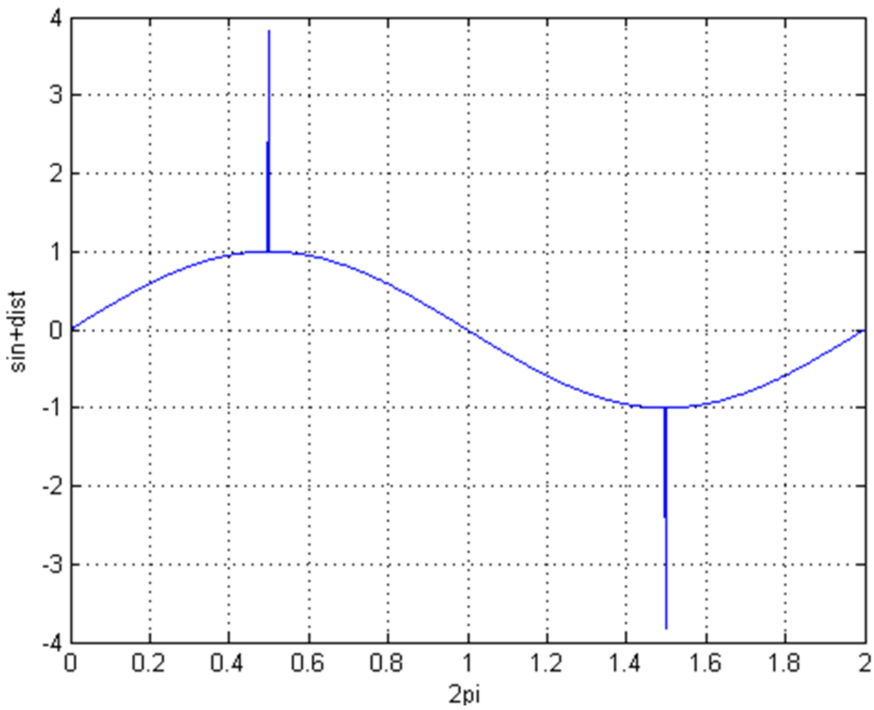

Figure 3.7. Distorted sine function.

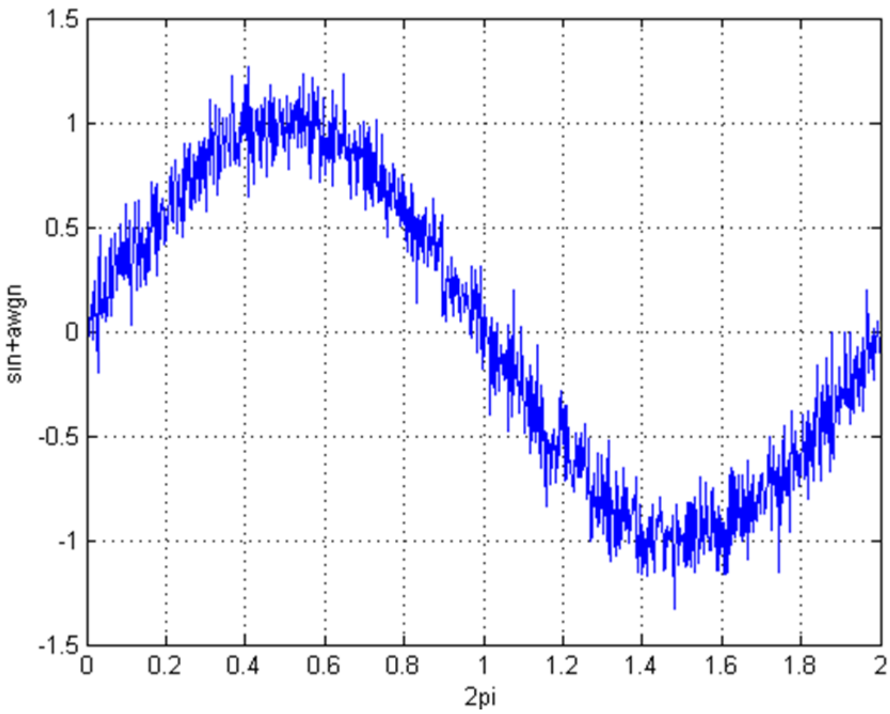

Figure 3.8. Sine function in the presence of additive white Gaussian noise with 15dB SNR.

If we look at the figures 3.7 and 3.8, it is apparent that the figure 3.7 is less distorted and resembles to the original signal in figure 3.6 more. If we compute DC of the signal in figure 3.7, we get the following results; for p=2, p=1.5, p=1 and p=0.5, DCs are 15.99, 9.51, 5.65 and 3.36, respectively. In words, the DC in p=2 norm has the highest result and the it diminishes for the decreasing values of p. However, if take the signal in figure 3.8 into account, which resembles to the original signal far more, we get the results; for p=2, p=1.5, p=1 and p=0.5, DCs are 15.99, 38.43, 99.89 and 289.76, respectively. Performance of the norm p=2 remains same. However, for the decreasing value of p, we get higher results and this is in harmony with our intuitive expectation; the signal in figure 3.8 resembles the original one less in comparison with that in figure 3.7. Thus, we can maintain that for the lower values of p, we get better results.

Let us consider another scenario in which we add a small bias 0.1 to the original signal and get the signal in figure 3.9.

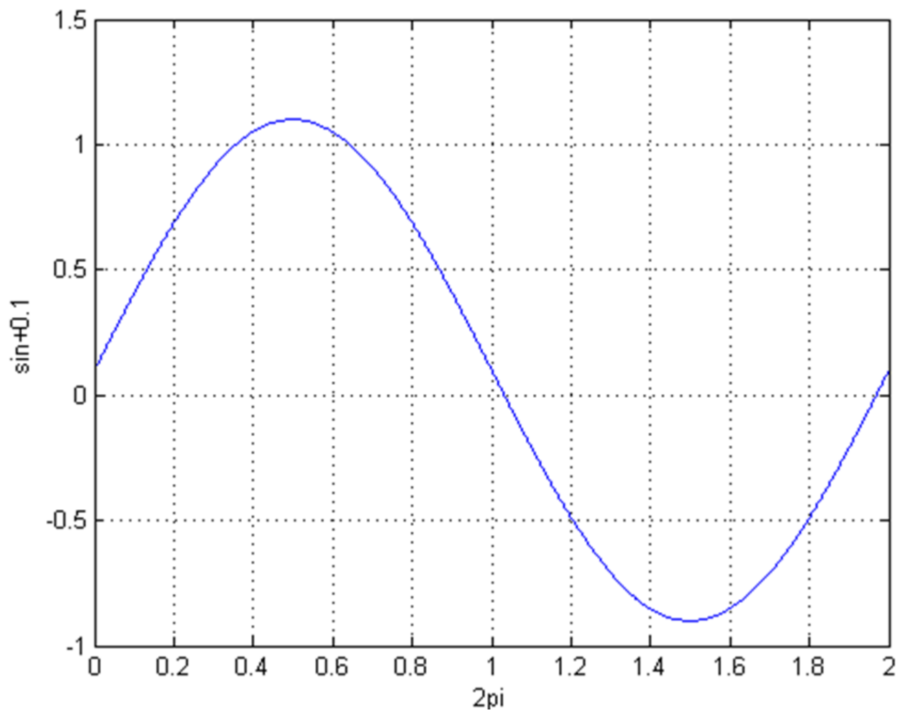

Figure 3.9. Sine function with the addition of bias 0.1.

For this scenario, DCs for each p value are as follows. DC=10 for p=2, DC=31.62 for p=1.5, DC=100 for p=1, DC=316.22 for p=0.5. From this trend, we can infer that even for the small values of bias, performance of the measurements decreases considerably for the decreasing values of p. Therefore, there occurs a tradeoff with respect to values of p and type of distortion. In the presence of distortion with high amplitude but having infrequent occurrence, the lower p values result in better performance. Nevertheless, the presence of even a small bias can alter this trend conversely and lower p values exhibits inferior performance. As a consequence selection of p profoundly depends on type of noise.

3.4. SACs Based on Statistics of Data

We previously examined the SACs based on data intensities and presented their behavior considering different types of them. As it will be shown afterwards however, they require iso-intensity property of data and even small magnification or fading in magnitude of one data cluster disrupt performance of them drastically. To overcome this problem, statistics of the clusters can be rather fruitful in measuring similarities of data clusters. In this respect, we, in this section, consider cross entropy (CE), joint entropy (JE), relative entropy (RE) and mutual information (MI). Before giving details of them, a brief discussion upon some basics is more appropriate.

3.5. Entropy of a Data Cluster

In assessing similarities of data clusters, evaluation of the information they share with each other can be very profitable. Information measure of a data cluster intrinsically exhibits how many different partition the cluster has. Intuitively, it can be said that if two data clusters share more data, their joint information measure should decline. In measuring information of a data cluster, the most famous one in signal processing is Shannon-Wiener entropy measure H (Shannon 1948, 1949) and defined as follows.

$$H = -p_i \sum_{i=1}^{N} p_i. \tag{3.9}$$

H is the information measure of a data cluster given its probability distribution p_i. This formula is based on three assumptions (Hill 2001);

1. The function should be continuous for p_i.
2. The function should be monotonically increasing when N increase for all equally probable $p_i = \frac{1}{N}$.
3. For the fragmented choice should satisfy the equality; $\mathrm{H}(p_1, p_2, p_3) = \mathrm{H}(p_1, p_2 + p_3) + (p_2 + p_3)\mathrm{H}(p_2/(p_2 + p_3), p_3/(p_2 + p_3))$.

It was proven by Shannon that the unique solution for the three conditions given is 3.9 and when all partitions in the cluster have equi-probability H is maximum. Additionally, in case of all partitions in the cluster have the same value, H is zero which signifies the absence of information.

Further, p_is are found from the histogram bin entries of the data cluster and width of each bin is user selective. For example, histogram and probability distribution of the data in figure 3.6 for 0.1 bin width is given by figure 3.10.

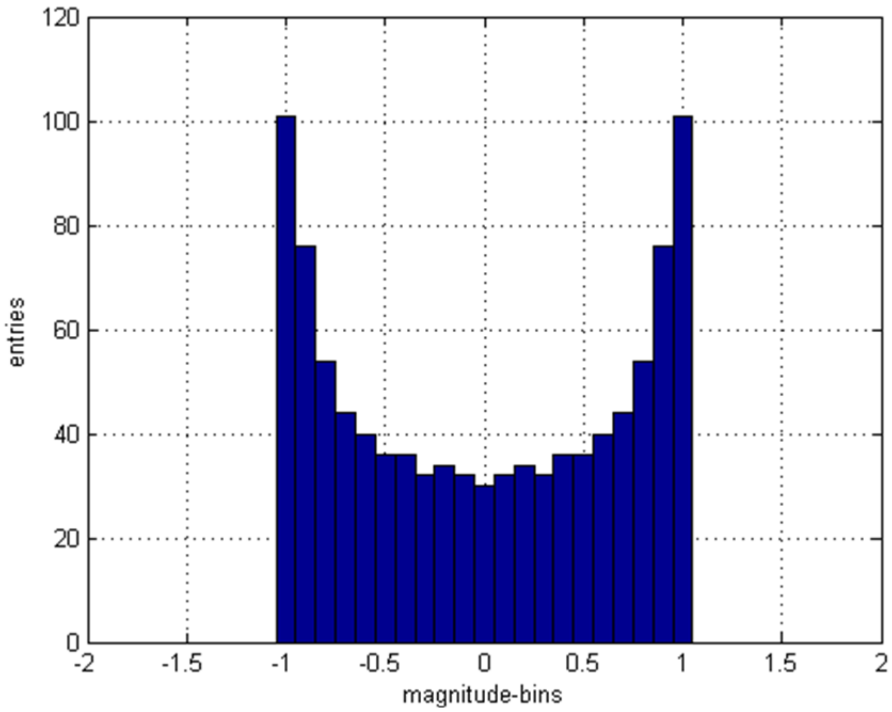

Figure 3.10. Histogram of sine function for the bin width 0.1.

3.6. Cross Entropy

Cross entropy is a measure of information shared by two data clusters and defined as follows;

$$H(x,y) = -\sum_{i=1}^{N} p_{x\,i} log_2 p_{y\,i}. \tag{3.10}$$

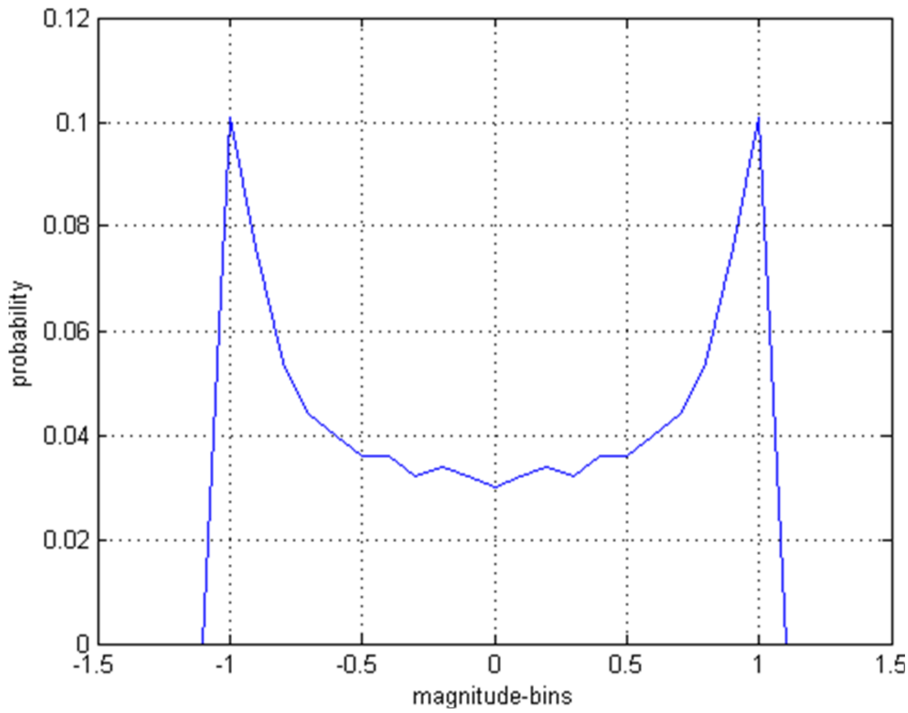

Figure 3.11. Probability distribution of histogram of sine function for the bin width 0.1.

3.7. Joint Entropy

Joint entropy is a measure of information shared by two data clusters and defined as follows;

$$H_{x\,y} = -\sum_{y=1}^{M} \sum_{x=1}^{N} p_{xy} log_2 p_{xy}. \tag{3.11}$$

3.8. Relative Entropy (Also Named as Kullback-Leibler Divergence)

Relative entropy is a measure of information shared by two data clusters and defined as follows;

$$H(x,y) = -\sum_{i=1}^{N} p_{x\,i} log_2(\frac{p_{x\,i}}{p_{y\,i}}). \tag{3.12}$$

3.9. Mutual Information

Mutual information is a measure of information shared by two data clusters and defined as follows;

$$H_{x\,y} = -\sum_{y=1}^{M} \sum_{x=1}^{N} p_{xy} log_2(p_{xy} / p_x p_y). \tag{3.13}$$

3.10. Performances of SACs in the Presence of Non Iso-intensity Data Clusters

In the previous sections, we presented intensity based SACs and compared their performances for different norms considering iso-intensity data clusters. However, as mentioned earlier, these measures generate unreliable results in the presence of mean intensity difference and such a problem can be surmounted by deploying information based SACs. To show this, we fade the intensity of signal given in figure 3.6 with 0.5, get the signal in figure 3.12 and compare the results of different SACs either intensity or information based. In table Table 3.1 we see different SACs and their results (b,c,d and e in the table respectively) for sine signal given in figure 3.6 and its distorted forms given in figures 3.7 (additive two spikes; b in the table), 3.8 (additive Gaussian noise; c in the table), 3.9 (addition of bias; d in the table), 3.12 (fading; e in the table). Qualitatively, it is obvious that the order of resemblance of the distorted signals to the original one is Fig.3.12 (also equal to Fig. 3.9), Fig. 3.9, Fig. 3.7 and Fig. 3.8.

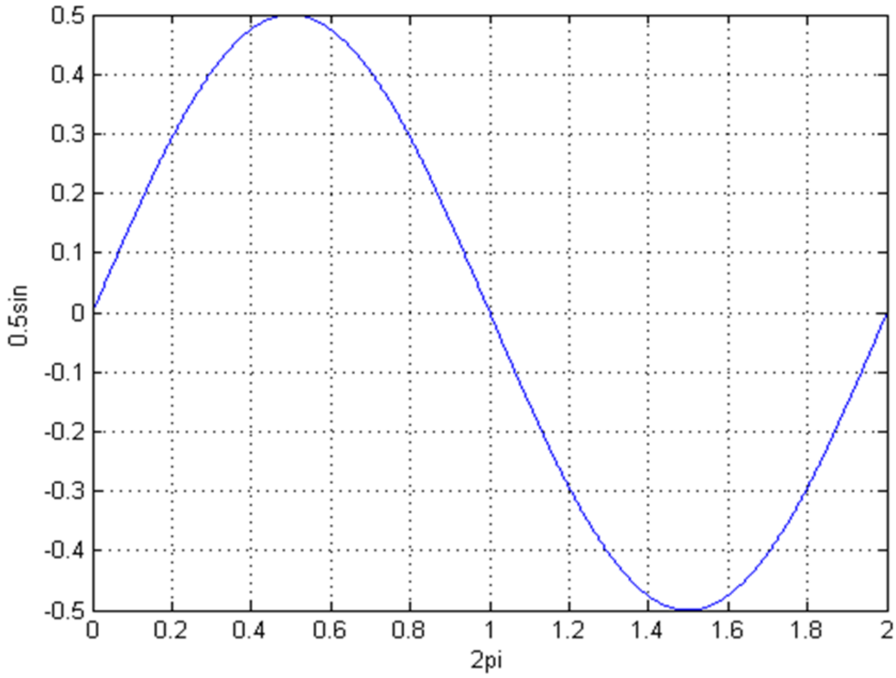

Figure 3.12. Sine function with amplitude 0.5.

As expected, vectoral distances in different norms fail in the presence of bias or amplitude magnification or fading and this trend can easily be seen in the table (d and e). However, this problem is surmounted by introducing statistics based SACs. There exists no difference in results of them when we compare the original signal with those with fading (e), with bias (d) and with itself (a in the table).

In the next section, we will contemplate on the synthetic data clusters and will compare the SACs using them.

Table 3.1. Different SACs and their results for sine signal with respect to its distorted forms.

SAC/Signal	a	b	c	d	e
MI	3,15	1,86	1,89	3,14	3,15
CE	3,15	2,57	3,83	3,15	3,15
JE	3,15	3,29	4,45	3,16	3,15
RE	0	-1,07	-0,69	0	0
L2 dist	0	15,9	15,78	10	125
L1.5 dist	0	9,47	38,02	31,62	196,72
L1 dist	0	5,64	99,41	100	318,31
L0.5 dist	0	3,36	290,24	316,23	539,31

4. RESULTS

In this section, we present the results of different SACs given the 19 channels data clusters which have 100 trials for each channel. The data cluster is synthesized by AR process (where coefficients a are zero) on Gaussian noise in accordance with the form given in Figure 4.1 and graph of its first trial for the first channel is given in Figure 4.2. We followed such a manner to imitate real EEG signals which are heavily contaminated by noise and such a trend can be seen in Figure 4.2. The Gaussian noise in the AR process has a normal distribution.

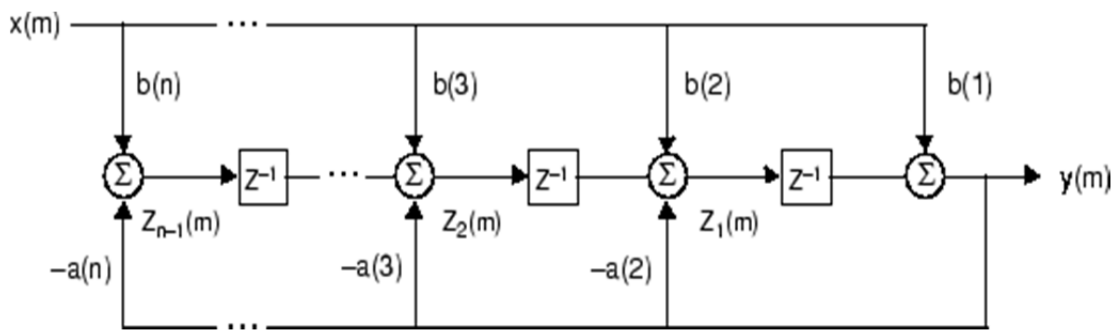

Figure 4.1. ARMA process applied on Gaussian noise in getting synthetic data sets (taken from MathWorks).

After creation of the synthetic data, we also added some sorts of distortions to the synthetic data so as to find results of the SACs for different types of distortions. These distortions are addition of Gaussian noise, addition of bias and fading. We initially present the results of SACs for the synthetic data in comparison with itself. In doing this, the main aim is to show that there is no distance between the two equal vectors in any norm given and statistics of two equal vectors is smoother than the other comparisons. In Table 4.1, we see the different SACs' results for the synthetic signal in comparison with itself.

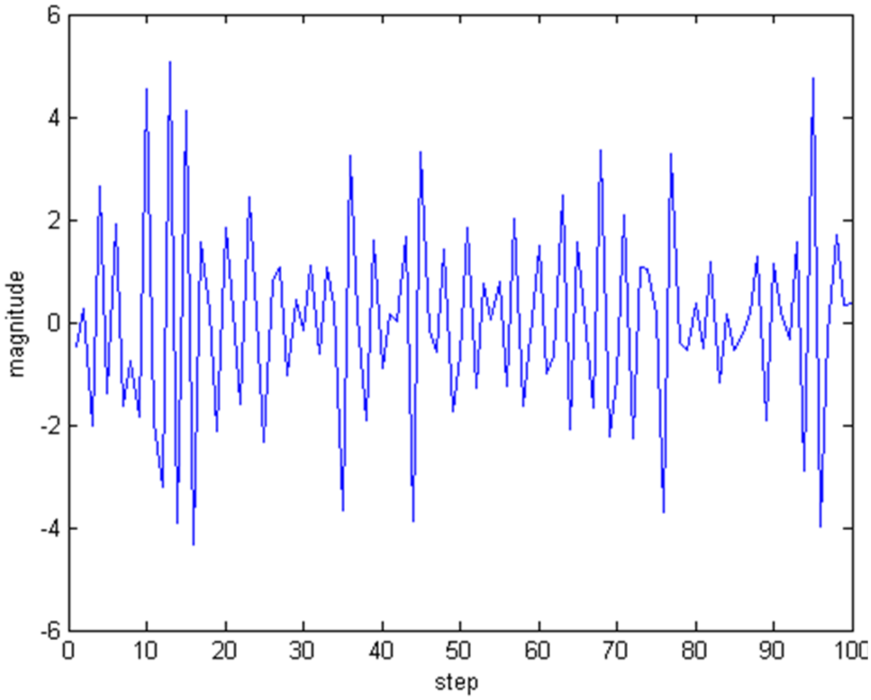

Figure 4.2. First trial of the first channel of the synthetic data.

Table 4.1. Comparison results of different SACs for the synthetic data and itself.

Channels	L0.5 Dist.	L1 Dist.	L1.5 Dist.	L2 Dist.	CrossEnt.	JointEnt.	M.I.	Rel.Ent.
1	0	0	0	0	4,14	4,14	4,14	0
2	0	0	0	0	4,1	4,1	4,1	0
3	0	0	0	0	4,14	4,14	4,14	0
4	0	0	0	0	4,14	4,14	4,14	0
5	0	0	0	0	4,11	4,11	4,11	0
6	0	0	0	0	4,13	4,13	4,13	0
7	0	0	0	0	4,11	4,11	4,11	0
8	0	0	0	0	4,14	4,14	4,14	0
9	0	0	0	0	4,13	4,13	4,13	0
10	0	0	0	0	4,14	4,14	4,14	0
11	0	0	0	0	4,15	4,15	4,15	0
12	0	0	0	0	4,09	4,09	4,09	0
13	0	0	0	0	4,12	4,12	4,12	0
14	0	0	0	0	4,12	4,12	4,12	0
15	0	0	0	0	4,12	4,12	4,12	0
16	0	0	0	0	4,14	4,14	4,14	0
17	0	0	0	0	4,11	4,11	4,11	0
18	0	0	0	0	4,12	4,12	4,12	0
19	0	0	0	0	4,13	4,13	4,13	0

As expected, result of such a comparison is always equal to zero considering the distance in any given norm. However, there occurs a different trend for the statistics based SACs; they are not equal to zero. This issue arises from the fact that even if the vector compared with itself, entropy of its distribution is not equal to zero and this causes non-zero values in most of statistics based SACs.

The inharmonious one is relative entropy solely which results in zero value due to its formula given in 3.12; both probability distributions are equal, their divisions are one and thus logarithms of them are zero. Further, due to the structures of the formulae of the MI, JE, and CE, their results are the same for the two vectors to be compared.

We also compared the synthetic data with its Gaussian noise with 15dB SNR added form given in Figure 4.3and results are given in Table 4.2. In the presence of noise, the synthetic data vector is perturbed and it diverges from its start point. Thus, a distance between the two forms occurs in any given norm.

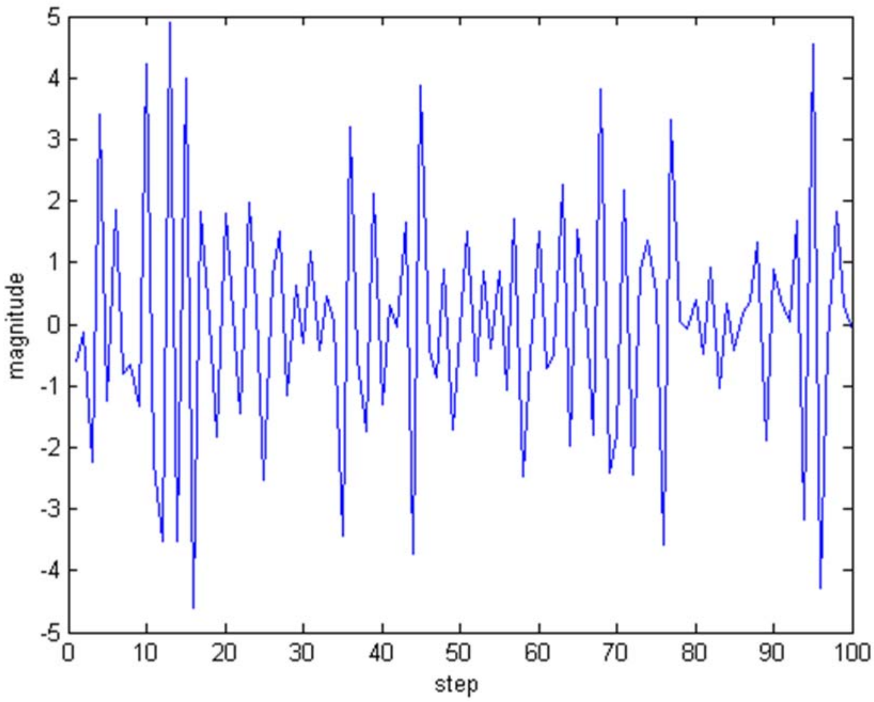

Figure 4.3. First trial of the first channel of the Gaussian noise contaminated synthetic data.

With the increasing distance values of norm degree, results raise and this is what is expected. However, this perception also involves a flaw for some cases. If noise is small in amplitude and scattered throughout the data, results of the norms having less degree would be more in quantity. This issue has an implicit but strong relationship with sparsity and has analogy with such a scenario; assume that we take square of a number between 0 and 1 and result is less than the number. Conversely, if we take square root of such a number, we get the result greater than the number. Additionally, by the introduction of noise, marginal entropy of noisy vector and (therefore) joint entropies of the two vectors increase. Hence save for the MI, results of other statistics based SACs increase and this trend can be straightforwardly seen in Table 4.2.

Table 4.2. Comparison results of different SACs for the synthetic data and its noise contaminated form.

Channels	L0.5 Dist.	L1 Dist.	L1.5 Dist.	L2 Dist.	CrossEnt.	JointEnt.	M.I.	Rel.Ent.
1	0,53	0,33	0,23	0,17	4,2	5,5	2,77	0,19
2	0,53	0,33	0,23	0,17	4,14	5,43	2,74	0,17
3	0,53	0,33	0,23	0,17	4,2	5,51	2,77	0,19
4	0,52	0,32	0,22	0,16	4,19	5,49	2,77	0,18
5	0,53	0,33	0,22	0,17	4,18	5,45	2,75	0,2
6	0,52	0,32	0,22	0,16	4,16	5,48	2,77	0,19
7	0,52	0,33	0,23	0,17	4,18	5,46	2,76	0,18
8	0,52	0,32	0,22	0,17	4,22	5,5	2,78	0,2
9	0,52	0,32	0,22	0,16	4,17	5,47	2,77	0,19
10	0,52	0,32	0,22	0,17	4,2	5,49	2,77	0,19
11	0,53	0,33	0,23	0,17	4,22	5,52	2,77	0,2
12	0,52	0,32	0,21	0,16	4,14	5,45	2,72	0,18
13	0,52	0,32	0,22	0,16	4,19	5,48	2,75	0,18
14	0,53	0,33	0,23	0,17	4,17	5,48	2,76	0,18
15	0,52	0,33	0,22	0,17	4,17	5,48	2,76	0,17
16	0,52	0,32	0,21	0,16	4,19	5,49	2,77	0,17
17	0,52	0,32	0,22	0,16	4,19	5,47	2,75	0,2
18	0,52	0,32	0,22	0,16	4,17	5,47	2,77	0,17
19	0,52	0,32	0,22	0,16	4,2	5,47	2,77	0,19

Value of CE increases slightly, but the converse is true for JE and MI. Ignorable increase in CE stems from the fact that the CE method handles the two vectors separately (either in logarithm part or multiplier part) and generates results partially independent of them. The condition in which it alters is increase in the information content of at least one vector, and addition of Gaussian noise does not change the content since the original one is also a AR process for Gaussian noise. For this case, the more sensitive are JE and MI in comparison with others.

In Table 4.3., we present the results for the addition of bias having magnitude 0.1 to the original data set and we see the signal in Figure 4.4 for the first trial of the first channel. At first sight, we see that there is no distinction for the statistics based SACs in results given in Table 4.1 and Table 4.3.

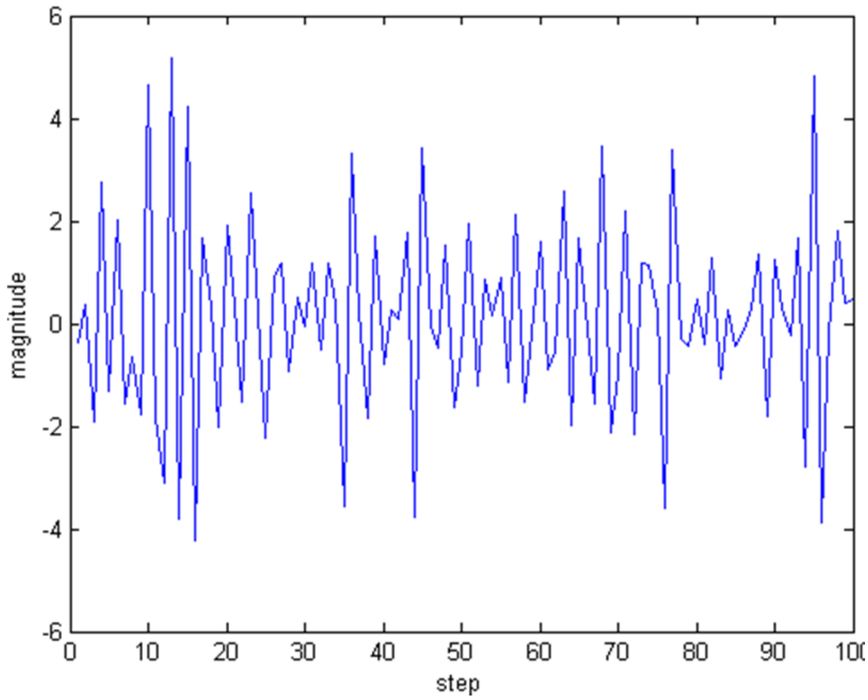

Figure 4.4. First trial of the first channel of the biased synthetic data.

Table 4.3. Comparison results of different SACs for the synthetic data and its bias added form.

Channels	L0.5 Dist.	L1 Dist.	L1.5 Dist.	L2 Dist.	CrossEnt.	JointEnt.	M.I.	Rel.Ent.
1	0,32	0,1	0,03	0,01	4,14	4,14	4,14	0
2	0,32	0,1	0,03	0,01	4,1	4,1	4,1	0
3	0,32	0,1	0,03	0,01	4,14	4,14	4,14	0
4	0,32	0,1	0,03	0,01	4,14	4,14	4,14	0
5	0,32	0,1	0,03	0,01	4,11	4,11	4,11	0
6	0,32	0,1	0,03	0,01	4,13	4,13	4,13	0
7	0,32	0,1	0,03	0,01	4,11	4,11	4,11	0
8	0,32	0,1	0,03	0,01	4,14	4,14	4,14	0
9	0,32	0,1	0,03	0,01	4,13	4,13	4,13	0
10	0,32	0,1	0,03	0,01	4,14	4,14	4,14	0
11	0,32	0,1	0,03	0,01	4,15	4,15	4,15	0
12	0,32	0,1	0,03	0,01	4,09	4,09	4,09	0
13	0,32	0,1	0,03	0,01	4,12	4,12	4,12	0
14	0,32	0,1	0,03	0,01	4,12	4,12	4,12	0
15	0,32	0,1	0,03	0,01	4,12	4,12	4,12	0
16	0,32	0,1	0,03	0,01	4,14	4,14	4,14	0
17	0,32	0,1	0,03	0,01	4,11	4,11	4,11	0
18	0,32	0,1	0,03	0,01	4,12	4,12	4,12	0
19	0,32	0,1	0,03	0,01	4,13	4,13	4,13	0

However, while the norm degree rises, distances in the norms decreases and this is consistent with the expectation. Addition of bias does not change the statistics of the data and do the same any assessment method based on it. This is a very desirable feature because by subtraction of the estimated bias from the biased data, it is probable to recover the signal. Nevertheless, it is also obvious that relying only upon the statistics based SACs remains misleading in the presence of bias due to unchanged statistics. Fortunately, this issue can be overcome by simultaneously considering distance of data vectors in any norm given. If there is a distance between two data vectors but no statistical change, this signifies the existence of bias and this bias can be eliminated by subtraction.

In Figure 4.5, we see the two times amplified form of the synthetic signal and the results are presented in Table 4.4. Again since the magnification does not change statistics of data vectors, results of CE, JE, MI and RE remains the same as if the original signal is compared with itself. On the hand, magnification causes the data vector to change and this change results in a distance between the original signal and its magnified form.

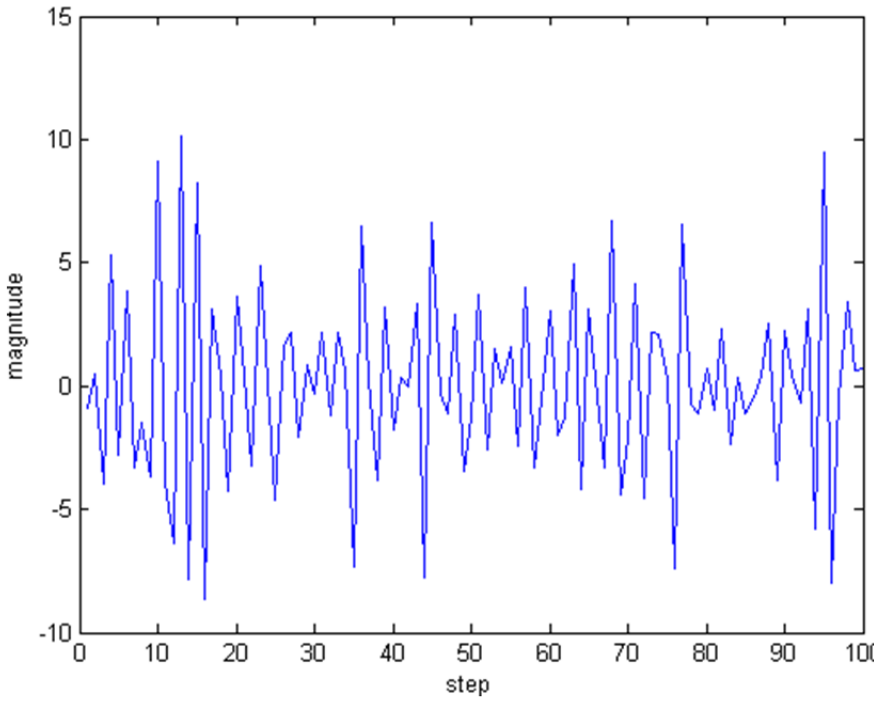

Figure 4.5. First trial of the first channel of the magnified synthetic data.

Thus, vectoral distances for different degrees become different from zero. Additionally, magnification of the vector causes the difference to be large and therefore as the norm degree rises, the distances also increase. We see this trend in Table 4.4 manifestly. Considering this, we can assert that in magnification case, sole use of statistics based SACs may be misleading and use of one of the intensity based SAC would be beneficial. Further, in the magnification, most sensitive one is L2 norm distance and the result is proportional to the norm degree.

Table 4.4. Comparison results of different SACs for the synthetic data and its two times magnified form.

Channels	L0.5 Dist.	L1 Dist.	L1.5 Dist.	L2 Dist.	CrossEnt.	JointEnt.	M.I.	Rel.Ent.
1	1,26	1,87	3,09	5,49	4,14	4,14	4,14	0
2	1,25	1,86	3,07	5,47	4,1	4,1	4,1	0
3	1,25	1,86	3,05	5,39	4,14	4,14	4,14	0
4	1,25	1,83	2,97	5,18	4,14	4,14	4,14	0
5	1,24	1,82	2,96	5,19	4,11	4,11	4,11	0
6	1,24	1,82	2,94	5,14	4,13	4,13	4,13	0
7	1,24	1,82	2,97	5,23	4,11	4,11	4,11	0
8	1,25	1,82	2,96	5,18	4,14	4,14	4,14	0
9	1,25	1,83	3	5,28	4,13	4,13	4,13	0
10	1,24	1,83	2,98	5,25	4,14	4,14	4,14	0
11	1,26	1,86	3,04	5,37	4,15	4,15	4,15	0
12	1,23	1,78	2,89	5,04	4,09	4,09	4,09	0
13	1,24	1,8	2,91	5,08	4,12	4,12	4,12	0
14	1,25	1,86	3,05	5,41	4,12	4,12	4,12	0
15	1,25	1,84	3,02	5,36	4,12	4,12	4,12	0
16	1,24	1,82	2,96	5,19	4,14	4,14	4,14	0
17	1,24	1,8	2,92	5,11	4,11	4,11	4,11	0
18	1,24	1,82	2,97	5,21	4,12	4,12	4,12	0
19	1,25	1,83	2,98	5,22	4,13	4,13	4,13	0

In other probable case, the data clusters are contaminated by some sorts of spiky noise and we obtained the contaminated data clusters only adding two spikes to the 25^{th} and 75^{th} elements. Magnitudes of spikes are equal and 2.5. In this case, we expect slight variations in statistics of the data because we only perturb 2 elements out of 100 and this trend can be seen in Table 4.5 for the statistics based SACs. In Figure 4.6, we also see the distorted signal and it is almost similar to the original one.

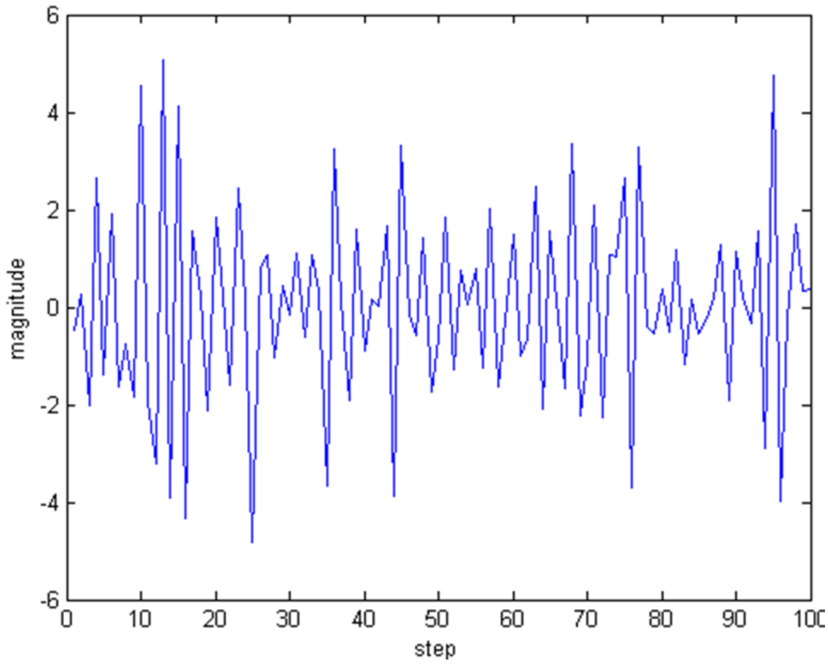

Figure 4.6. First trial of the first channel of the spiky synthetic data.

Table 4.5. Comparison results of different SACs for the synthetic data and its spike added form.

Channels	L0.5 Dist.	L1 Dist.	L1.5 Dist.	L2 Dist.	CrossEnt.	JointEnt.	M.I.	Rel.Ent.
1	0,03	0,05	0,08	0,13	4,11	4,43	3,82	0,08
2	0,03	0,05	0,08	0,13	4,08	4,32	3,87	0,06
3	0,03	0,05	0,08	0,13	4,15	4,36	3,92	0,07
4	0,03	0,05	0,08	0,13	4,15	4,38	3,9	0,07
5	0,03	0,05	0,08	0,13	4,13	4,36	3,84	0,07
6	0,03	0,05	0,08	0,13	4,1	4,37	3,87	0,06
7	0,03	0,05	0,08	0,13	4,08	4,35	3,85	0,07
8	0,03	0,05	0,08	0,13	4,13	4,39	3,87	0,07
9	0,03	0,05	0,08	0,13	4,08	4,38	3,84	0,07
10	0,03	0,05	0,08	0,13	4,15	4,4	3,85	0,09
11	0,03	0,05	0,08	0,13	4,15	4,38	3,9	0,07
12	0,03	0,05	0,08	0,13	4,07	4,3	3,86	0,05
13	0,03	0,05	0,08	0,13	4,12	4,4	3,82	0,1
14	0,03	0,05	0,08	0,13	4,1	4,36	3,88	0,07
15	0,03	0,05	0,08	0,13	4,15	4,34	3,9	0,08
16	0,03	0,05	0,08	0,13	4,16	4,36	3,92	0,06
17	0,03	0,05	0,08	0,13	4,08	4,35	3,84	0,09
18	0,03	0,05	0,08	0,13	4,12	4,32	3,91	0,07
19	0,03	0,05	0,08	0,13	4,14	4,3	3,96	0,04

On the other hand, since magnitudes of the spikes are greater than one, expected distances in different norms are proportional to the norm degree and this is what is present in the Table 4.5; as the norm degree rises, distances for the norm given also increase.

In the other scenario, we imitated the insufficiency of notch filter which is used to suppress power grid noise and added the power grid noise given in Figure 4.7 to the data set. The power grid electricity has frequency values 50Hz (in most of the world) and 60Hz (in especially the USA) and we chose the 50Hz signal pocket. In this case, both statistics and intensities of the data clusters alter considerably and this can be seen in the Table 4.6. Since the noise involves element greater than 1 in magnitude, as the norm degree rises, distances found in that norm also go higher.

Table 4.6. Comparison results of different SACs for the synthetic data and its power grid distorted form.

Channels	L0.5 Dist.	L1 Dist.	L1.5 Dist.	L2 Dist.	CrossEnt.	JointEnt.	M.I.	Rel.Ent.
1	1,05	1,26	1,57	2	4,23	6,06	2,22	0,28
2	1,05	1,26	1,57	2	4,22	6,04	2,18	0,31
3	1,05	1,26	1,57	2	4,23	6,05	2,24	0,27
4	1,05	1,26	1,57	2	4,28	6,06	2,23	0,3
5	1,05	1,26	1,57	2	4,23	6,05	2,2	0,28
6	1,05	1,26	1,57	2	4,27	6,07	2,22	0,32
7	1,05	1,26	1,57	2	4,23	6,04	2,19	0,3
8	1,05	1,26	1,57	2	4,25	6,06	2,22	0,29
9	1,05	1,26	1,57	2	4,24	6,06	2,2	0,28
10	1,05	1,26	1,57	2	4,23	6,06	2,2	0,27
11	1,05	1,26	1,57	2	4,26	6,07	2,23	0,27
12	1,05	1,26	1,57	2	4,24	6,03	2,18	0,31
13	1,05	1,26	1,57	2	4,26	6,05	2,2	0,31
14	1,05	1,26	1,57	2	4,25	6,04	2,22	0,3
15	1,05	1,26	1,57	2	4,22	6,04	2,22	0,28
16	1,05	1,26	1,57	2	4,25	6,07	2,2	0,29
17	1,05	1,26	1,57	2	4,22	6,04	2,19	0,28
18	1,05	1,26	1,57	2	4,22	6,04	2,21	0,28
19	1,05	1,26	1,57	2	4,25	6,07	2,2	0,29

On the other hand, the information included by the signal increases due to addition of the noise and this can be seen from the results of the statistics based SACs. We here perceive that both JE and MI information are sensitive to such type of noise considerably which are also sensitive to additive Gaussian noise.

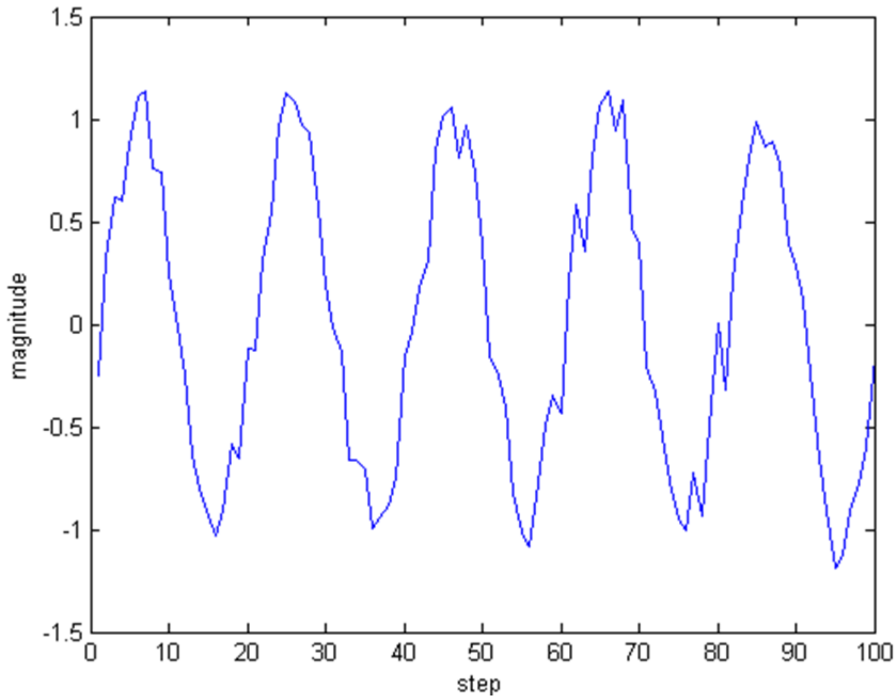

Figure 4.7. Imitated power grid noise.

In Figure 4.8, the second synthetic data cluster independent of the first one is presented and results of the SACs are given in Table 4.7. As seen, the signal is totally different from the base signal and this is reflected to the distance/dissimilarity measures. With regard to the intensity based SACs, distance measures increase with the increasing degree of the norm and the largest distance is ascertained in the L2 norm. Considering the statistics based SACs, since the two data being compared are distinct from each other and there is no relationship between them, their information measures get higher. For this case, the most sensitive ones are JE and MI. Both synthetic data clusters are generated by a Gaussian process as mentioned earlier and due to this and formulization of CE, results of CE remains unaltered. This is also true for the RI and we can easily assert that the most dependable measures are JE and MI.

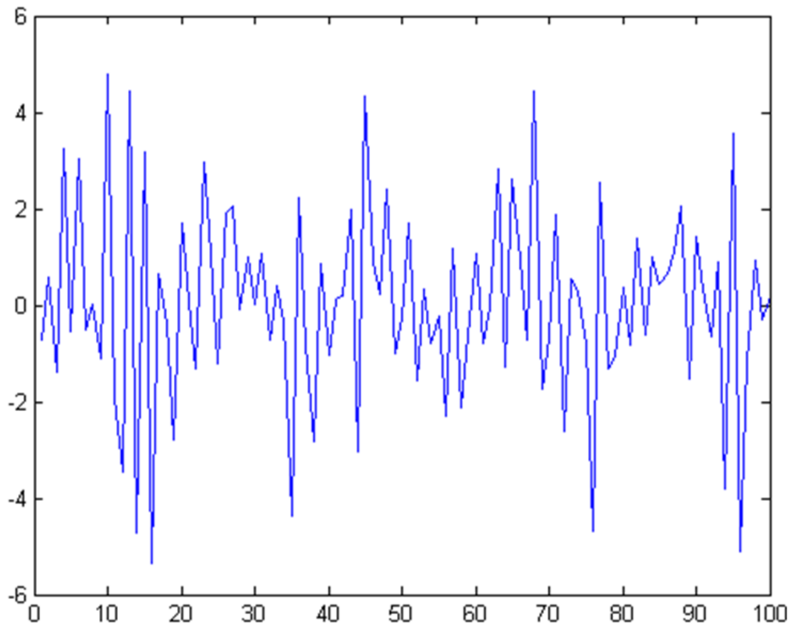

Figure 4.8. First trial of the first channel of power grid contaminated data.

Table 4.7. Comparison results of different SACs for two different synthetic data vectors.

Channels	L0.5 Dist.	L1 Dist.	L1.5 Dist.	L2 Dist.	CrossEnt.	JointEnt.	M.I.	Rel.Ent.
1	1,57	2,9	6	13,36	4,15	6,32	1,92	0,27
2	1,57	2,92	6,03	13,44	4,18	6,31	1,92	0,3
3	1,55	2,86	5,85	12,9	4,21	6,33	1,97	0,28
4	1,58	2,95	6,12	13,69	4,19	6,33	1,95	0,28
5	1,57	2,92	6,02	13,38	4,18	6,32	1,9	0,28
6	1,57	2,93	6,06	13,49	4,21	6,31	1,95	0,29
7	1,56	2,88	5,94	13,22	4,14	6,31	1,9	0,28
8	1,56	2,88	5,95	13,3	4,1	6,31	1,91	0,27
9	1,57	2,92	6,02	13,41	4,12	6,33	1,9	0,29
10	1,57	2,92	6,04	13,49	4,21	6,33	1,96	0,28
11	1,57	2,92	6,02	13,37	4,13	6,33	1,91	0,25
12	1,56	2,88	5,96	13,36	4,15	6,3	1,9	0,3
13	1,58	2,94	6,09	13,63	4,23	6,34	1,93	0,31
14	1,57	2,93	6,07	13,61	4,16	6,32	1,92	0,28
15	1,57	2,9	5,97	13,27	4,15	6,33	1,92	0,28
16	1,57	2,9	5,95	13,15	4,18	6,32	1,94	0,29
17	1,55	2,85	5,82	12,83	4,15	6,31	1,91	0,29
18	1,55	2,85	5,81	12,77	4,21	6,32	1,93	0,31
19	1,57	2,92	6,01	13,37	4,2	6,32	1,95	0,3

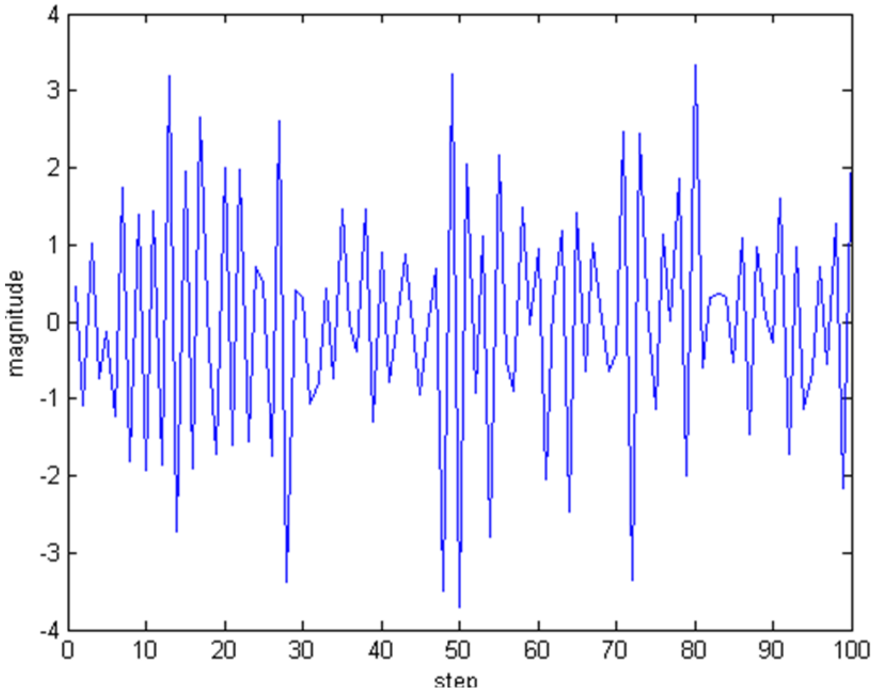

Figure 4.9. First trial of the first channel of the second synthetic data.

In the final scenario, we generated a data cluster for 19 channels using 2D AR model given by equation (4.x). This approach establishes relationships between all neighboring samples and channels, and sample length for a channel is 100.

$$y(i,j) + 0.1y(i-1,j) + 0.5y(i,j-1) - 0.05y(i-1,j-1) = w_{i\,j} \qquad (4.x)$$

Where $w_{i\,j}$ is i.i.d. normal Gaussian noise, i stands for sample index and j denotes channel index.

We see the clusters in Figure 4.10.

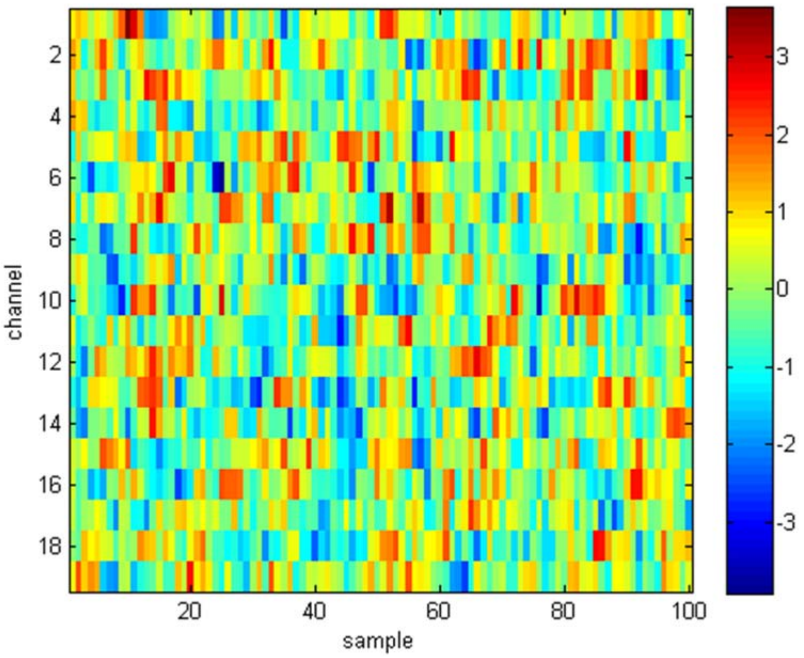

Figure 4.10. 19x100 synthetic EEG signal cluster gotten through 2D AR process.

To compare SACs' performance, we got results of SACs for the synthetic data in comparison with itself in Table 4.8, we added 15dB i.i.d. Gaussian noise to the cluster and get the noisy cluster given in Figure 4.11. Results of this scenario are presented in Table 4.9.

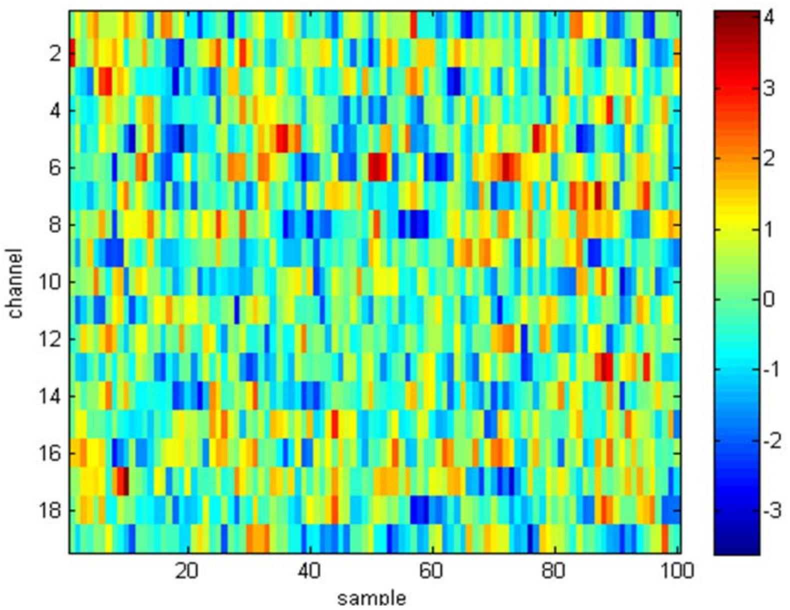

Figure 4.11. The 19x100 synthetic EEG signal cluster contaminated by 15dB i.i.d. Gaussian noise.

Table 4.8. Comparison results of different SACs for the 2D AR synthetic data and itself.

Channels	L0.5 Dist	L1 Dist	L1.5 Dist	L2 Dist	CrossEnt.	JointEnt.	M.I.	Rel.Ent.
1	0.00	0.00	0.00	0.00	3.12	3.12	3.12	0.00
2	0.00	0.00	0.00	0.00	2.92	2.92	2.92	0.00
3	0.00	0.00	0.00	0.00	2.85	2.85	2.85	0.00
4	0.00	0.00	0.00	0.00	2.98	2.98	2.98	0.00
5	0.00	0.00	0.00	0.00	2.87	2.87	2.87	0.00
6	0.00	0.00	0.00	0.00	3.08	3.08	3.08	0.00
7	0.00	0.00	0.00	0.00	3.07	3.07	3.07	0.00
8	0.00	0.00	0.00	0.00	3.17	3.17	3.17	0.00
9	0.00	0.00	0.00	0.00	3.07	3.07	3.07	0.00
10	0.00	0.00	0.00	0.00	2.86	2.86	2.86	0.00
11	0.00	0.00	0.00	0.00	3.06	3.06	3.06	0.00
12	0.00	0.00	0.00	0.00	3.04	3.04	3.04	0.00
13	0.00	0.00	0.00	0.00	2.70	2.70	2.70	0.00
14	0.00	0.00	0.00	0.00	2.92	2.92	2.92	0.00
15	0.00	0.00	0.00	0.00	2.89	2.89	2.89	0.00
16	0.00	0.00	0.00	0.00	2.96	2.96	2.96	0.00
17	0.00	0.00	0.00	0.00	2.70	2.70	2.70	0.00
18	0.00	0.00	0.00	0.00	2.96	2.96	2.96	0.00
19	0.00	0.00	0.00	0.00	3.10	3.10	3.10	0.00

As in the previous cases, all lp distances becomes zero if we compare the data with itself (Table 4.8). However, by the introduction of noise such distance measures rise. Additionally, we see the possible highest results for different statistics based SACs in Table 4.8 and the results of the SACs for the noise contaminated data in Table 4.9. Since noise distort the data mutual information and relative entropy between the original and contaminated data decreases. Further, noise augments information content of the data and this can be seen in Tables 4.8 and 4.9. The most sensitive SACs are mutual information, joint entropy and L2 distance.

Table 4.9. Comparison results of different SACs for the 2D AR synthetic data and noisy form.

Channels	L0.5 Dist.	L1 Dist.	L1.5 Dist.	L2 Dist.	CrossEnt.	JointEnt.	M.I.	Rel.Ent.
1	0.16	0.18	0.20	0.22	3.31	4.07	2.05	-0.18
2	0.14	0.16	0.18	0.20	3.03	3.94	1.90	-0.11
3	0.16	0.18	0.20	0.22	2.97	3.71	1.90	-0.13
4	0.14	0.16	0.19	0.21	3.01	4.03	1.88	-0.03
5	0.16	0.18	0.20	0.22	2.93	3.69	2.05	-0.05
6	0.14	0.16	0.17	0.19	3.11	3.73	2.39	-0.03
7	0.13	0.16	0.18	0.21	3.01	3.94	2.09	-0.01
8	0.12	0.14	0.16	0.18	3.24	4.03	2.21	-0.07
9	0.15	0.17	0.19	0.21	3.09	4.22	1.96	-0.02
10	0.14	0.17	0.19	0.21	2.92	3.91	1.92	-0.06
11	0.15	0.17	0.19	0.21	3.13	4.14	1.95	-0.08
12	0.13	0.15	0.17	0.19	3.08	4.10	1.87	-0.04
13	0.14	0.17	0.19	0.21	2.75	3.71	1.80	-0.06
14	0.14	0.17	0.19	0.22	3.00	4.09	1.87	-0.09
15	0.16	0.18	0.20	0.22	2.97	4.01	1.76	-0.08
16	0.15	0.18	0.20	0.21	3.00	4.12	1.91	-0.04
17	0.14	0.16	0.18	0.20	2.49	3.59	1.69	0.03
18	0.13	0.16	0.18	0.20	2.99	3.96	1.92	-0.03
19	0.15	0.17	0.19	0.21	3.21	4.07	1.98	-0.11

5. CONCLUSION

In this thesis, we compared several distance/dissimilarity measures used for classification EEG signal events. We initially introduced the basics in biological signals concept and then explained the properties of EEG signals.

Following that, we referred to the past studies having relationships with this study and gave explanations for them. Whether they are on intensity or statistics or other approaches based SACs they may give precious insight to the reader and in this regard, they are strongly recommended to be read.

In the third section, we gave the basics of vectors and distance measurement norms for different values and tried to be as elaborative as we could by giving many graphs for different cases. Additionally, we followed the same way in setting forth the fundamentals of statistics based SACs and bolstered explanation by giving several examples for different cases. The SACs were L0.5, L1, L1.5 and L2 norms for the intensity based and cross entropy, joint entropy, mutual information and relative entropy for the statistics based.

In the following section, we presented a multitude of results for several scenarios and got able to develop an insight of the SACs and their behavior. In the presence of Gaussian type noise having small variance (15dB SNR), addition of small random values to the signal is done and such addition happens for the all of data segments. Therefore, distance measures in lower norm degree become higher. However, by the increasing variance of the noise, this trend is easy to be altered. Considering, the statistics based SACs, since addition of Gaussian type noise causes information content to be higher, entropies get higher and this is seen in the results.

In the other scenario, we added a small bias to all elements in the signal cluster and obtained results of all SACs for all data channels. As explained previously, addition of a bias does not alter statistics of the signal and it is normal to see no change in the statistics based SACs. However, such an addition generates a distance between the two vectors in any norm and result has an inversely proportional trend with respect to the norm degree.

We also magnified the signal two times and compared it with itself to examine the responses of the SACs to multiplication. Since magnification or fading do not change structural or statistics of data, information based SACs remained unchanged and generated results as if we compare two equal signals. This is similar to the addition of bias and in some cases may be misleading. To overcome such a drawback, attention on the distance measures also recommended.

We also perturbed the signal by adding two spikes having magnitude 2.5 to 25^{th} and 75^{th} elements of the signal. In this case, both statistics and distance of data vector change. Since magnitudes of spikes are 2.5 and exceed 1, distances in larger norms are bigger than the others. Further, as changing two elements out of a hundred slightly affects statistics of the data, the SACs based on statistics do not generate considerably different values and this is case presented.

In the penultimate scenario, we tried to imitate power grid noise and added it to the signal. Since the noise contains considerable information, it also augments information of the signal after addition and both intensity and statistics based SACs generate results in harmony with this; MI decreases and all others increase.

We also compared two data sets totally distinct from each other. In this case, both types of SACs signify a big difference between the two compared data sets by giving high dissimilarity measures.

Considering all, we can assert that in case of big noise addition both statistics and intensity (especially for higher degrees) based SACs are sensitive to the presence of noise and in case of small but prevalent noise, intensity based with lower degrees becomes more sensitive. Due to their intrinsic properties, both CE and RI remain inferior to MI and JE considering statistics SACs and in some cases mentioned afore, vector distance measures may be misleading. Therefore, both JE and MI can be assuredly recommended for all cases. However, simultaneous use of them with a distance measure is also recommended because statistics based SACs are insensitive to the presence of bias, magnification and fading.

REFERENCES

AUBIN, J. P., 1999. Applied Functional Analysis. JOHN WILEY and SONS, ISBN 0471-17976-0.

BAILLET, S., GARNERO, L., MARIN, G., and HUGONIN, J. P., 1999. Combined MEG and EEG Source Imaging by Minimization of Mutual Information. IEEE Transactions on Biomedical Engineering, v.46, p.522-534.

BRONZINO, J. D., 1995. Principles of Electroencephalography. In: J.D. Bronzino ed. The Biomedical Engeneering Handbook, pp. 201-212, CRC Press, Florida.

GRAMFORT, A. and KOWALSKİ, M., 2009. Improving M/EEG Source Localization with an Inter-Condition Sparse Prior. International Symposium on Biomedical Imaging, IEEE, p.141-144.

HILL, D. L. G., BATCHELOR, P. G., HOLDEN, M. and HAWKES, D. J., 2001. Medical image registration, Physics in Medicine and Biology, V.46, R1-R45.

JASPER, H.H.. 1958. The ten-twenty electrode system of the International Federation. Electroencephalography and Clinical Neurophysiology, p.371-375.

LAN, T., ERDOGMUS, D., ADAMI, A. and PAVEL, M., 2005. Feature selection by independent component analysis and mutual information maximization in EEG signal classification. International Symposium on Neural Networks, pp. 3011-3016.

LI, D. H. and FUKUSHIMA, M., 2001. A modified BFGS method and its global convergence in nonconvex minimization, J. Comp. Appl. Math., vol. 129, pp. 15–35.

PICTON, T.W., et. al. 2000. Guidelines for using event-related potentials to study cognition: Recording standards and publication criteria. Psychophysiology, vol.37, p.127-152.

SHANNON C. E., 1948. The mathematical theory of communication (parts 1 and 2)Bell Syst. Tech. J.27379–423, 623–56

SHANNON C. E., 1949. Communication in the presence of noise. Proc. IRE3710–21 (reprinted in Proc. IEEE86447–57).

STRANG, G., 2009. Linear Algebra. 4th Edition. Thomson Learning Inc. ISBN-13: 860-1200456875.

TEPLAN, M., 2002. Fundamentals of EEG Measurement. Measurement Science Review, Vol.2, Section.2.

VUGT, M. K.,SEDERBERG, P. B. and KAHANA, M. J., 2007. Comparison of spectral analysis methods for characterizing brain oscillations. Journal of Neuroscience Methods. V.162, p.49-63.

XU, P., TIAN, Y., CHEN, H., and YAO, D., 2007. Lp Norm Iterative Sparse Solution for EEG Source Localization. IEEE Transaction On Biomedical Engineering, V.54, p.400-409.

WENDLING F,2009, ANSARI-ASL K, BARTOLOMEİ F, SENHADJIL,2009. From EEG signals to connectivity: A model-based evaluation of interdependence measures. Journal of Neuroscience Methods. V.183, p.9-18.

ARNOLD J, GRASSBERGER P, LEHNERTZ K, ELGER C. E, 1999. A robust method for interdependences: Application to intracranially recorded EEG. Physica D, V.134,p.419-430.

LI XIAOLI, OUYANG GAOXIANG,2010. Estimating coupling direction between neuronal populations with permutation conditional mutual information. Neurolmage, V.52, p.497-507.

OUYANG GAOXIANG, DANG C, RICHARDS D.A, XIAOLI L, 2010. Ordinal pattern based similarity analysis for EEG recordings. Clinical Neurophysiology, V.121, p.694-703.

BIOGRAPHY

Erçin ÖZCAN was born in Kahramanmaraş, Turkey in 1985. He received her B.S. degree in Electrical and Electronics Engineering Department from Mersin University in 2010. After completion his B.S. education, he started MSc education in Electrical and Electronics Engineering Department in Çukurova University. He has been working as a Research Assistant in Electrical and Electronics Engineering Department of the ÇukurovaUniversity since 2011. His research areas are Signal Processing and Telecommunication Techniques.

APPENDIX

```
% Lp metrics
function d = lpdist(x, y,p)
n = length(x);
d = (sum(abs(x - y).^p)/n);%%^(1/p);
```

```
function info = mutualinfo(x, y, s)

[Px, Py, Pxy] = prbdst(x, y, s);
P = Px.' * Py;
m = (Pxy > 0)&(P > 0);
K = zeros(size(Pxy));
K(m) = Pxy(m) .* log2(Pxy(m) ./ P(m));
info = sum(sum(K));
```

```
function [k l]=All(E1,E2,ch,p)

sz=size(E1);

k=zeros(8,sz(3));%%rows for distance measure, columns for eeg channels
%%row1:Lp norm, row2:eucledian dist, row3:chebyshev, row4:cross entropy,
%%row5:joint entropy, row6:mutual info, row7:relative entropy
for j=1:sz(3)
   for i=1:sz(2)
      k(1,i,j)=lpdist(E1(:,i,j),E2(:,i,j),p);%% p is  norm value and to be entered
during implementation
      k(2,i,j)=lpdist(E1(:,i,j),E2(:,i,j),1);%% p=1 for chebyshev distance
      k(3,i,j)=lpdist(E1(:,i,j),E2(:,i,j),1.5);%% p=1 for chebyshev distance
      k(4,i,j)=lpdist(E1(:,i,j),E2(:,i,j),2);%% p=2 for euclidian distance
      k(5,i,j)=crosse(E1(:,i,j),E2(:,i,j),numel(E1(:,i,j))/4);%% cross entropy
      k(6,i,j)=jointe(E1(:,i,j),E2(:,i,j),numel(E1(:,i,j))/4);%% joint entropy
      k(7,i,j)=mutualinfo(E1(:,i,j),E2(:,i,j),numel(E1(:,i,j))/4);%% mutual info
      k(8,i,j)=relativee(E1(:,i,j),E2(:,i,j),numel(E1(:,i,j))/4);%% relative entropy
   end

   l(1,j)=sum(sum(k(1,:,j)));
   l(2,j)=sum(sum(k(2,:,j)));
   l(3,j)=sum(sum(k(3,:,j)));
   l(4,j)=sum(sum(k(4,:,j)));
   l(5,j)=sum(sum(k(5,:,j)));
   l(6,j)=sum(sum(k(6,:,j)));
   l(7,j)=sum(sum(k(7,:,j)));
   l(8,j)=sum(sum(k(8,:,j)));
end
l=l'/100;
end
```

```
function c = crosse(x, y, s)

[Px, Py] = prbdst(x, y, s);
n = (Px > 0)&(Py > 0);

K = zeros(size(Px));
K(n) = Px(n) .* log2(Py(n));
c = - sum(K); % cross entropy
```

```
function x = ensemble(n, m, mu, Sigma)
% ---
if nargin == 0
    n = 100;
    m = 2;
    mu = [1 2];
    Sigma = [4 0.6; 0.6 1];
end
z = randn(n, m);

zbar = repmat(mean(z),n,1);
S = cov(z);
z0 = (z-zbar) / chol(S); % zero mean, identity cov
x = z0 * chol(Sigma) + repmat(mu,n,1); % specified mean and cov
```

```
% Euclidean Distance

% Interval data. Euclidean distance, squared Euclidean distance,
Chebychev, block, Minkowski, or customized.
% Count data. Chi-square measure or phi-square measure.
% Binary data. Euclidean distance, squared Euclidean distance, size
difference, pattern difference, variance, shape, or Lance and Williams.

% Correlation coefficient

% Joint Entropy
% Entropy
% Conditional Entropy
% Mutual Information

% Kullback-Leibler divergence (relative entropy)

% t-test


% ARTIFICAL SIGNALS
%

p = 10;

% Distinct frequencies
deuc = zeros(1, 3);
dlp = zeros(1, 3);
dcheb = zeros(1, 3);
w = wnd(1, 32);
for k = 1:3;
```

```
[x, y] = asignals(k);
deuc(k) = eucdist(x.*w, y.*w);
dlp(k) = lpdist(x.*w, y.*w, p);
dcheb(k) = chebdist(x.*w, y.*w);
end

[x, y] = asignals(3);
s = 8;
[Px, Py, Pxy, binsx, binsy] = prbdst(x, y, s);
figure(1); plot(0:length(x)-1, x, 'b', 0:length(y)-1, y, 'r')
figure(2);
subplot(221); bar(binsx, Px, 0.4); axis([binsx(1) - 0.25, binsx(end) + 0.25, -
0.25, 1.25]);
xlabel('x'); ylabel('Px')
subplot(223); bar(binsy, Py, 0.4); axis([binsy(1) - 0.25, binsy(end) + 0.25, -
0.25, 1.25]); xlabel('y'); ylabel('Py')
subplot(222); imagesc(binsy, binsx, Pxy); axis xy
xlabel('y'); ylabel('x'); title('Pxy')
colormap(gray);

[x, y] = asignals(3);
s = 8;
H = jointe(x, y, s);
info = mutualinfo(x, y, s);
r = relativee(x, y, s) + relativee(y, x, s);
c = crosse(x, y, s);

%
n = 128;
m = 19;
mu = zeros(1, m);
```

```
rho = 0.9;
r = rho.^(0:m-1);
Sigma = toeplitz(r);
x = ensemble(n, m, mu, Sigma);
B = extend(x, 7);
```

```
function E1 = gencone(n,m,t)

bcol = [1, 4/3*cos(pi/3), 4/9];
brow = [1, -3/2*cos(pi/8), 9/16];

E1 = zeros(n, m, t);
for k = 1:t
x = randn(n, m);
E1(:,:,k) = filter(brow, 1, x);
E1(:,:,k) = filter(bcol, 1, E1(:,:,k)')';
end
```

```
function E2 = genctwo(n,m,t)

bcol = [1, -5/3*cos(pi/5), 25/36];
brow = [1, -1.8*cos(pi/3), 0.81];

E2 = zeros(n, m, t);
for k = 1:t
x = randn(n, m);
E2(:,:,k) = filter(brow, 1, x);
E2(:,:,k) = filter(bcol, 1, E2(:,:,k)')';
end
```

```
function H = jointe(x, y, s)
[~, ~, Pxy] = prbdst(x, y, s);
m = Pxy > 0;
K = zeros(size(Pxy));
K(m) = Pxy(m) .* log2(Pxy(m));
H = - sum(sum(K));
```

```
function E2=noise(E1,SNR)

sz=size(E1);

for k=1:sz(3)
   E2(:,:,k)=awgn(E1(:,:,k),SNR,'measured');
end
```

```
function [Px, Py, Pxy, binsx, binsy] = prbdst(x, y, s)

Maxx = max(x);
Maxy = max(y);
Minx = min(x);
Miny = min(y);


Dx = (Maxx-Minx)/s;
Dy = (Maxy-Miny)/s;

Ex = Minx : Dx : Maxx;
Ey = Miny : Dy : Maxy;

binsx = Ex(1:end-1);
binsy = Ey(1:end-1);

h = zeros(s, s);

for k = 1:s
   t = (Ex(k) <= x)&(x < Ex(k + 1));
   for n = 1:s
      v = (Ey(n) <= y)&(y < Ey(n + 1));
      h(k, n) =  sum(double(t & v));
   end
end

Pxy = h / sum(sum(h));

hx = zeros(1, s);
for k = 1:s
```

```
    t = (Ex(k) <= x)&(x < Ex(k + 1));
    hx(k) =  sum(double(t));
end

hy = zeros(1, s);
for k = 1:s
    t = (Ey(k) <= y)&(y < Ey(k + 1));
    hy(k) =  sum(double(t));
end

Px = hx / sum(hx);
Py = hy / sum(hy);



function y = tdaoone(u, v)
% 2-D auto-regressive process
% u : pole  at  vertical direction
% v : pole  at  vhorizontal direction
syms zx zy
% zeros of the characteristic polynominal is inside the unit unit circle
p = expand((1 - u*zx)*(1 - v*zy));
a = coeffs(p);
b = 1;
y = zeros(20, 101);
for n = 2:20
for m = 2:101
    y(n, m) = - a(2)*y(n - 1, m) - a(3)*y(n, m - 1)  - a(4)*y(n - 1, m - 1) +
b*randn;
end
end
y = y(2:end,2:end);
```

```
function y = vaoone(a, b)
% Vector auto-regressive process
% a : standart deviation
% b : mean
lambda = rand(1, 19);
D = diag(lambda);

K = a*randn(19, 100) + b;
[V, ~] = eig(K*K.');
% eigen-values of coefficient matrix is in [0, 1]
A = V*D*V.';

y = zeros(19, 101);
for k = 2:101
    y(:, k) =  A*y(:, k - 1) + randn(19, 1);
end
y = y(:, 2:end);
```

Publication Arose Form This Thesis Work