UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

GUILHERME MARON

Analysis and Classification of Spatial Cognition using Non-Linear Analysis and Artificial Neural Networks

Thesis presented in partial fulfillment
of the requirements for the degree of
Master of Computer Science

Prof. Dr. Dante Augusto Couto Barone
Advisor

Porto Alegre, January 2014

CIP – CATALOGING-IN-PUBLICATION

Maron, Guilherme

Analysis and Classification of Spatial Cognition using Non-Linear Analysis and Artificial Neural Networks / Guilherme Maron. – Porto Alegre: PPGC da UFRGS, 2014.

101 f.: il.

Thesis (Master) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR–RS, 2014. Advisor: Dante Augusto Couto Barone.

1. Non-linear analysis. 2. Lyapunov exponents. 3. Signal processing. 4. Biological signal processing. 5. EEG analysis. 6. Cognitive profiles. 7. Spatial cognition. 8. Classification algorithms. 9. Artificial neural networks. 10. MLP. I. Barone, Dante Augusto Couto. II. Título.

CONTENTS

LIST OF ABBREVIATIONS AND ACRONYMS

AUC	Area Under Curve
ANN	Artificial Neural Networks
BIC	Bayesian Information Criteria
BPR-5	Battery of reasoning tests
BPR-5 SR	Spatial Reasoning sub-test of BPR-5
CARVI	Vineyards Region Campus (Campus da Região dos Vinhedos)
CD	Correlation Dimension
CENT	Center of Exact, Nature and Technology Sciences (Centro de Ciências Exatas, da Natureza e Tecnológicas)
EEG	Electroencephalogram
ERP	Event-Related Potential
FD	Fractal Dimension
FFT	Fast Fourier Transform
FNN	False Nearest Neighbor [Method]
FT	Fourier Transform
H	Hurst Exponent
KNN	K-Nearest Neighbors
KS	Kolmogorov-Sinai Entropy
LE	Lyapunov Exponent
LLE	Largest Lyapunov Exponent
MLP	Multi-Layer Perceptron ANN
PPGC	Post-Graduate Program in Computer Science (Programa de Pós-Graduação em Computação)
RBF	Radial Base Function ANN
SI-	Group of individuals with any alleged higher degree of development of their spatial cognition
SI+	Group of individuals that have an alleged higher degree of development of their spatial cognition

SVM	Support Vector Machine
UCS	Caxias do Sul University (Universidade de Caxias do Sul)
UFRGS	Federal University of Rio Grande do Sul (Universidade Federal do Rio Grande do Sul)
VP	Voted Perceptron ANN

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The main objective of the present work is to propose, develop, test, and show a method for classifying the spatial cognition degree of development on different individuals. Thirty-Seven undergraduate students had their electroencephalogram (EEG) recorded while engaged in 3-D images mental rotation tasks. Their spatial cognition degree of development was evaluated using a BPR-5 psychological test. The Largest Lyapunov Exponent (LLE) was calculated from each of the 8 electrodes recorded in each EEG. The LLEs were used as input for five different classifiers: i) multi-layer perceptron artificial neural network, ii) radial base functions artificial neural network, iii) voted perceptron artificial neural network, iv) support vector machines, and v) K-Nearest Neighbors. The best result was achieved by using a RBF with 4 clusters and Puk kernel function.

Also a statistical analysis of the brain activity, based in the calculated LLEs, differences between two interest groups: SI+ (participants with an alleged higher degree of development of their spatial cognition) and SI- (control group) during the performing of mental rotation of tridimensional images tasks was done.. An average difference of 16% was found between both groups.

The proposed classification method can contribute and interact with other processes in the analysis and study of human spatial cognition, as in the understanding of the human intelligence at all. A better understanding and evaluation of the cognitive capabilities of an individual could suggest him elements of motivation, ease or natural inclinations, possibly affecting the decisions of his life and carrier positively.

Keywords: Non-linear analysis, Lyapunov exponents, signal processing, biological signal processing, EEG analysis, cognitive profiles, spatial cognition, classification algorithms, artificial neural networks, MLP.

Análise e classificação da capacidade cognitiva espacial utilizando técnicas de análise não-linear e redes neurais artificiais

RESUMO

O principal objetivo do presente trabalho é propor, desenvolver, testar e apresentar um método para a classificação do grau de desenvolvimento da capacidade cognitiva espacial de diferentes indivíduos. 37 alunos de graduação tiveram seus eletroencefalogramas (EEGs) capturados enquanto estavam engajados em tarefas de rotação mental de imagens tridimensionais. Seu grau de desenvolvimento da capacidade cognitiva espacial foi avaliado utilizando-se um teste psicológico BPR-5. O maior expoente de Lyapunov (LLE) foi calculado a partir de cada um dos 8 canais dos EEGs capturados. OS LLEs foram então utilizados como tuplas de entrada para 5 diferentes classificadores: i) perceptron de múltiplas camadas, ii) rede neural artificial de funções de base radial, iii) perceptron votado, iv) máquinas de vetor de suporte, e v) k-vizinhos. O melhor resultado foi obtido utilizando-se uma RBF com 4 clusters e a função de kernel Puk.

Também foi realizada uma análise estatística das diferenças de atividade cerebral, baseando-se nos LLEs calculados, entre os dois grupos de interesse: SI+ (indivíduos com um suposto maior grau de desenvolvimento da sua capacidade cognitiva espacial) e SI- (grupo de controle) durante a realização de tarefas de rotação mental de imagens tridimensionais. Uma diferença média de 16% foi encontrada entre os dois grupos.

O método de classificação proposto pode vir a contribuir e a interagir com outros processos na analise e no estudo da capacidade cognitiva espacial humana, assim como no entendimento da inteligência humana como um todo. Um melhor entendimento e avaliação das capacidades cognitivas de um indivíduo podem sugerir a este elementos de motivação, facilidade ou de inclinações naturais suas, podendo, provavelmente, afetar as decisões da sua vida e carreira de uma forma positiva.

Palavras-chave: análise não-linear, expoentes de Lyaunov, processamento de sinais, processamento de sinais biológicos, análise de EEGs, perfis cognitivos, perfil cognitivo espacial, algoritmos de classificação, redes neurais artificiais, MLP, RBF, SVM, teoria do caos, eletroencefalografia.

1 INTRODUCTION

Several different reasons can lead to a slower development of the sciences at all. Monetary resources can be scarce. Lack of human resources in the research departments is a problem too, mainly in countries where the investment in Research and Development are low. Ethical problems (not implying if it is bad or not) and the bureaucracy related to the permissions to execute some studies are barriers too. Even if available, accessible technologies can halt the rhythm of science. But sometimes science itself limits its development speed. It happens when there are not available theories and adequate tools for the analysis of some problems or phenomena.

The problem of lacking the right tools for analyzing a problem was a major issue even for Isaac Newton the father of Modern Science who needed to develop the infinitesimal Calculus [1] a fit tool for manipulating the variation of mathematical values to perform his studies over Classical Physics, involving moving bodies. Other of his large contributions for science is the Newton's Universal Gravitation Law. But these studies needed to wait for like 20 years to become possible because of the lacking of scientific tools. The calculation of the exact radius of Earth made by Jean Picard in 1671 was the lacking piece for Newton's Universal Gravitation Law puzzle, published in 1687 (MARTINS, 2012; BAKER, 2007; CRILLY, 2008).

Even with the contributions from Newton, the study of Celestial Mechanics was far away from being a dry area. Newton's Universal Gravitation Law addresses the problem of the gravitational interaction between two bodies. But it is widely known that this is a hypothetical scenario that is far away from reality. Even here in our own Solar System we cannot calculate the gravitational interaction between Earth and Sun without taking the Moon into account. Then another problem is faced the Three-Bodies Problem (or $n-$bodies problem, with $n > 2$). This problem was not correctly tackled by the method used by the researchers of that time, the Perturbation Theory [2]. That was so important that King Oscar II from Sweden made a challenge for the scientists of the world, in the event of his 60th birthday: to solve 4 different scientific problems, being one of them the 3-bodies problem (CFTCS, 2012).

Henry Poincare earned the offered prize, even without completely solving the $3-$bodies problem, for his new approach: a qualitative approach, unlike the used to date, the Per-

[1] Infinitesimal Calculus is the techniques of Integral Calculus and Different Calculus combined. Isaac Newton and Gottfried Leibniz made the same conclusions, almost at the same time, but trailing different ways for it.

[2] Using the Perturbation Theory to solve the $n-$bodies problem consisted in treating the interaction between the celestial bodies as perturbations in the current state of the system, and using that new state as input for a new perturbation. The weak point of this technique is that it worked for very limited intervals of time, and was impossible to infer over future states of the system.

turbation Theory, a numeric approach. Poincare's approach consisted in using a plane of lower dimension than the original system a bidimensional plane in the $3-$bodies problem, which is tridimensional and observing when an object crosses that plan, registering the point in which it happened. In the case in which an object crosses 2 or more times the plane at the same point, a periodic orbit can be assumed for this object. This technique, showed in figure 1.1 (CFTCS, 2012) is known as Poincare Section, and is used up today in dimensionality reduction problems. One interesting result of this approach was the discovery that the bodies sometimes do not reached a stable state, moving, for instance, in periodic orbits, but in orbits very close of it, i.e. quasi-periodic orbits. However the most interesting behavior found was that even very small variations in the initial conditions caused a large difference in the system state in a short time. Poincare has discovered one of the major characteristics of a chaotic system, the Sensitivity [3], in the $3-$bodies problem.

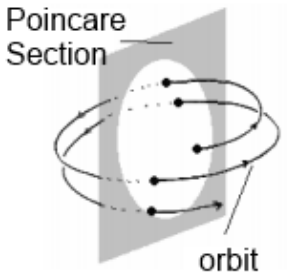

Figure 1.1: Poincare Section in the $3-$bodies problem.

Although, even with Poincare's findings, the science was not able to deal with Chaos, and many other Dynamical Systems [4] that show characteristics similar to the ones in the $3-$ bodies problem, and could not be easily modeled as a set of linear difference equations. These dynamical systems where treated as noisy, or spurious cases, even being the majority of the cases in natural systems (GLEICK, 1989). This approach kept being used up to 1963, when Edward Lorenz faced a problem that qualitatively resembled the behavior found in the $3-$bodies problem, but in a completely different domain: meteorological simulation (GLEICK, 1989).

Lorenz created a set of equations to try mimicking metereological behavior. It is assumed that he was running his simulations and for some reason had to stop them. But he decided to not restart it from scratch, but using an intermediary point of the simulation, using the variable's values he printed before stopping the simulation. But after coming back from a cup of coffee, as it is said, he got a big surprise: the state of the simulation was completely different from the expected one (as we can see in picture 1.2. His first reaction was to check if he did not inputted wrong values for the new simulation, and after several revisions he assumed that the inputs were indeed right. After a lot of thinking he concluded that the problem was the decimal precision of the variables, his simulator used $6-$ digit decimal precision, but the output was printed only using $3-$ digit decimal

[3] Sensitivity is an inherent characteristic of chaotic systems. Systems that show a high sensibility to its initial state, showing large (exponential) deviations in future states even with little perturbation on its initial state.

[4] Dynamical Systems are systems that evolve in time in the form $X_{t+1} = f(X_t)$, where f is a set of rules, usually differential equations that model the system state transition.

precision. But the question he did was, **how can be a very little difference in the initial state cause a large deviation in the system state in so little time?** This was the kick starter for the study of the Chaos Theory (GLEICK, 1989; CRILLY, 2008).

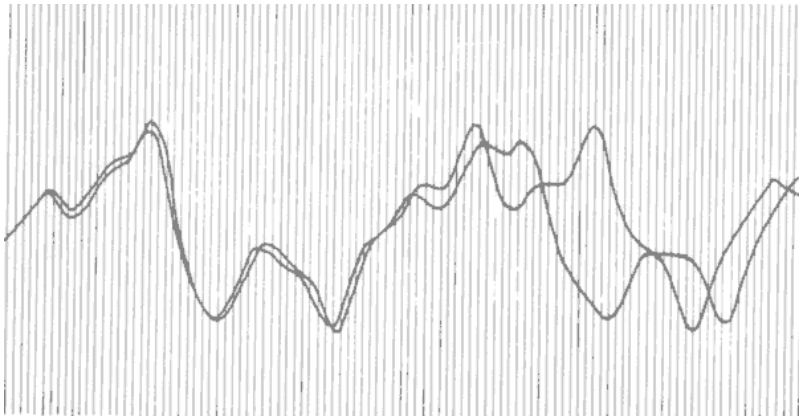

Figure 1.2: The difference between Lorenz's meteorological simulation.

Intrigued with this behavior, Lorenz decided to study this system deeply. He reduced his simulation to a simpler form, using only the convection of the air in the atmosphere, as we can see in figure 1.3. This system is modeled by the following set of non-linear different equations (ABARBANEL et al., 1993):

$$\frac{dx(t)}{dt} = \sigma(y(t) - x(t)) \tag{1.1}$$

$$\frac{dy(t)}{dt} = x(t)z(t) + \rho x(t) - y(t) \tag{1.2}$$

$$\frac{dz(t)}{dt} = x(t)y(t) - \beta z(t) \tag{1.3}$$

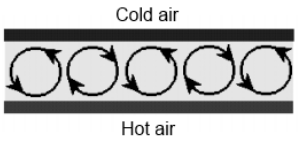

Figure 1.3: Scheme of the air convection in the atmosphere.

The Lorenz System behavior is dictated by the variation of the parameters σ [5], ρ [6] e β [7]. For the values of $\sigma = 16$, $\rho = 45.92$, and $\beta = 4$ 4 the Lorenz system shows the attractor seen in figure 1.4. It is possible to analyze the behavior of the Lorenz dynamical system, shown in figure 1.4, by measuring one of the system's dimensions. For example, measuring only **x(t)**, resulting in a time series, as we can see in figure reffig:lorenz-series.

Unfortunately this time series shows a big inconvenience. It could not be adequately analyzed by using the Linear Analysis techniques available by the time, based in the

[5] σ - Prandtl's Number: reason between the termal and the viscous dissipation.

[6] ρ - Rayleigh's Number: reason between Rayleigh's number and the critical value where convection begins

[7] β - Size of the convection cell.

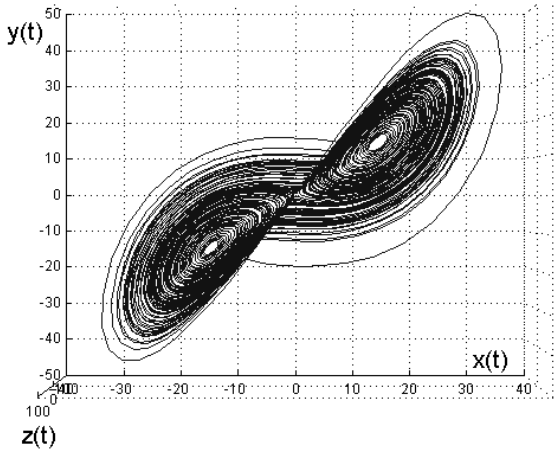

Figure 1.4: Lorenz Attractor for the values $\sigma = 16$, $\rho = 45.92$, and $\beta = 4$.

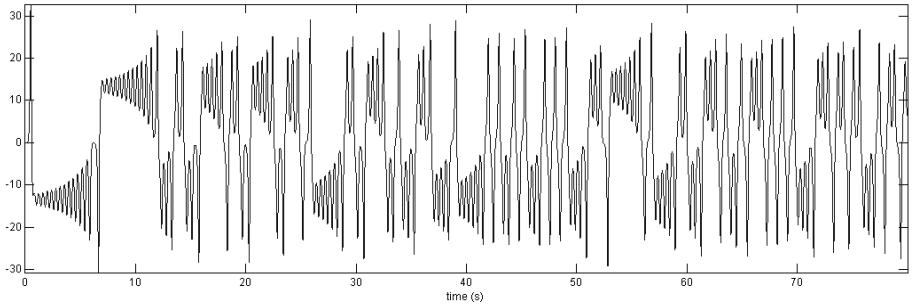

Figure 1.5: Time series measured from **x(t)** axis of Lorenz attractor.

Fourier Transform. For example, a signal that is adequately analyzed by using a Fast Fourier Transform (FFT) is a sinusoidal signal, like the one showed in figure 1.6. The FFT analysis of this signal is shown in figure 1.7, where the characteristics of a well-behaved signal (in the Linear Analysis point of view) can be seen: i) a narrow band, and ii) accentuated and narrow peaks. But when a FFT analysis of the time series measured from **x(t)** axis of the Lorenz attractor is performed it shows the same behavior showed by noise: i) wide bande, and ii) absence of peaks or peaks surrounded by a continuous band. The result of a FFT analysis over the time series measured from **x(t)** axis of the Lorenz Attractor can be seen in the figure 1.8. This signal would be understood as noise, and would be discarded by using linear analysis techniques. So a new approach would be needed for studying it (ABARBANEL et al., 1993).

For Abarbanel *et al.* (1993) the great revolution for the studying of systems that behaved like Lorenz system, the Chaotic Systems, was analyzing them by using Information Theory. A chaotic system is the realization of an Ergodic Information Source [8] , as in Shannon's theories (ABARBANEL et al., 1993), implying that the system orbits tend to occupy all the phase-space [9] in a small time. This occupation, which is independent of the initial state of the system, is known as Mixing, and is another characteristic shown by all Chaotic Systems. It was an encouraging example of the fitting of the non-linear techniques for the studying of this kind of problems, let's see figure 1.9. This figure shows the

[8] Ergodic Information Source: see (ABARBANEL et al., 1993), page 1333 for more references.

[9] Known as state-space too is the set of possible states for a given system. Its reconstruction from the time series is the first step for the non-linear analysis of the system (ABARBANEL et al., 1993; SUBHA et al., 2010).

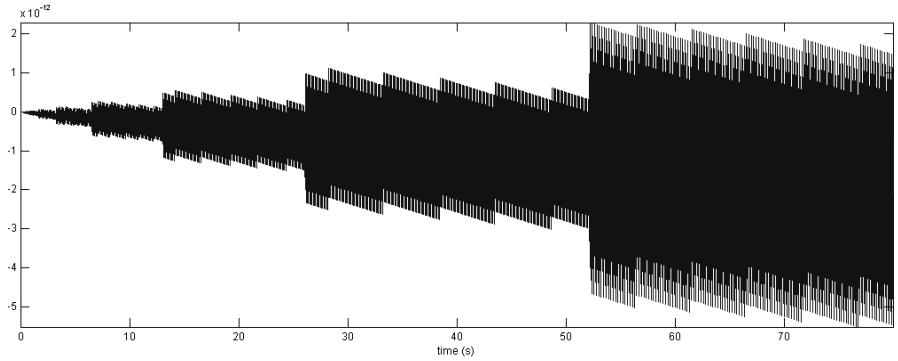

Figure 1.6: Time series measured from a sinusoidal signal with frequency $= 100Hz$ and time $= 80$ seconds.

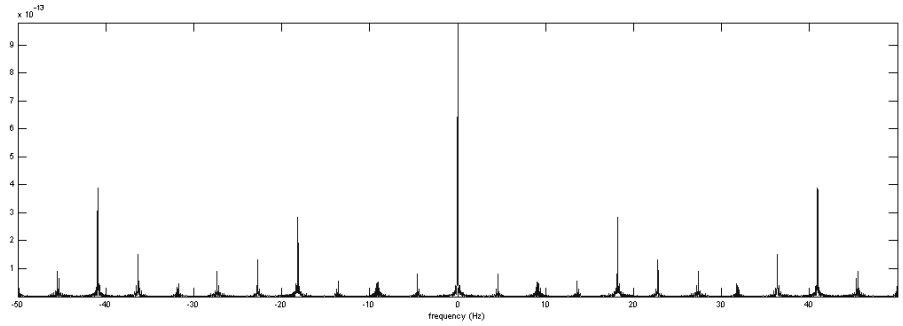

Figure 1.7: FFT analysis of a sinusoidal signal.

reconstructed Lorenz attractor from the time series measured from its x(t) axis the time series shown in figure 1.5 using a sampling delay (τ) of 7 samples, estimated as ideal by using Mutual Auto Information. These techniques will be detailed in chapter 2.

Under this new vision, all the theoretical tools available for the Information Theory researchers were available for the studying of Chaotic Dynamical Systems, and so several measures known as Dynamical Invariants [10] , as:

- **Linear auto-correlation and Mutual Information:** how much each measure of the system is affected by its past states.
- **Fractal dimension:** the spatial complexity of the manifold, the real dimension in which the signal is, the reason it grows **(FD)**.
- **Correlation Dimension:** the density of the attractor that generated the signal **(CD)**.
- **Kolmogorov-Sinai Entropy:** the signal predictability, is equal the sum of the positive Lyapunov exponents **(KS)**.
- **Lyapunov Exponents:** used to characterize chaotic behavior, because they always show positive values for this measure **(LE)**, the Largest **(LLE)**.
- **Hurst Exponent:** evaluates the long rage dependencies of the signal **(H)**.

[10] Dynamical Invariants have this name because its values do not depend on the initial state of the system, in opposition of the orbits, which diverge exponentially in the phase-space if the system is in chaotic regime.

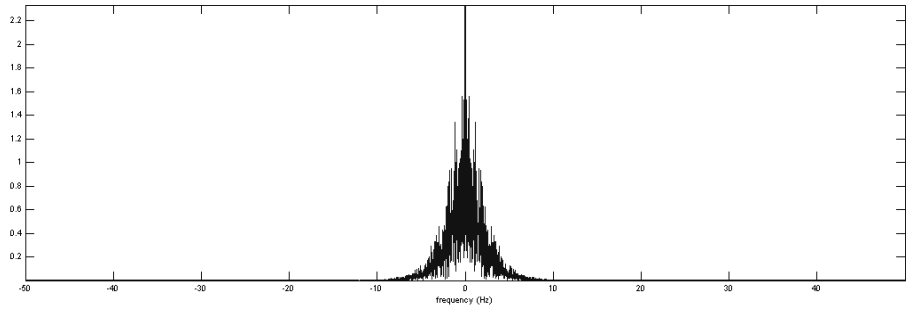

Figure 1.8: FFT analysis of the time series measured from **x(t)** axis of Lorenz attractor.

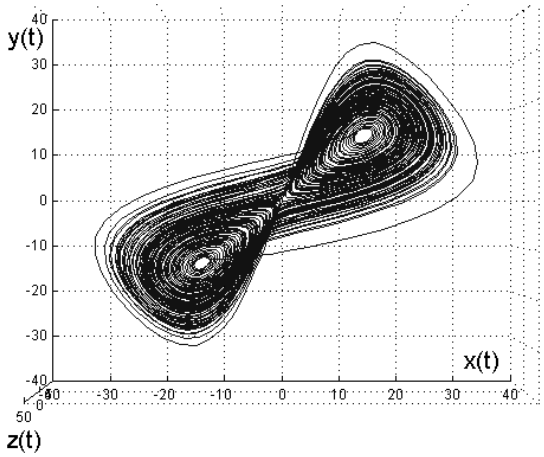

Figure 1.9: Lorenz attractor reconstructed from **x(t)**, $\tau = 7$.

With these theoretical tools it became now possible to deal with problems that under the linear analysis optic were considered untreatable, random, noisy, stochastic, exceptions, or even were stripped off from their chaotic nature because linear analysis techniques were not able to tell the difference between a chaotic system and noise. For Gleick (1989) the more emblematic problems of this initial phase of chaotic systems studies were:

- The meteorological studies of Lorenz. In his paper Deterministic Nonperiodic Flow he showed the existence of non-periodic systems that were deterministic too. Not being random or stochastic, only too complex for the linear analysis tools, so the name Deterministic Chaos.

- The turbulence dynamics studies of David Ruelle.

- The Jupiter's great red spot and its strange behavior.

- The modeling of animal populations growing of Robert May, that here shown one of the most simple known chaotic systems: $x_{t+1} = \alpha x_t(1 - x_t)$

- James Yorke and his paper Period three implies Chaos.

- Smale's horseshoe map and how something as simple as folding something in half recursively can lead to deterministic chaos.

- Benoit Mandelbrot (creator of the Mandelbrot Set, the most known fractal) and Hendrik Houthakker reaching the same results studying separately the wealth distribution and the cotton prices, discovering the nongaussianity of these systems, and the symmetry between the small changes (daily) and the large ones (over the years) by using computational methods for these analysis.

Several of the aforementioned studies were performed over time series, which are of major interest of several scientific areas. The observation of the measurements of some variable from a system, to allow the analysis, modeling, classification and prediction are common routines in science. The importance of time series generated by biological systems, known as biological signals as well, which show very complex dynamics, has long been recognized by non-linear analysis researchers. Several approaches used in this area were used to detect the important dynamical properties of the physiological phenomena involved (SUBHA et al., 2010). Many different studies show the presence of deterministic chaos in different biological signals, as, for example:

- **Lung sounds:** Chaotic dynamics of respiratory sounds (AHLSTRÖM et al., 2006).
- **Vocal tract:** Chaos in voice, from modeling to measurement (JIANG; ZHANG; MCGILLIGAN, 2006).
- **Electroencephalograms:** Is there chaos in the brain? II. Experimental evidence and related models (KORN, 2003), Dynamical paradigm in psychopathology: "chaos theory", from Physics to Psychiatry (PEZARD; NANDRINO, 2001), Is there chaos in the brain? Concept of nonlinear dynamics and methods of investigation (FAURE; KORN, 2001).

Signals that show a high complexity (as biological signals do) were for a long time studied using linear analysis techniques without success. Linear analysis techniques tend to obscure and distort fast transients of the signal. Linear analysis techniques are effective only if the analyzed signal has slow and/or smooth transients or if the generating system can be correctly modeled by a set of linear different equations (CHUNG; KENNEDY, 1991). Non-linear analysis techniques are more efficient for the analysis of biological signals than linear analysis techniques. The reason behind this is the high non-stationarity, non-gaussianity and the non-linear nature of the biological signals (SUBHA et al., 2010). Non-linear analysis techniques are not only more adequate to biological signals analysis, but are even capable of enhancing the detection and visualization of important parameters for physiological studies and clinical protocols (CERUTTI et al., 1996).

In the last four decades there was an increasing interest in the neural processes and brain signals (object of study of the computational neuroscience), especially the electroencephalogram (EEG), in the context of non-linear analysis and Deterministic Chaos Theory context (PALUS, 1998; BASAR, 2006). Because of the advances in this area science have considered the EEG as a powerful tool for the diagnosis of diseases like: epilepsy, tumor, cerebrovascular lesions, depression and traumas. The brain activity of a healthy person is easily distinguished from the abnormal brain activity of a unhealthy person, by using the right signal processing techniques (SUBHA et al., 2010).

Many different works in the nonlinear analysis area were developed in the **PPGC UFRGS**. Focusing mainly in works oriented by Prof. Dr. Dante Augusto Couto Barone we can cite: PhD Thesis **(PPGC UFRGS)** of Ricardo Custódio defended in 1999 - Nonlinear analysis in sound pattern recognition: case study of lung sounds (CUSTODIO,

1999). Some aspects of this research were published as a paper, as well, in Lung sounds analysis with time dependent fractal dimensions (CUSTODIO, 1999).

In 2000, in his Master Thesis **(PPGC UFRGS)**: Analysis and classification of non-stationary time series with nonlinear methods (THIELO, 2000), Marcelo Resende Thielo have studied electroencephalograms (EEGS), that are the focus of the present work. He performed the clustering of sleep stages using dynamical invariants and Genetic Algorithms. He also wrote the paper Treatment of local minima in a chaotic data clustering task with an extremely simple genetic algorithm (THIELO; BARONE, 2000).

Adriano Petry in 2003 defended his PhD Thesis **(PPGC UFRGS)**: Automatic speaker recognition using nonlinear dynamical features (PETRY, 2002), where He showed that using nonlinear dynamical invariants, more precisely the Fractal Dimension and Lyapunov Exponents, in addition to cepstral and melcepstral coefficients typically used, could enhance in approximately 17% the speaker identification rates. The results of this work were also published in many papers: Fractal dimension applied to speaker identification (PETRY; BARONE, 2001), Text-dependent speaker verification using Lyapunov exponents (PETRY; BARONE, 2002a), Speaker identification using nonlinear dynamical features (PETRY; BARONE, 2002b), and Preliminary experiments in speaker verification using time-dependent largest Lyapunov exponents (PETRY; BARONE, 2003).

More recently, Guilherme Maron published a series of papers: Measuring the differences between Spatial Intelligence in different individuals using Lyapunov Exponents (MARON; BARONE; RAMOS, 2012), nominated for the Best Paper Award in the 7th International Conference on Mass-Data Analysis of Images and Signals (MDA 2012), in which an statistical analysis of the brain activity of the participants during a spatial cognition task was performed. Then Guilherme Maron proceeded to the automatic classification of individuals regarding their Spatial Cognition degree of development, showing a method in the paper Spatial cognition degree of development classification Using largest Lyapunov exponents (MARON; BARONE; RAMOS, 2013a), and enhancing this method in the paper Spatial cognition degree of development classification Using artificial neural networks and largest Lyapunov exponents (MARON; BARONE; RAMOS, 2013b).

The aforementioned works developed by Guilherme Maron aimed to extend and enhance the works developed by Spíndola. The main inspiration of his works was Spíndola's PhD Thesis, oriented by Prof. Dr. Milton Antonio Zaro, developed in **PPGIE UFRGS**: *Spatial Cognition ability: measured using electroencephalography* [11] (SPINDOLA, 2010). Many papers wrote or co-authored by Spíndola were also an inspiration for the present work: Desenvolvimento de um protótipo EEG como ferramenta para caracterização de sinais cerebrais em atividades relacionadas a raciocínio lógico (CARRA et al., 2007), Em busca de um padrão cognitivo na Engenharia (CHIARAMONTE et al., 2007a,b) , A study of brain reaction to spatial stimuli in students with different background knowledge (SPINDOLA et al., 2010), and Cognitive Measure on Different Profiles (SPíNDOLA et al., 2010).

However, the vast majority of works in the Computational Neuroscience applied to cognitive investigation was done relying on the linear analysis theoretical tools, as will be shown in the related works section. Even the works from Spíndola were done over FFT analysis. This is an opposition to the works in Computational Neuroscience applied to Medicine, in which there is a large portion of the works developed under the non-linear analysis vision. So the present Master thesis aims to fill this gap, using non-linear analysis techniques in cognitive computational neuroscience investigation.

[11] A literal translation from Portuguese: Habilidade cognitiva especial: medida com eletroencefalografia.

1.1 Objectives

The conceptual map [12] shown in the figure 1.10 summarizes the conceptual assumptions, objectives, and hypotheses that guide the present Master thesis.

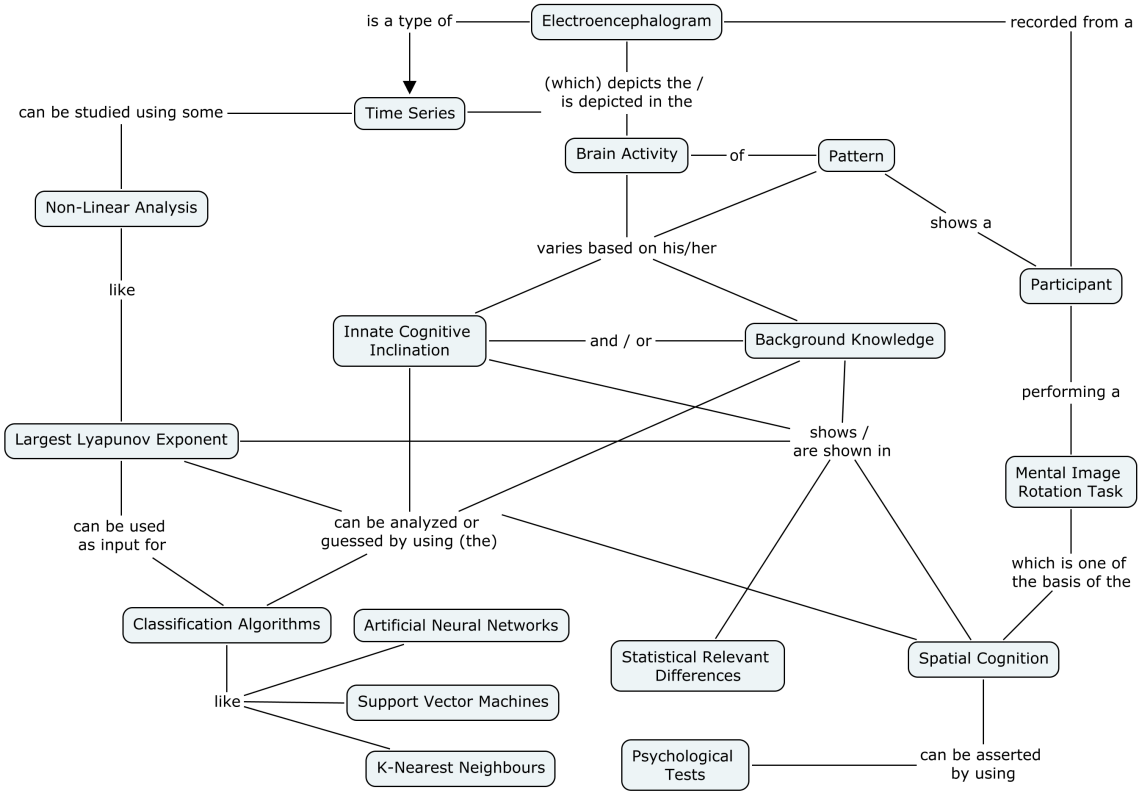

Figure 1.10: Conceptual Map that summarizes the conceptual assumptions, objectives, and hypotheses that guide the present Master thesis.

The electroencephalograms (EEGs), recorded from the participants are a type of time series, that describes the brain activity. The EEGs in the present experiment were recorded during the performing of a Mental tridimensional Image Rotation task, like the one described in (SHEPARD; METZLER, 1971) - which is one of the bases of the Spatial Cognition - which is the aspect of the intelligence concerned in imagination and mental manipulation of space and geometry. The Spatial Cognition degree of development can be evaluated using psychological tests, like BPR-5, which has a sub-test for Spatial Reasoning, BPR-5 SR (PRIMI; ALMEIDA, 2000). It is known that are observable differences in the brain systems that give support to the work memory and attention control. It results in general differences between the cognitive capabilities of different individuals (GEVINS; SMITH, 2000). The brain activity patterns of one varies, based on his background knowledge and/or innate inclinations, and this differences are statistically relevant (CHIARAMONTE et al., 2007a,b; SPINDOLA et al., 2010; SPíNDOLA et al., 2010; SPINDOLA, 2010; MARON; BARONE; RAMOS, 2012). As said before, the EEG is a type of time series, more precisely a biological time series. Complex signals as an EEG are better analyzed using non-linear analysis techniques (CHUNG; KENNEDY, 1991; ABARBANEL et al., 1993; CERUTTI et al., 1996; SUBHA et al., 2010). The non-linear analysis technique that will be used in the present work is the Largest Lyapunov Exponent (), which

[12] For references about Conceptual Maps see section 2.12 in chapter 2.

will be calculated from each EEG. The choice of using the LLE was done based in the fact that the LLE is related to the energy that is dissipated by the system, if the system is dissipative, like the brain is. More dissipated energy means more brain effort in performing the task. The energy spent is a good measure of the effort done by the brain to perform the task (SPINDOLA, 2010; MARON; BARONE; RAMOS, 2012). The LLEs then will be used as input for different classification algorithms: i) Artificial Neural Networks, ii) Support Vector Machines, and iii) K-Nearest Neighbors; aiming to guess the spatial cognition degree of development of each participant.

1.1.1 General Objective

The objective of the present Master Thesis is filling a gap in cognitive computational neuroscience the lack of works under the non-linear analysis vision, which is widely used and successful in medical computational neuroscience works. The goal in the present Master Thesis is to develop and present a method for measuring, analyzing and classifying the spatial cognition degree of development of different individuals, using non-linear analysis techniques.

1.1.2 Specific Objectives

- Propose and develop a method for measuring, analyzing and classifying the differences of the spatial cognition degree of development amongst different individuals.
- Use non-linear analysis techniques to process the recorded electroencephalograms (EEG), calculating the Largest Lyapunov Exponents (LLEs) from each EEG.
- Do a statistical analysis of the differences of brain activity of these individuals during a mental image rotation task.
- Develop a new classifier for WEKA (WAIKATO, 2013): a MLP that uses *tanh* (tansig) transfer function.
- Perform an automatic classification of these individuals regarding their spatial cognition degree of development.
- Do a comparison between the results achieved by using different classification algorithms.
- Discuss the achieved results with researchers more focused on the psychological and biological aspects of the neurosciences.

1.2 Motivation

The proposed statistical analysis and classification method can contribute and interact with other processes in the analysis and study of human spatial cognition, as in the understanding of the human intelligence at all. There are signals of the using of the Intraparietal Sulcus, one area of the brain related to numbers and quantities estimation, in this kind of spatial activity, even by kids 3 or 4 years old. They have shown the capability to, even in a limited way, estimate quantities, by choosing bags with more candies or showing surprise by finding less objects behind another one than they saw be hidden there. This notion of quantity is a building block of the logical-mathematical cognitive profile. When humans grow up they only apply more complex operations over this aspect of their intelligence

(DEHAENE; , EDITORS). In the present work, an investigation over the mental rotation of 3-D images capabilities, which must be one of the building blocks of the spatial cognitive profile, is performed. Previous works in this area can be seen in section 1.3, and are the basis for the present proposed method. A better understanding and evaluation of the cognitive capabilities of an individual could suggest him elements of motivation, ease or natural inclinations, possibly affecting the decisions of his carrier positively.

1.3 Related works

It is known that are observable differences in the brain systems that give support to the work memory and attention control. It results in general differences between the cognitive capabilities of different individuals (GEVINS; SMITH, 2000). Gevins and Smith have used a Wechsler Adult Intelligence Scale test (WAIS-R) to classify individuals in 2 groups: i) verbal cognitive style, and ii) non-verbal cognitive style. Then the researchers have analyzed the parietal activation differences among the individuals during the performing of work memory tasks. They concluded that the individuals in the non-verbal cognitive style group tended to have more right parietal activation, while the individuals in the verbal cognitive style group tended to have more left parietal activation (GEVINS; SMITH, 2000). Glass and Riding analyzed individuals in two dimensions regarding their cognitive style: i) wholist-analytic, and ii) verbal-imagery. The verbal-imagery dimension indicates how the individual tends to represent the information while thinking: verbally (verbal) or using mental images (imagery). The conducted experiments consisted in showing of different words for the analyzed individuals, using different exhibition rates, and then the individuals should press a button if the showed word could be classified in a previously known conceptual category. During this task the individuals EEGs were recorded from the midline, paramedial and lateral electrode clusters. The EEGs were then processed using FFT and the differences between the cognitive styles were analyzed and compared (GLASS; RIDING, 1999).

The analysis of brain activity during mental image rotation tasks using EEG is the focus of several papers. Roberts and Bell found in (ROBERTS; BELL, 2003) that 2D mental image rotation causes more activation on the left side of the brain than activation on the right side of the brain, while in women the opposite happens. During 3-D mental image rotation there is more activity on the right side of the brain in both genders. This differences were also evaluated between genders and between children and college students in (ROBERTS; BELL, 2000). Li and O'Boyle studied how gender, native language, and college major relates to the cognitive strategies used during 3-D mental rotation in (LY; O'BOYLE, 2008). The participants engaged in one task of mental image rotation while maintaining a concurrent verbal or spatial memory load. The physical sciences college students had better result while the concurrent task was one of spatial memory, while the opposite happened for the social sciences college students.

In previous works Spindola *et al.* studied the cognitive differences among engineering college students and social and human science college students (CHIARAMONTE et al., 2007a,b; SPINDOLA et al., 2010; SPíNDOLA et al., 2010; SPINDOLA, 2010), using Fast-Fourier Transform (FFT) to analyze the participants' EEGs during image rotation tasks. Another work that has studied the spatial cognition, but this time using LLE was conducted by Maron, Barone, and Ramos in (MARON; BARONE; RAMOS, 2012). They found that participants with an alleged higher degree of development of their spatial cognition showed 5% to 25% (average 15%) smaller averages in the LLEs calculated from

their EEGs during mental image rotation tasks.

Lyapunov Exponents are broadly used to analyze EEGs, but usually in search of medical information, as Swiderski *et al.* used to characterize epileptic seizures in (SWIDERSKI; OSOWSKI; RYSZ, 2005). Acharya *et al.* have studied the sleep stages with several techniques, and one of them are the LLE measured from EEGs captured while the participants were sleeping. They concluded that the LLE tends to have larger values at phases 3 and 5 of the sleeping, which are the states with larger variations (ACHARYA et al., 2005). To study the differences in brain behaviors during meditation the LLEs were used by Joseph, Kannathal and Acharya (2007) to show that meditation leads to a relaxed brain state, where the LLEs and other dynamical features show smaller values then while the participants were not meditating (JOSEPH; KANNATHAL; ACHARYA, 2007), the same conclusions were obtained by Patil and Bormane while studying Bramani meditation (PATIL; BORMANE, 2006). In cognitive studies the previously cited work of Maron, Barone and Ramos used the LLE to analyze the EEGs captured from subjects during mental image rotation tasks (MARON; BARONE; RAMOS, 2012). Maron, Barone and Ramos also proposed a method for the analysis and classification of spatial cognition degree of development using LLEs and Artificial Neural Networks in (MARON; BARONE; RAMOS, 2013a), and also have improved this method in (MARON; BARONE; RAMOS, 2013b). For other works with EEG analysis using LLE and Artificial Neural Networks the reading of the survey wrote by Suhba *et al.* (2010) is recommended.

1.4 Dissertation structure

This Master thesis begins with a fast and brief overview of the non-linear analysis history, from the studies of Poincare in the $3-$bodies problem to its modern applications in Computational Neuroscience. Chapter 1 also shows the objectives, motivation and a brief overview of the main concepts involved in the present work, as an overview of related works. Chapter 2 shows the theoretical basis of the present work, involving cognitive investigation, the mathematical basis of non-linear analysis, and the different computational tools used, mainly the classification algorithms applied to this problem. The developed method for analyzing and classifying the spatial cognition degree of development is shown in chapter 3, where the participants, equipment, data and processes involved in the present research are detailed. The results achieved are shown in chapter 4, where the statistical analysis of the spatial cognition differences amongst the groups is shown, followed by an analysis of the performance of the different classifiers in the studied problem. The last chapter, chapter 5, brings the conclusions obtained through the analysis of the results achieved, and a brief discussion of their implications. Also this chapter shows some possible future works that can be done based on the present one.

The present chapter had shown a brief overview of the history of non-linear analysis. Starting with Poincare's studies in the 3-bodyes problem, passing through Lorenz's studies in Meteorological simulation and other emblematic ones in the area, and reaching the studies of the brain activity, measured using electroencephalography, with a special highlight for the works in non-linear analysis and cognitive investigation developed in UFRGS. Also the main concepts involved in the present work are briefly explained and linked together, to give an overview of the objectives, motivations, theoretical basis and hypothesis that guided the present work. This chapter also gives a fast overview of previous works on spatial cognition investigation, the use of Largest Lyapunov Exponents and Artificial Neural Networks in electroencephalogram analysis.

Next chapter will show the theoretical basis of the present work, mainly: spatial cognition, non-linear analysis of signals, and classification algorithms.

2 THEORETICAL BASIS

In this chapter the theoretical basis of the present work will be shown. Starting with the non-linear analysis techniques used in the present problem:

- Mutual Information to estimate the time delay (τ).
- False Nearest Neighbors Method to estimate the embedding dimension.
- Reconstruction of the phase-space.
- Estimation of the Largest Lyapunov Exponent.

After this the main algorithms involved in the present method are briefly explained:

- Clustering algorithms
- Artificial Neural Networks
  - Multi-layer Perceptron (MLP)
  - Radial Base Function (RBF)
  - Voted Perceptron (VP)
- Support Vector Machines (SVM)
- K-nearest Neighbors (KNN)

The chapter ends with a brief overview about the physiological, psychological and pedagogical background on spatial cognition investigation.

2.1 Non-linear (chaotic) systems

A dynamical system is a system, usually described by a set of differential equations that evolve trough time, or some variable that looks like it, usually in a self-feedback process. One example of a dynamical system is the Logistic Map, that can be seen in equation 2.1, and was shown by Robert May while studying models of animal population grows. It is one of the simplest dynamical systems known.

$$x_{t+1} = \alpha x_t(1 - x_t) \tag{2.1}$$

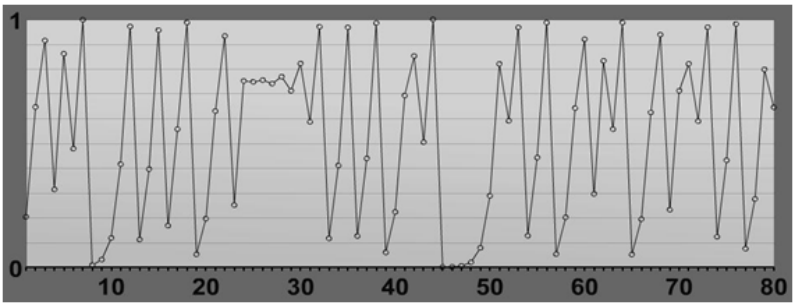

Figure 2.1: Logistic Map with α = 4, showing a chaotic behavior.

This system feedbacks its value in x_t to advance to its value in x_{t+1}. But this system, and others that are known as non-linear, or chaotic systems have some interesting characteristics. The first of them is their **Sensitivity** to it initial conditions. The Logistic Map behavior is mainly influenced by its α parameter. If α is set to 4, for example, a Chaotic Behavior arises, as can be seen in figure 2.1, its initial position x_0 is 0.2027.

But a little change in α parameter can change the behavior of the Logistic Map to a periodic one, as can be seen in figure 2.2, where α = 3.742718 and its initial position x_0 is 0.2027, as in the above example shown in figure 2.1.

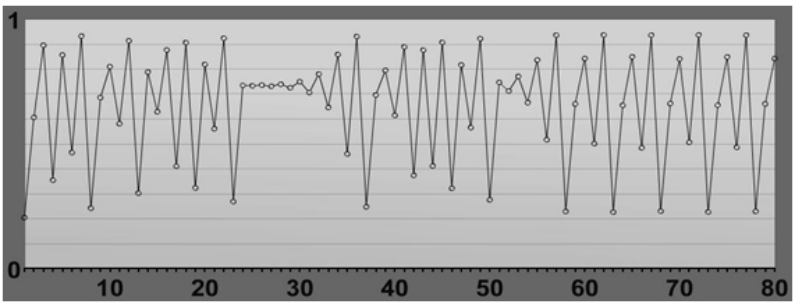

Figure 2.2: Logistic Map with α = 3.742718, showing a stable periodic behavior.

When α parameter is set to 2.75 the Logistic Map reaches a stable state, converging to a fix-point attractor, and there is no more influence of the initial position x_0 in the long run, as can be seen in figures 2.3 and 2.4, where the initial positions x_0 are 0.22 and 0.90, respectively.

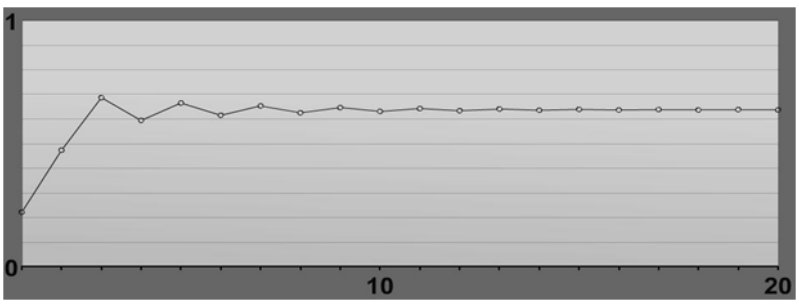

Figure 2.3: Logistic Map with α = 2.75, and x_0 = 0.22 showing a stable (fixed-point) behavior.

From now the focus will be in the chaotic behavior of the Logistic Map, with α = 4, as shown in 2.1. One way of understanding the affirmation: the perturbations (error) evolve

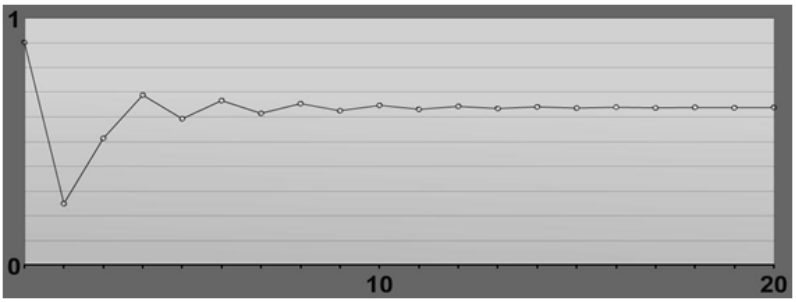

Figure 2.4: Logistic Map with α = 2.75, and x_0 = 0.9 showing a stable (fixed-point) behavior.

exponentially in a chaotic system; is to observe the difference in how the error evolves in a linear system and a chaotic system. The linear system which equation can be seen in equation 2.2 is shown in figure 2.5.

$$x' = xc, with\ c\ >\ 1 \tag{2.2}$$

With different values for c_0: 1.5, which is the original system; and 1.6, which will be considered the perturbed system, there is an initial error $E_0 = \epsilon$, 0.1 in this example. In n iterations of this system, taking in account the initial perturbation the observed value u_n will be $u_n = c_n(x_0+E_0)$. To calculate the error value at n a simple algebraic manipulation is done:

$$\begin{aligned} E_n &= u_n - x_n \\ u_n - x_n &= c_n(x_0 + E_0) - c_n x_0 \\ c_n(x_0 + E_0) - c_n x_0 &= c_n E_0 \end{aligned}$$

Therefore the error E propagates linearly trough the system, as can be seen in figure 2.5, with a reason c.

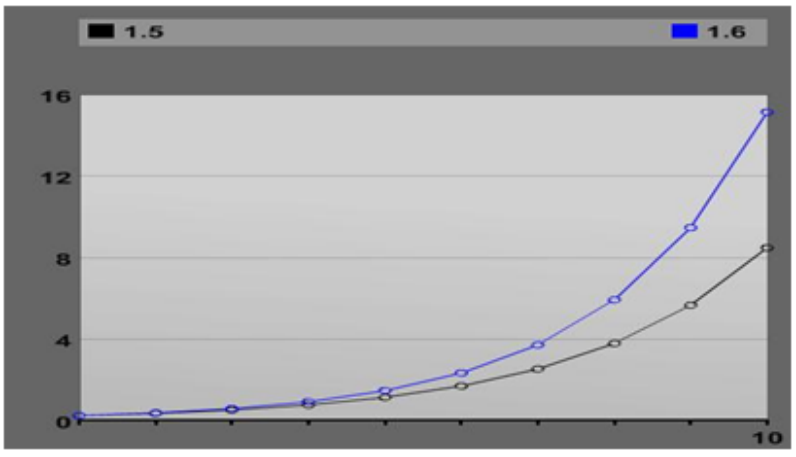

Figure 2.5: Error propagation in the linear system $x' = xc;\ c\ >\ 1$. X_0 = 0.22 and c_0 is 1.5 (dark gray line) or 1.6 (blue line).

In the other hand, the error propagates exponentially in a deterministic chaotic system, looking more like an entire new system. Returning to the Logistic Map one can see the error propagation analyzing the following scenarios: figure 2.6 shows the original system (x), with α = 4 and x_0 = 0.202; figure 2.7 shows the perturbed system (x'), with α = 4 and

x_0 = 0.202001; and figure 2.8 shows the behavior of the error itself, that can be calculated by doing $E_n = |x_n - x'_n|$, which was 0.000001 in t = 0.

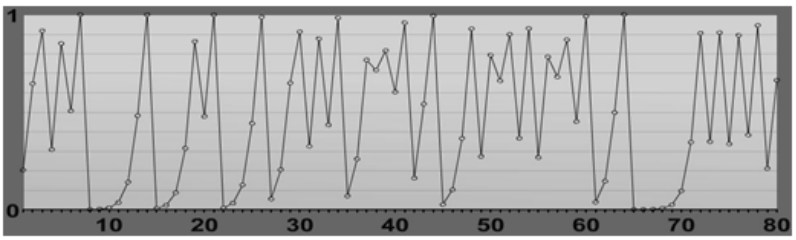

Figure 2.6: Logistic Map with α = 4, and x_0 = 0.202, the original system x.

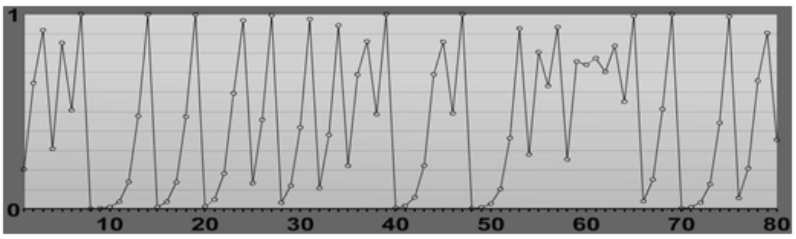

Figure 2.7: Logistic Map with α = 4, and x_0 = 0.202001, the the perturbed system x'.

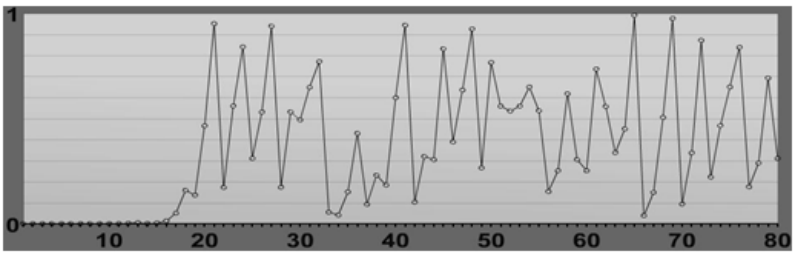

Figure 2.8: Error propagation in the Logistic Map, $E_n = |x_n - x'_n|$, E_0 = 0.000001.

Other characteristic shown by the Chaotic Systems is the **Mixing**, also known as topological transitivity, which in one characteristic that leads the orbits to occupy the entire phase-space in little time. In other words one can say that "in a Chaotic System you can start from almost anywhere and get to almost anywhere in a short time". A more formal formulation of this statement is to say that from 2 open intervals: I and J (greater than 0) it is possible to find initial values in I which when iterated will eventually lead to J. A simple experiment can be performed over the Logistic Map to show this characteristic. 10.000 equally spaced points will be generated, ranging from 0 to 1 and the ones that hit the interval [0.68, 0.69] by the iteration of the Logistic Map will be discarded. Figure 2.9 shows the survivors, points that did not iterated to the selected interval. The number of hits, points that reached the [0.68, 0.69] interval are shown in figure 2.10.

The third important characteristic shown by all Chaotic Systems is the presence of infinite arbitrarily close **Periodic Points** within each orbit. The demonstration of this characteristic is out from the scope of the present work, but a little experiment using the Logistic Map can be shown. Figure 2.11 shows the behavior for the Logistic Map x with α = 4 and $x_0 = sin^2\left(\frac{\pi}{7}\right)$ which is approximately 0.1885509907063. This x_0 value is a periodic point of the Logistic Map. Now perturbing this system, changing x_0 to 0.188550

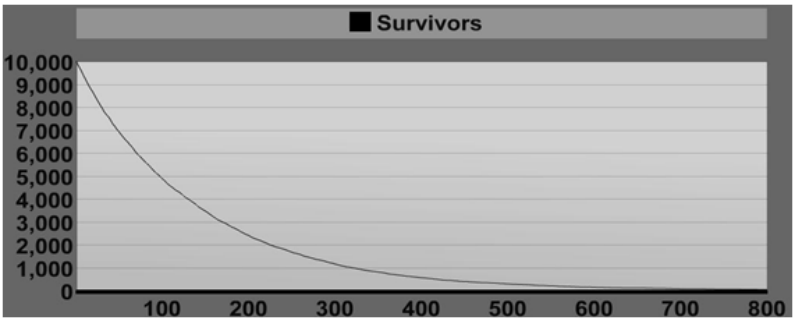

Figure 2.9: Number of survivors over the iterations of the Logistic Map, survivors are points that did not iterated over the [0.68, 0.69] interval.

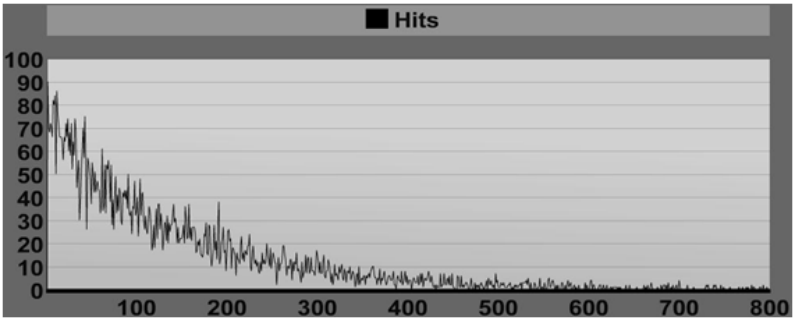

Figure 2.10: Number of hits in the [0.68, 0.69] interval by iterating the Logistic Map over 10.000 equally spaced points ranging from 0 to 1.

will result in x', it will show again a chaotic behavior, as can be seen in figure 2.12. The error propagation of this little divergence is shown in figure 2.13, with $E_n = |x_n - x'_n|$.

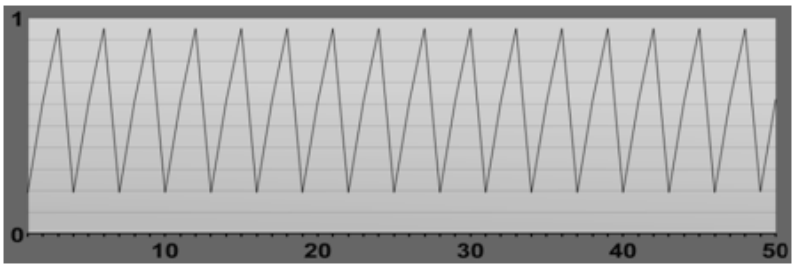

Figure 2.11: Logistic Map x with α = 4 and $x_0 = sin^2\left(\frac{\pi}{7}\right) \approx 0.1885509907063$.

In the nature several different phenomena show complex behaviors that make its analysis hard. Classic examples are: water flowing, air convection in the atmosphere and the time series generated by measuring vocal or brain activity. As an example one can think about the laminar water flowing: it will be very hard to determinate the position of each water particle through time, even them following simple and known physics rules (PETRY, 2002). The influence of very little perturbations can lead to very large deviations in the position and speed of the particles in a short time. Systems like that usually show the aforementioned characteristics. When facing systems with these characteristics one should use the theoretical tools of the non-linear analysis, also known as (Deterministic) Chaos Theory.

From now on the term signal will take the place of the term system, because the focus

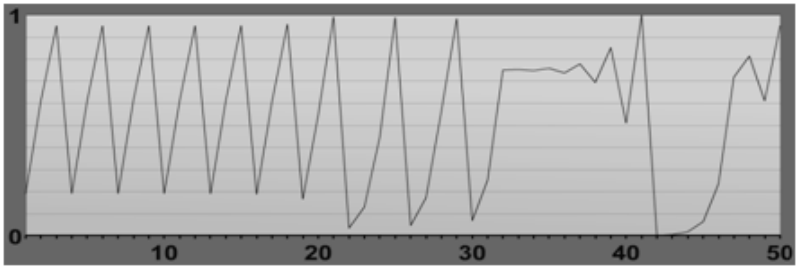

Figure 2.12: Logistic Map x' with α = 4 and x_0 = 0.188550.

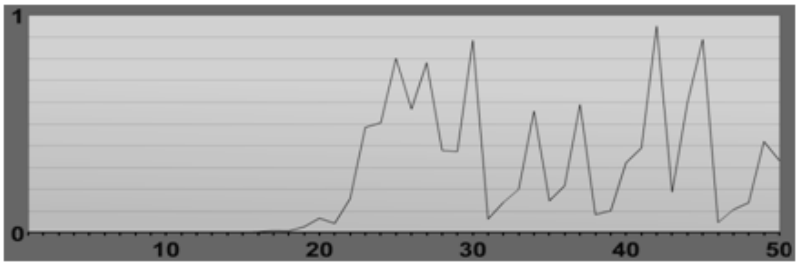

Figure 2.13: Error propagation in the Logistic Maps x and x', $E_n = |x_n - x'_n|$.

of the present work is non-linear analysis of biological time series. A time series is a set of measures $S = \{s_0, s_1, \cdots, s_n\}$ of the value if a variable within a finite time. This is how usually the information about the state of a system is measured, by means of one or more of its variables. In the present work the focus will be univariate time series.

2.2 Chaos in time series

To show the non-linear techniques used in the present work the Lorenz System will be used as example. This system was described by Lorenz in his paper Deterministic Nonperiodic Flow (LORENZ, 1963). This system is a simplified version of the air convection in the atmosphere, as shown in the figure 2.14, and is popularly known as Butterfly Effect. This system is modeled by the following set of non-linear different equations (ABARBANEL et al., 1993):

$$\frac{dx(t)}{dt} = \sigma(y(t) - x(t)) \tag{2.3}$$

$$\frac{dy(t)}{dt} = x(t)z(t) + \rho x(t) - y(t) \tag{2.4}$$

$$\frac{dz(t)}{dt} = x(t)y(t) - \beta z(t) \tag{2.5}$$

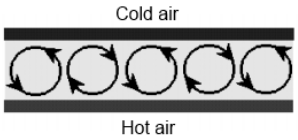

Figure 2.14: Scheme of the air convection in the atmosphere.

The Lorenz System behavior is dictated by the variation of the parameters σ [1], ρ [2] e β [3]. For the values of $\sigma = 16$, $\rho = 45.92$, and $\beta = 4$ 4 the Lorenz system shows the attractor seen in figure 2.15.

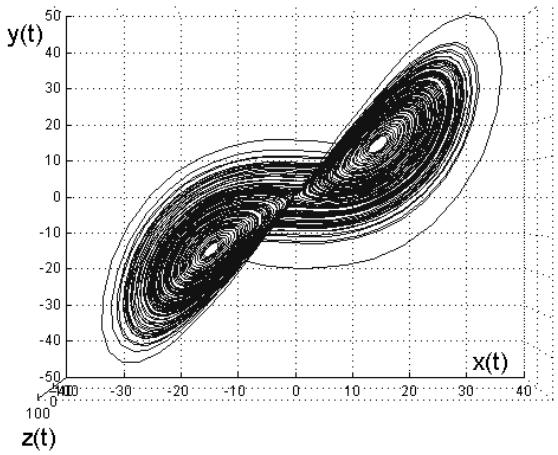

Figure 2.15: Lorenz Attractor for the values σ = 16, ρ = 45.92, and β = 4.

It is possible to analyze the behavior of the Lorenz dynamical system, shown in figure 2.15, by measuring one of the system's dimensions. For example, measuring only **x(t)**, resulting in a time series (signal), as we can see in figure 2.16.

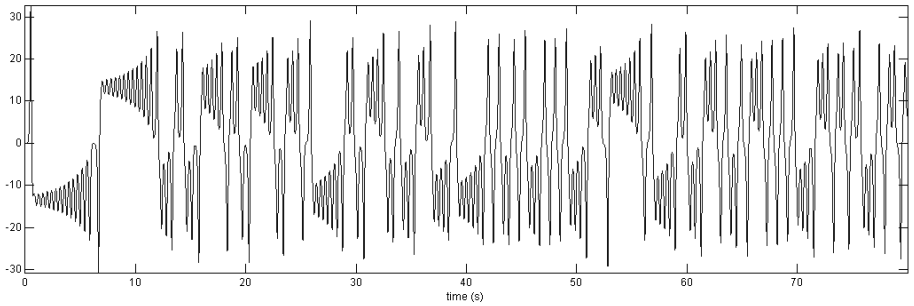

Figure 2.16: Time series measured from **x(t)** axis of Lorenz attractor.

The linear analysis techniques, based in the Fourier Transform, generally use the spectral peaks of the signal, the signal resonance frequencies to classify them. But, as said in chapter 1, this approach does not work with chaotic signals. Chaotic signals have the same behavior as the noise does when analyzed this way, as can be seen in figure 2.18, that shows the Fast Fourier Transform (FFT) of a sinusoidal signal 2.17, where the characteristics of a well-behaved signal (in the Linear Analysis point of view) can be seen: i) a narrow band, and ii) accentuated and narrow peaks.

But when applied to the time series generated by the **x(t)** axis of the Lorenz attractor the FFT analysis shows the same behavior showed by noise: i) wide bande, and ii) absence of peaks or peaks surrounded by a continuous band, as can be seen in figure 2.19.

The non-linear analysis relays on calculating different measures, the non-linear invariants, to correctly analyze a chaotic signal. Non-linear invariants are independent of

[1] σ - Prandtl's Number: reason between the termal and the viscous dissipation.

[2] ρ - Rayleigh's Number: reason between Rayleigh's number and the critical value where convection begins

[3] β - Size of the convection cell.

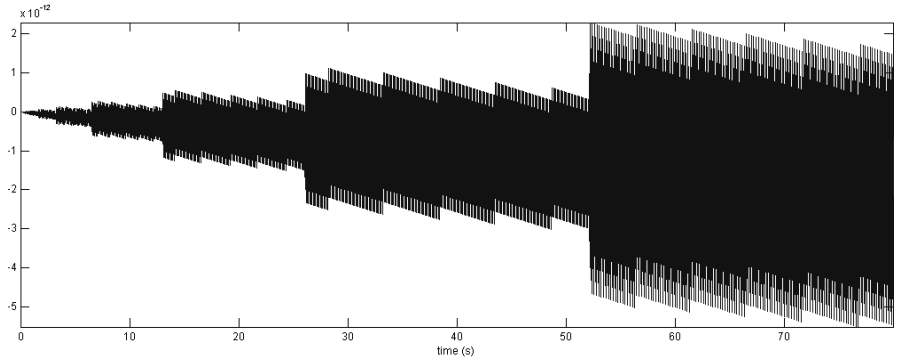

Figure 2.17: Time series measured from a sinusoidal signal with frequency = $100Hz$ and time = 80 seconds.

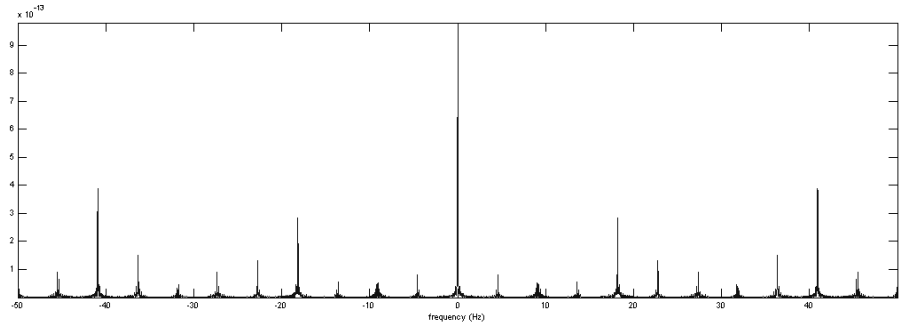

Figure 2.18: FFT analysis of a sinusoidal signal.

the initial state of the system, and are not affected by small deviations in the orbits, unlike the orbits themselves, that are exponentially sensible to these changes. This intrinsic instability turns impossible the tasks of comparing orbits experimentally or with artificially generated ones, because they do not have correlation between themselves and even the finite numeric precision used in the process will turn the orbits into different ones (ABARBANEL et al., 1993). The most used non-linear invariants in electroencephalogram (EEG) analysis are (ROBERT; GAUDY; LIMOGE, 2002; SUBHA et al., 2010):

- **Fractal dimension:** the spatial complexity of the manifold, the real dimension in which the signal is, the reason it grows **(FD)**.
- **Correlation Dimension:** the density of the attractor that generated the signal **(CD)**.
- **Kolmogorov-Sinai Entropy:** the signal predictability, is equal the sum of the positive Lyapunov exponents **(KS)**.
- **Lyapunov Exponents:** used to characterize chaotic behavior, because they always show positive values for this measure **(LE)**, the Largest **(LLE)**.
- **Hurst Exponent:** evaluates the long rage dependencies of the signal **(H)**.

The present work will focus on the Largest Lyapunov Exponent (LLE). The first step to calculate the LLE is the reconstruction of the attractor, and this will be the goal of the following sections.

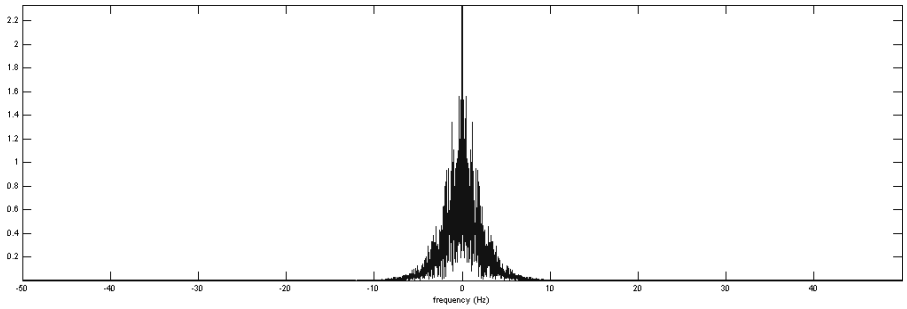

Figure 2.19: FFT analysis of the time series measured from **x(t)** axis of Lorenz attractor.

2.3 Detecting chaotic behavior in signals

The most used technique to evaluate if a signal is chaotic or not is using Surrogate Data analysis. This technique consists in generate datasets with the same mean, standard deviation and variance of the original signal, by shuffling its points. The resulting signals will have the same spectral characteristics of the original one (ACHARYA et al., 2006). Then some non-linear measure is taken from them, for example, Correlation Dimension (CD). If the difference between the original signal CD and the shuffled ones are above 50% the signal is considered chaotic (THEILER et al., 1992; SUBHA et al., 2010). A deeper study of this technique is out from the scope of this work.

2.4 Reconstructing the attractor

Attractor is the set of states to which the system tends to evolve during its observation in the phase-space. In other words, dynamical systems are attracted to its attractors. The term phase-space is used to describe the p-dimensional space where a system evolves. Abarbanel et al (1993) says that the relevant dynamics of a system are not in the same dimension as the system itself, which in certain cases can show an infinite number of degrees of freedom, but in geometries of much lower dimensionality, the attractors (ABARBANEL et al., 1993). When the original phase-space is projected into a time series, a 1-D space, the result is a time series obtained by measuring one of the system's variables, as can be seen in figure 2.20, adapted from (PETRY, 2002). In the method developed in the present work: electroencephalograms (EEGs) captured using the Event-related potentials (ERP).

Takens' theorem proofs that an attractor topologically identical to the original attractor which generated the signal can be reconstructed from the measured time series (TAKENS, 1981). This reconstructed attractor will have non-linear invariants with values very close to the original one's, and are relevant and valid for studying the original system. The technique showed by Takens, the time-delays, is valid if the dimension m respects the condition:

$$m \geq 2D + 1,$$

with D being the dimension (usually fractal) from the real system. After Sauer et al. (1991) demonstrated that the above condition can be changed by

$$m > 2D_0$$

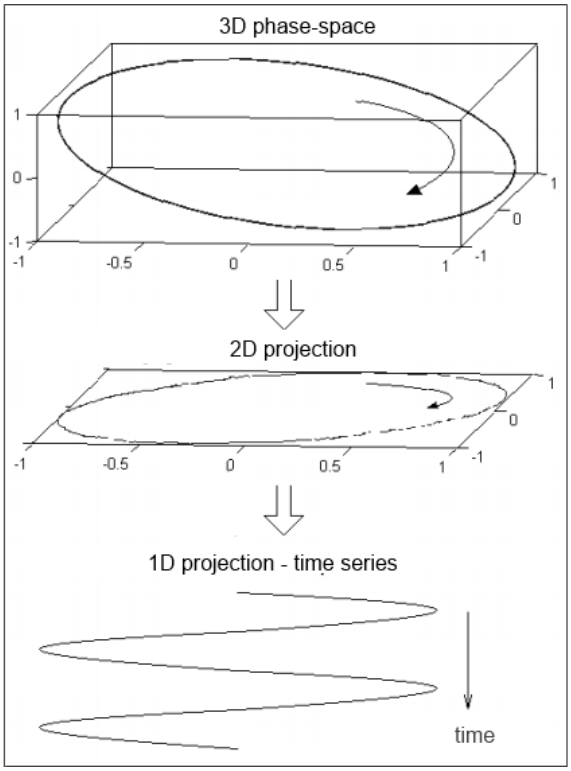

Figure 2.20: Temporal projection (measured time series) of a system that evolves in a tridimensional phase-space.

where D_0 is the box-counting dimension of the attractor in D. The simplicity of this method made it largely diffused and applied to several different knowledge areas.

There are 4 different types of attractors that can be seen in dynamical systems (CAMBEL, 1992), the figure 2.21 was adapted from (HUDSON, 2000):

- **Fixed-point:** attractors that tend to a stable state, not showing sensibility to the initial states and not leading the system to a chaotic behavior, figure 2.21 **(A)**.
- **Limit cycles:** attractors that show a defined cycle, small perturbations cannot take the system out of the attractor because the perturbed orbit tends to return to it, figure 2.21 **(B)**.
- **Torus attractor**: attractors that looks like "wool balls" where there are energy conservation, they usually happen in systems with a high number of degrees of freedom, figure 2.21 **(C)**.
- **Strange attractor:** attractors with complicated geometric properties, with fractal geometry and highly irregular. Are found in non-periodic dissipative dynamical systems, figure 2.21 **(D)**.

The first step to calculate the LLE of a signal, which is the objective of this and the following sections, is to reconstruct the system's attractor. Transforming the 1-D time-series

$$s(t) = \{s(t_1), s(t_2), \cdots, s(t_n)\},$$

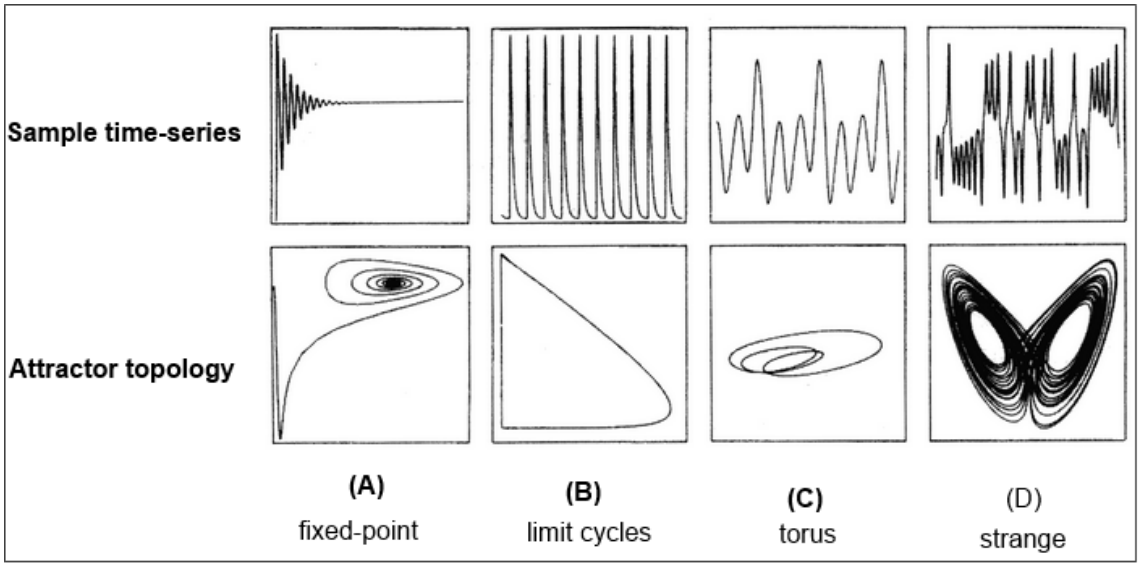

Figure 2.21: Types of attractors that can be seen in dynamical systems, and sample series that could have been generated by them.

measured from the system into an m-dimensional attractor in the $\Re^m$ Euclidean Space. To create an attractor that preserves the topological properties of the original unknown attractor the most common method is the method of delays introduced by Takens (TAKENS, 1981)(SUBHA et al., 2010). The reconstructed m-dimensional vectors are in the form:

$$x(t) = \{x(t_i), x(t_i + \tau), x(t_i + 2\tau), ..., x(t_i + (m-1)\tau)\},$$

where $s(n)$ is the original time-series, m is the embedding dimension and τ is the embedding delay. For reconstructing the time-delayed vectors, knowing the delay and the embedding dimension, one can use the delay program in the TISEAN suite (HEGGER; KANTZ; SCHREIBER, 2007). The next section will detail the method used to estimate the τ-delay.

2.4.1 Estimating the embedding delay (τ)

To estimate the embedding delay τ (in samples) that will be used to reconstruct the attractor the first minimum of the Mutual Information function, as suggested by Fraser and Swinney (FRASER; SWINNEY, 1986), is widely used, and is calculated as:

$$S = -\sum_{ij} p_{ij}(\tau) ln\left(\frac{p_{ij}}{p_i p_j}\right)$$

where p_i is the probability of finding a value from the time series in the i-th interval and $p_{ij}(\tau)$ is the joint probability that an observation is done in the i-th interval and τ time later in the j-th interval. This technique takes the non-linear correlations in account (SUBHA et al., 2010). As t increases, S decreases, and then rises again, so the optimal τ-delay is the first minimum of the Mutal Information function (FRASER; SWINNEY, 1986; SUBHA et al., 2010). In figure 2.22 the mutual information function of the Lorenz time-series is shown. The first minimum is at 0.16 seconds, which is approximately, 7 samples, because the Lorenz signal was sampled at 50Hz.

The right estimation of τ-delay is very important; otherwise the reconstructed attractor will not hold the same topological properties as the original attractor. We can see the

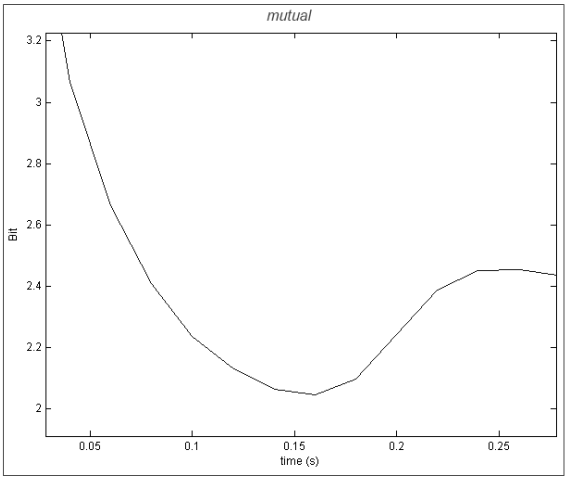

Figure 2.22: Mutual information in Lorenz system. First minimum at 0.16 seconds, or $\approx$ 7 samples.

results of choosing different τ-delays for reconstructing the Lorenz attractor. In figure 2.23 **(A)** the reconstruction is done using a delay that is smaller than the optimal one: 3 samples. 2.23 **(B)** shows the attractor reconstructed using the optimal delay: 7. Finally, 2.23 **(C)** shows the attractor reconstructed using a delay larger than the optimal one: 13. Compare the original Lorenz attractor and the reconstructed one in figure 2.24, where **(A)** is the original attractor, and **(B)** is the attractor reconstructed from x(t) axis time series using a τ-delay of 7. This kind of analysis can be done using mutual program in the TISEAN suite (HEGGER; KANTZ; SCHREIBER, 2007).

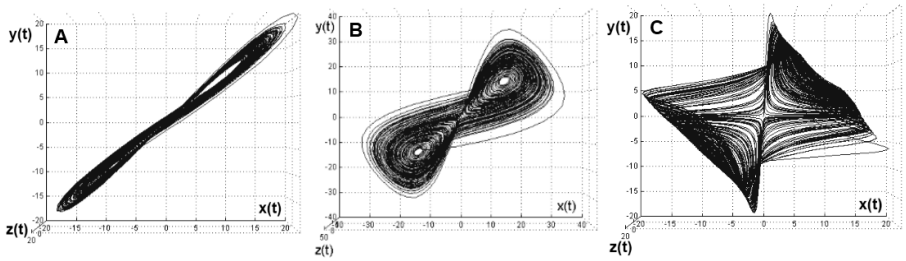

Figure 2.23: The effect of using different τ-delays to reconstruct the Lorenz attractor.

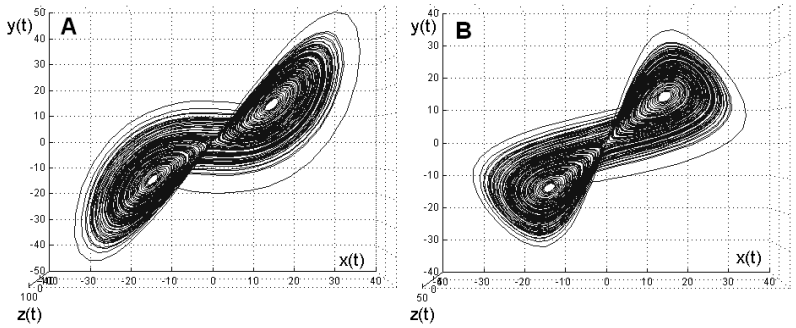

Figure 2.24: **(A)** the original Lorenz attractor, and **(B)** the reconstructed attractor.

2.4.2 Estimating the embedding dimension (m)

Other important step to reconstruct the attractor is the estimation of the embedding dimension. If the choosing embedding dimension *m* is smaller than the minimal embedding dimension m_0 some points will be projected over neighborhoods they should not be mapped into, ruining the topological structure of the reconstructed attractor. If the attractor is reconstructed using an adequate dimension, the reconstructed attractor will be a one-to-one image of the attractor in the original phase-space. Thus the neighborhoods will be mapped onto their neighborhoods again (KENNEL; BROWN; ABARBANEL, 1992; SUBHA et al., 2010).

To estimate the embedding dimension (m) the most used method is the False Nearest Neighbor (FNN), suggested by Kennel et al. (KENNEL; BROWN; ABARBANEL, 1992). For each point in the time-series $\vec{s_i}$, let $\vec{s_j}$ be the look for its nearest neighbor in m-dimensional space. The distance between them is given by

$$\|\vec{S_i} - \vec{S_i}\|$$

Iterate both points and compute:

$$R_i = \frac{|S_{i+1} - S_(j+1)|}{\|\vec{S_i} - \vec{S_i}\|}.$$

R_t is the threshold point which has a nearest false-neighbor. One should choose a dimension high enough to make fractional points for which $R_t > R_t$ is zero or very small. This kind of analysis generates a graphic like the one shown in figure 2.25, in the case of the Lorenz Attractor an embedding dimension of 3 was chosen, when only 2.8% of the points were false-neighbors. This kind of analysis can be done using the false_nearest program in the TISEAN suite, for example.(HEGGER; KANTZ; SCHREIBER, 2007).

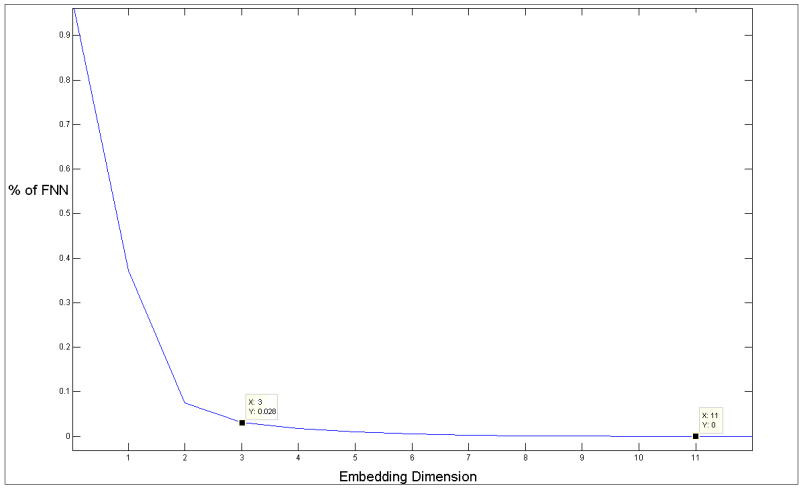

Figure 2.25: False-nearest neighbors for Lorenz system, with m = 3 the % of false-neighbors is 2.8%.

2.5 Largest Lyapunov Exponent (LLE)

The Lyapunov Exponent (λ) is the measure that represents the rating in which two initially arbitrarily close orbits diverge in the phase space, in other words, the rating of

information loss in the system (SUBHA et al., 2010). The LLE is widely used to study chaotic dynamical systems: if a system has one or more positive LE it is (most likely) a non-linear (chaotic) system (without knowing the system initial condition we cannot predicted its future state), stable attractors give an LLE of 0 and trajectories will have a negative LLE (SUBHA et al., 2010).

To calculate the LLE it is necessary to reconstruct the attractor first, this process was detailed in the previous sections. By using the reconstructed attractor the LLE can be calculated using, for example, the method proposed by Wolf (WOLF et al., 1985): having a point X_0 and another point $X_0 + \delta_{X_0}$. Both points will create orbits in the phase-space, δ_X away from one another, which can be written as $\delta_X(X_0, t)$. The largest divergence rating is the LLE (λ) and can be computed as:

$$\lambda = \lim_{t \to \infty} \frac{1}{t} \, ln \, \frac{|\delta_X(X_0, t)|}{|\delta_{X_0}|}.$$

This kind of analysis can be done, for example, using lyap_k program in the TISEAN suite (HEGGER; KANTZ; SCHREIBER, 2007), which uses a method proposed in (KANTZ, 1994). The LLE calculated for the Lorenz attractor using lyap_k was 0.90487.

The recommended embedding dimension for reconstructing EEGS ranges from 5 to 20, and a time delay of 1 is recommended (DIAS-TOSTA et al., 1999), but maybe this delay should be increased if a sampling rate higher than $200Hz$ is used, which was the sampling rate used in (DIAS-TOSTA et al., 1999).

2.6 Clustering algorithms

As pointed out by Backer and Jain, “in cluster analysis a group of objects is split up into a number of more or less homogeneous subgroups on the basis of an often subjectively chosen measure of similarity (i.e., chosen subjectively based on its ability to create “interesting” clusters), such that the similarity between objects within a subgroup is larger than the similarity between objects belonging to different subgroups” (BACKER; JAIN, 1981) apud (XU; WUNSCH D., 2005). In other words, clustering is grouping objects (in this case, data points) in groups that have objects that have large similarity between them and a small similarity with the ones outside its cluster. In figure 2.26 an example of 2 different clusterings performed over the same data is shown, both using 3 different clusters.

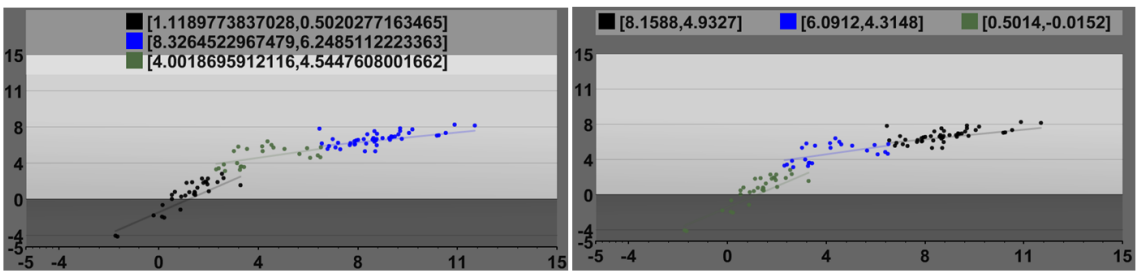

Figure 2.26: Example of 2 different clusterings over the same data.

The most used and one of the simplest, clustering algorithms is the K-Means, which is shown in (HARTIGAN, 1975) (the oldest reference found) and can be calculated as:

$$arg_s min \sum_{i=1}^{k} \sum_{x_j \in S_i} \|x_j - \mu_i\|^2,$$

which means: minimize the sum of the square of the distances (Euclidean) between k centroids and the nearest points of each of them (also known as WCSS: within-cluster sum of squares).

The K-Means cluster algorithm implementation present in WEKA (WAIKATO, 2013), called X-Means was used in the method shown in chapter 3, as can be seen in section 3.5 in chapter 3. This algorithm was proposed by Pelleg and More in (PELLEG; MOORE, 2000).

2.7 Classification algorithms

Classifications algorithms are basically of 2 types:

- Supervised learning algorithms
- Unsupervised learning algorithms

The last ones are usually called clustering algorithms, and were already described in previous sections. The first ones are mapping of the inputs vectors $x \in \Re^d$, where d is the input space dimensionality, to a discrete finite set of label, modeled by a mathematical function that usually requires training to minimize some empirical error or risk function (BISHOP, 1995; PELLEG; MOORE, 2000).

The main goal of a classification algorithm is to recognize objects with features that are close to features shown by known objects and assign the same label (class) to them. In the next sections the different classification algorithms used in the present work will be detailed.

2.8 Artificial Neural Networs (ANN)

Artificial Neural Networks (ANNs) is a knowledge area of the Computational Intelligence also known as Conexionist. The main inspiration comes from the biological neurons and the biological nervous system. The ANNs aim to reproduce, in different granularities, the behavior shown by them, and applying this solutions to solve computational problems. Haykin (1998) says that ANNs are widely used for problems of time series analysis, classification, prediction and pattern recognition. The main reason for that is the emergent capability of learning from the input data, and extrapolate them (HAYKIN, 1998). The next sections will describe the ANNs types used in the present work, that are the most used in the EEG analysis literature (SUBHA et al., 2010).

2.8.1 Perceptron

An Elementary perceptron is a non-linear artificial neuron based on the McCulloch-Pitts model: a linear integrator followed by an abrupt limiter (usually a sigmoid) that is the signal function (HAYKIN, 1998). A scheme of an elementary perceptron can be seen in figure 2.27.

The function implemented by an elementary perceptron is:

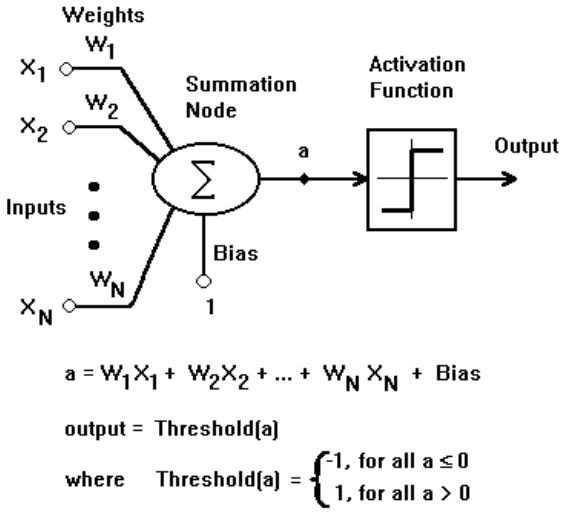

Figure 2.27: Scheme of an elementary perceptron.

$$a = \sum_{n=1}^{N} W_n X_n + Bias,$$

with N being the number of inputs in the neuron. X_n the input value present in the n-th input. W_n the weight associated with the n-th input, usually obtained through a training process done by a method of gradient descending in the mean quadratic error function. And $Bias$ is an external bias.The output of the neuron, a, is then evaluated against an activation function, usually a sigmoid like an Hyperbolic Tangent or a Logistic Function and if the value is high enough: > 0 for the Hyperbolic Tangent; or > 0.5 for the Logistic Function, the neuron output will be 1, otherwise it will be 0. This means that the perceptron creates an hyper-plan that divides the data-space in 2 parts This hyper-plan can be defined as:

$$\sum_{n=1}^{N} W_n X_n + Bias \; = \; 0.$$

Therefore the elementary perceptron is only capable of solving linearly separable problems (HAYKIN, 1998). One example of a linearly separable problem can be seen in figure 2.28.

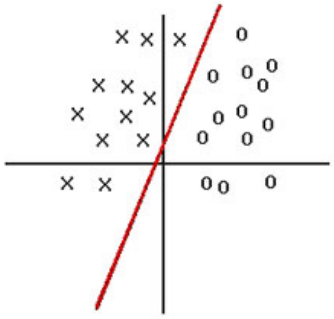

Figure 2.28: One example of a linearly separable problem.

2.8.2 Multi-layer perceptron (MLP)

As stated in the previous section, an elementary perceptron is only capable of resolving linearly separable problems. But if multiple perceptrons are used, grouped in different layers, as shown in figure 2.29, with 3 or more layers, being at least:

- **Sensory layer:** the sensory neurons that receive the input values.
- **Hidden Layer(s):** the computational neurons, that are in charge of adding non-linearity to the problem solution, by dividing the larger problem into smaller ones, that are linearly separable.
- **Output layer:** computational neurons that are in charge of labeling of mapping the outputs from the hidden layer(s).

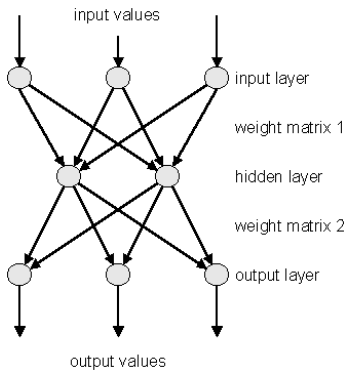

Figure 2.29: Scheme of a Multi-layer Perceptron (MLP).

if using a high enough number of neurons in the hidden layer, and maybe a high enough number of hidden layers, the result will be an Universal Function Approximator (HAYKIN, 1998). Is known that all the computational neurons in a MLP are elementary perceptrons, thus using sigmoid activation functions, so based on the Cybenko Theorem (CYBENKO, 1989) that says that, being $f(.)$ a continuous non-constant function which is limited and monotony crescent. Being I_p an unitary p-dimensional hyper-cube $(0,1)^p$. The space of the continuous functions in I_p is denominated $C(I_p)$. So, given any function $g \in C(I_p)$ and with $\epsilon > 0$, there is an integer M and a set of $\Re$ constants α_i and w_{ij}, where $i = \{1, \cdots, M\}$ and $j = \{1, \cdots, p\}$, that is possible to define:

$$F(x_1, x_2, \cdots, x_p) = \sum_{i=1}^{M} \alpha f(\sum_{j=1}^{p} w_{ij} x_j - w_{0i})$$

as an approximation of the function $g(.)$ such that,

$$\|F(x_1, x_2, \cdots, x_p) - g(x_1, x_2, \cdots, x_p)\| < \epsilon$$

for all $x_1, \cdots, x_p \in I_p$.

In the present work 2 different sigmoid transfer functions were evaluated within the MLP classifier:

- **Hyperbolic Tangent (tansig):** $n = \frac{2}{1+exp(-2*n)} - 1$

- **Logistic Function (logsig):** $n = \frac{1}{(1+exp(-n))}$

MLPs are usually trained using back-propagation, one algorithm that adjusts the errors between the MLP outputs and the expected outputs, by propagating the error backwards in the network, and adjusting the synaptic weights w_n by the amount of error that each wn weight is responsible for. Further detailing over the MLP training process is out of the scope of this work.

2.8.3 Radial Base Function (RBF)

While a MLP is composed of neurons (perceptrons) that calculate a non-linear function of the scalar product of its inputs and the synaptic weights associated with them, and the neurons in the hidden layer and in the output layer of the MLP are the same. A Radial Base Function ANN (RBF) shows neurons that use the concept of Euclidean Distance between the input and a prototype vector to decide the activation of its hidden neurons. And the output layer in a RBF is composed of linear elements (HAYKIN, 1998). The scheme of a RBF is shown in figure 2.30, took from (WOLFRAMRESEARCH, 2005)

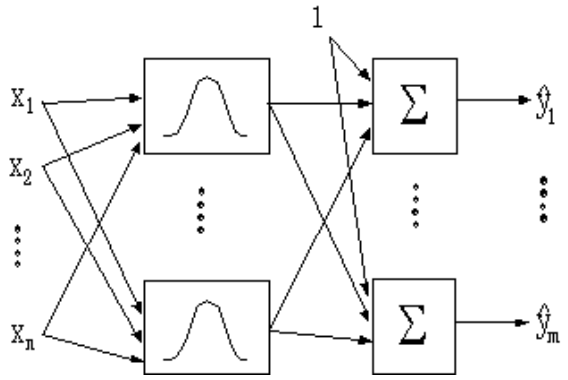

Figure 2.30: Scheme of a Radial Base Function (RBF) Artificial Neural Network.

A RBF, in an extremely simplified way, performs a non-linear mapping from the input to a high-dimensionality space, where the problem tends to be linearly separable, and after this maps them linearly to a single dimension: the output of the hidden layer neurons. A RBF with p inputs will perform a mapping that can be defined as:

$$s : \Re^p \mapsto \Re^1$$

The neurons in the hidden layer of a RBF keep the center of a radial base function, usually a Gaussian, that being spherical (mono-variate) is defined by:

$$\phi_j(x) = exp\left[-\frac{(x-c_j)^2}{2\sigma_j^2}\right]$$

being c the center of the function, and σ_j^2 its variance. The activation of this neuron, as an answer to the current input, is dependent of the Euclidean distance between the input and the center c of this neuron. Further detailing over the MLP training process is out of the scope of this work. Further detailing over the RBF working principles, theoretical basis and training process is out of the scope of this work.

2.8.4 Voted Perceptron (VP)

Voted perceptron (VP) is a technique that combines Rosenblatt's perceptron algorithm (ROSENBLAT, 1958; ROSENBLATT, 1962) with an adaptation of the batch learning algorithms with leave-one-out approach from Helmbold and Warmuth (HELMBOLD; WARMUTH, 1995). This technique exploits the fact that if the problem is linearly separable a perceptron will make a finite number of mistakes therefore, if the training data set is cycled enough, the perceptron will converge to a vector that will correctly classify all instances. The same trick that is used on Support Vector Machines (SVMs) of applying Kernel functions over the separation hyper-planes is used in Voted Perceptrons too.

The basic VP algorithm is rather simple, and is shown in figure 2.31, took from (FREUND; SCHAPIRE, 1999), and is based on keeping the voting vectors that generated a mistake, then count for how much time they "survived" without doing other mistake, when the larger time without a mistake, the larger the weight of this neuron vote.

Training

Input: a labeled training set $\langle(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)\rangle$
number of epochs T

Output: a list of weighted perceptrons $\langle(\mathbf{v}_1, c_1), \ldots, (\mathbf{v}_k, c_k)\rangle$

- Initialize: $k := 0$, $\mathbf{v}_1 := \mathbf{0}$, $c_1 := 0$.
- Repeat T times:
  - For $i = 1, \ldots, m$:
    - Compute prediction: $\hat{y} := \text{sign}(\mathbf{v}_k \cdot \mathbf{x}_i)$
    - If $\hat{y} = y$ then $c_k := c_k + 1$.
      else $\mathbf{v}_{k+1} := \mathbf{v}_k + y_i \mathbf{x}_i$;
      $c_{k+1} := 1$;
      $k := k + 1$.

Prediction

Given: the list of weighted perceptrons: $\langle(\mathbf{v}_1, c_1), \ldots, (\mathbf{v}_k, c_k)\rangle$
an unlabeled instance: $\mathbf{x}$

compute a predicted label $\hat{y}$ as follows:

$$s = \sum_{i=1}^{k} c_i \, \text{sign}(\mathbf{v}_i \cdot \mathbf{x}); \quad \hat{y} = \text{sign}(s).$$

Figure 2.31: The Voted Perceptron algorithm.

The Voted Perceptron was developed by Freud and Schapire, and described in (FREUND; SCHAPIRE, 1998, 1999). The VP implementation present in WEKA (WAIKATO, 2013) was used in the present work.

2.8.5 Support Vector Machines (SVM)

A Support Vector Machine (SVM) is a non-probabilistic binary linear classifier. Its main goal is to find a separation hyper-plane that divides both classes with the largest possible gap (BURGES, 1998). The points in this hyper-plane satisfy the following equation:

$$w.x + b = 0$$

where w is the normal (not necessarily normalized) vector to the hyper-plane, x is the point itself an $|b|/\|w\|$ is the perpendicular distance from the separation hyper-plan to the origin. In the simplest case of a linearly separable problem, the following conditions will be respected:

$$x_i.w + b \geq \ for\ y_i = +1$$
$$x_i.w + b \leq -1\ for\ y_i = -1$$

Figure 2.32, took from (BURGES, 1998), shows this case, the circled data points are the support vector.

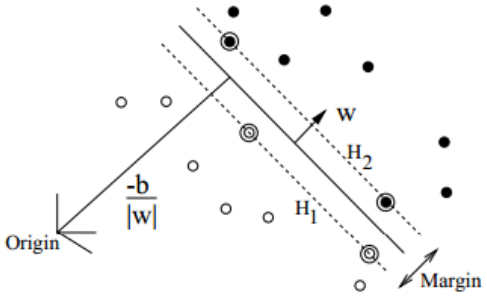

Figure 2.32: Using SVM in a linearly separable problem.

In problems that are not linearly separable the same trick used in the Voted perceptron is used: Kernel functions. The decision hyper-plane then is deformed by the kernel function, usually gaining generalization, figure 2.33, took from (BURGES, 1998), shows an example of how a decision hyper-plane could look like with the use of Kernel functions.

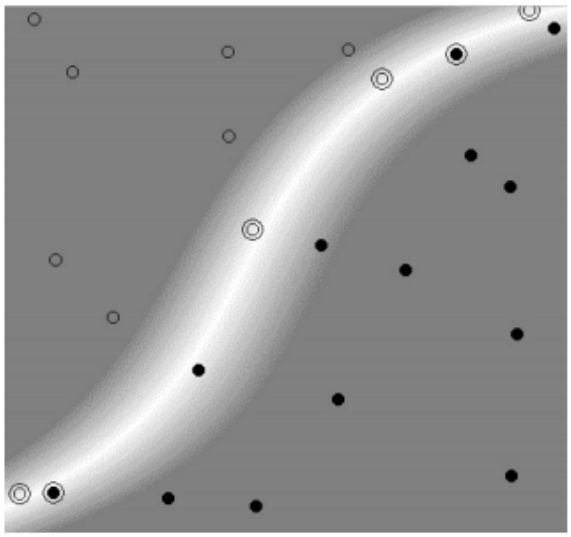

Figure 2.33: An example of how a decision hyper-plane could look like after the use of Kernel functions.

The implementation of SVM present in WEKA (WAIKATO, 2013), called SMO was used in the present method. The SMO algorithm was tested using the 4 different available Kernel functions:

- **Polynomial kernel:** $K(x,y) =< x,y >^p$ or $K(x,y) = (< x,y > +1)^p$; p is an arbitrary exponent.
- **Normalized Polynomial kernel:** $K(x,y) = \frac{<x,y>}{\sqrt{<x,x><y,y>}}$, where $< x,y >= PolynomialKernel(x,y)$.
- **RBF kernel:** $K(x,y) = e^{-(\gamma * <x-y,x-y>^2)}$, γ is an arbitrary parameter.
- **PUK kernel:** for references see (UESTUEN; MELSSEN; BUYDENS, 2006).

2.9 K-nearest neighbors (KNN)

The K-Nearest Neighbors (KNN) algorithm is one of the simplest types of instance-based learning algorithms. K nearest instances of the current instance are used to decide the class of the current instance, by a majority of votes, so the current instance class will be the most common class among its K nearest neighbors. Any distance function can be used to evaluate which are the K-nearest neighbors. In the present method the implementation of KNN present in WEKA (WAIKATO, 2013) was used, it is called IBK and based in the KNN implementation shown in (AHA; KIBLER; ALBERT, 1991).

2.10 Electroencephalogram (EEG)

An electroencephalogram (EEG) is a time series composed of measurements of electrical activity on the scalp due to brain activity trough time. These measures are done using electrodes that are attached to the individual scalp. These electric indicators are the excitatory and inhibitory post-synaptic electric potentials that occur during the neurons communication process. They sum up at the cortex and extend to the surface, where they can be measured using non-invasive techniques, as, for example, Event-related potentials (ERP). Usually EEG signals measured from the scalp have amplitudes that range from about $10\mu V$ to $100\mu V$, and frequencies that range from $1Hz$ to about $100Hz$ (SUBHA et al., 2010). An example of an EEG signal can be seen in figure 2.34.

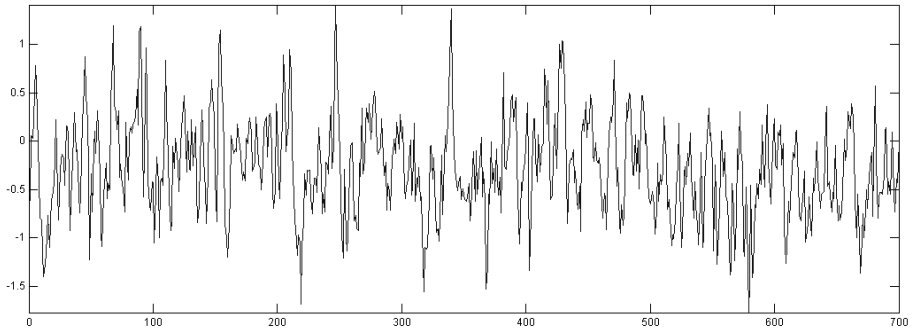

Figure 2.34: An example of an EEG signal.

2.11 Spatial Cognition

Excitatory and inhibitory post-synaptic electric potentials that occur during the neurons communication process sum up at the cortex and extend to the surface, where they

can be measured using an electroencephalograph, resulting in an EEG recording (SUBHA et al., 2010). This EEGs turn possible the analysis of verbal comprehension degree, perceptual organization, work memory and processing speed of healthy individuals, highlighting the verbal and visual capabilities of them (CUNHA, 2003).

The most accepted theories of how intelligence works are based on the modularity of the intelligence, and the brain itself, with different aspects, focused on different task which are conceptually closely related. The most known works in this aspect are from Howard Gardner (GARDNER, 1983, 1999, 2005), Steven Pinker (PINKER, 1999, 2000).

Spatial Cognition is the aspect of the intelligence concerned with the mental manipulation of geometries, colors, distances, positions, and the space itself. Gardner says that this aspect of the intelligence tends to be well developed, for example, in engineers, architects and plastic artists, because people who are in these careers or had a natural inclination for it, or developed this by training (GARDNER, 1983).

An individual1s spatial cognition can be evaluated by applying psychological tests, as, for example, the BPR-5 test, which has a subtest focused on Spatial Reasoning, the BPR-5 SR (PRIMI; ALMEIDA, 2000). This test is described in section 3.1 in chapter 3.

The spatial cognition is focused on processing visual inputs, so for the present work an experiment based in the one conducted by Shepard and Metzler in (SHEPARD; METZLER, 1971) was used. This experiment is described in section 3.2, in chapter 3. The use of this kind of tasks in Spatial Cognition investigation are aligned with previous researches: (ROBERTS; BELL, 2000, 2003; XU et al., 2006; LY; O'BOYLE, 2008; CHIARAMONTE et al., 2007a,b; SPINDOLA et al., 2010; SPíNDOLA et al., 2010; SPINDOLA, 2010; MARON; BARONE; RAMOS, 2012, 2013a,b).

The frontal and pre-frontal lobes of the brain are related with visual processing, by using the work memory for that. Experiments conducted in (GAZZANIGA; IVRY; MANGUN, 2006) show that spatial tasks tend to recruit more the right hemisphere pre-frontal lobe, but there are other studies that show the possibility of bi-laterality, maybe because of its wide communication with parietal and temporal lobes (KANDEL; SCHWARTZ, 1985).

The visual stimuli decoding happens in the occipital lobe, and after this follow 2 different paths, to be processed in the temporal and parietal lobes in each hemisphere of the brain, say Ungerleider and Mishkin (1982) *apud* (GAZZANIGA; IVRY; MANGUN, 2006). Each of this lobes is specialized in one different task:

- **Temporal Lobe:** perception and recognition of objects.
- **Parietal Lobe:** spatial perception, recognition of positions and spatial configuration.

In each hemisphere the goal is slightly different, says Warrington (1985) *apud* (GAZZANIGA; IVRY; MANGUN, 2006), in the right hemisphere there is the perceptive processing, to recognize the object. Meanwhile, in the left hemisphere there is the semantic processing, to name the object. Therefore the choice of the electrodes:

1. **FP1:** pre-frontal lobe (left)
2. **FP2:** pre-frontal lobe (right)
3. **F3:** frontal lobe (left)

4. **F4:** frontal lobe (right)
5. **T3:** temporal lobe (left)
6. **T4:** temporal lobe (right)
7. **P3:** parietal lobe (left)
8. **P4:** parietal lobe (right)

aims to cover the lobes that are used in the rotation, attention, comparison and recognition of geometries in the human brain, as described by the literature.

2.12 Conceptual Maps

Conceptual Maps are educational artifacts created by Joseph Novak (NOVAK; nAS, 2008), based on the Significant Learning Theory of David Asubel. Asubel stated that a new learning will only be really learned, and will have some meaning for the learner, if it was anchored with other concepts already understood by the learner, and if the learner was capable of correlate them with the new information (MOREIRA, 2006).

In a Conceptual Map, as the one shown in figure 1.10 in chapter 1, the concepts are shown within geometric figures (circles, ellipses, squares), connected by arrows that guide the reading in one direction or lines that connect two concepts, describing their relation with a succinct description. But for the correct interpretation of a Conceptual Map it is mandatory that it is followed by a descriptive text (MOREIRA, 2006).

In the present chapter the theoretical basis in which the present work was built over are shown, as the tools used for the process of classifying the spatial cognition degree of development from different individuals. The most important topics were:

- Non-linear analysis of signals.
- Classification algorithms.
- Spatial Cognition investigation.

In the next chapter the proposed method will be detailed in depth, showing each step of the proposed method.

3 METHOD

The method proposed, developed and shown in the present thesis is summarized in the scheme shown in figure 3.1. In the following sections every step of this scheme will be explained in details.

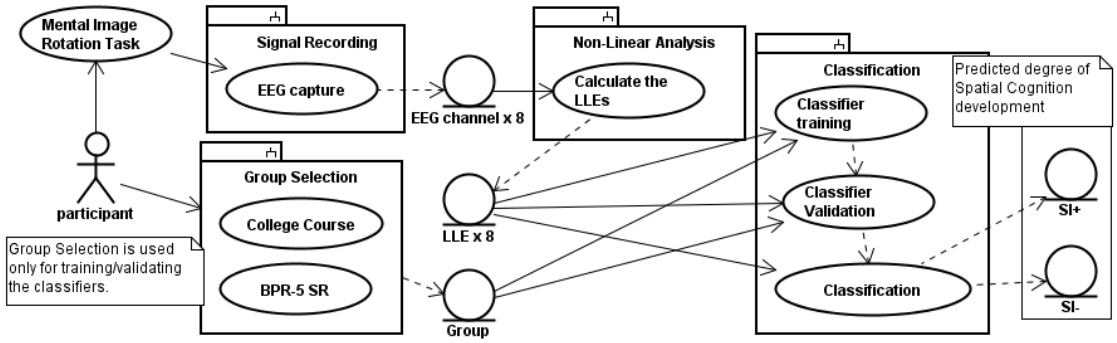

Figure 3.1: Scheme of the proposed method.

Section 3.1 will describe the **Participants** of the experiment, and how they were assigned in different groups of interest. Explaining the step named **Group Selection** in the scheme. The **Mental Image Rotation Task** performed by the participants is described in the section 3.2. The section 3.3 explains the equipment used for recording the EEGs, and the process itself, the step called **EEG Recording** in the scheme. The non-linear analysis performed over the EEGs recorded is explained in section 3.4, the step named **Non-Linear Analysis** in the scheme, as the tools used for it. The data sets sets generated by the previous steps of the method: **EEG Channels**, **LLES**, and **Group** are explained in the section 3.5. Finally, section 3.6 explains the process of automatic classification of the participants regarding their spatial cognition degree of development. Also section 3.6 shows a comparison between the used classification algorithms. This is the step called **Classification** in the scheme.

3.1 Participants

Thirty-seven undergraduate students (23 engineering students, 14 human and social science students), who were free of medications and neurological diagnosis volunteered to EEG measurement in UCS, in the state of Rio Grande do Sul, Brazil. They were then separated in 2 different groups: i) participants with an alleged higher degree of spatial cognition development (named **SI+**), and ii) a control group with no alleged special spatial cognition development (named **SI-**). There was no significant age difference between

both groups. This hypothesis of classification was based in their college course, and strengthened by applying a subset of the Battery of Reasoning Tests BPR-5 to them. This classification hypothesis was also used in (CHIARAMONTE et al., 2007a,b; SPINDOLA et al., 2010; SPíNDOLA et al., 2010; SPINDOLA, 2010; MARON; BARONE; RAMOS, 2012, 2013a,b).

The BPR-5 is a set of tests of Abstract, Verbal, Numerical, Spatial and Mechanical Reasoning used by Psychology professionals. Only the Spatial Reasoning subtest, BPR5-SR, was applied to the participants. The BPR5-SR subtest is composed of 20 questions. In each one of the questions there are cubes in different positions, indicating their movements. The cubes can show constant movements, like: always turn up, or alternated ones, like: turn up then turn right the cube. The task is composed of 2 parts: first, recognizing the movement pattern performed in the cube, and second, assigning what should be the next cube if the following movement was applied. There are 5 options in each question and the participants have up to 8 minutes to do the entire subtest. An example of a BPR5 - SR subtest question can be seen in figure 3.2.

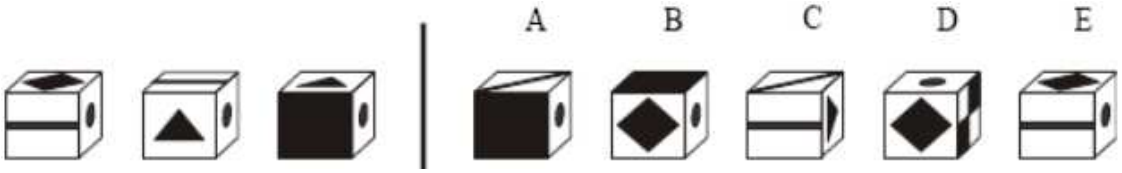

Figure 3.2: An example of a BPR-5 SR test question.

Table 3.1 and figure 3.3 show the results achieved by the participants in the BPR-5 SR subtest. An easily spotted difference between the SI+ and SI- groups' performance can be observed. A t test succeeded to reveal a statistically reliable difference between the number of right answers achieved by the 2 groups in the BPR-5 test: i) SI+ (M = 15.21, s = 3.26) and ii) SI- (M = 11.14, s = 4.52), t(37) = 3.181, p = .037, α= 0.05.

Table 3.1: The BPR-5 SR test results regarding groups SI+ and SI-.

GROUP	**right answers**	**% right**	**participants**	**% group**
SI+	20	100%	2	9%
	18+	90%	7	30%
	15+	75%	13	57%
	10+	50%	22	96%
SI-	20	100%	0	0%
	18+	90%	2	14%
	15+	75%	3	21%
	10+	50%	8	57%

The measurements were done after being approved by the ethic committee of UCS, in the scope of the project *Eletroencefalografia como instrumento avaliador de habilidade*

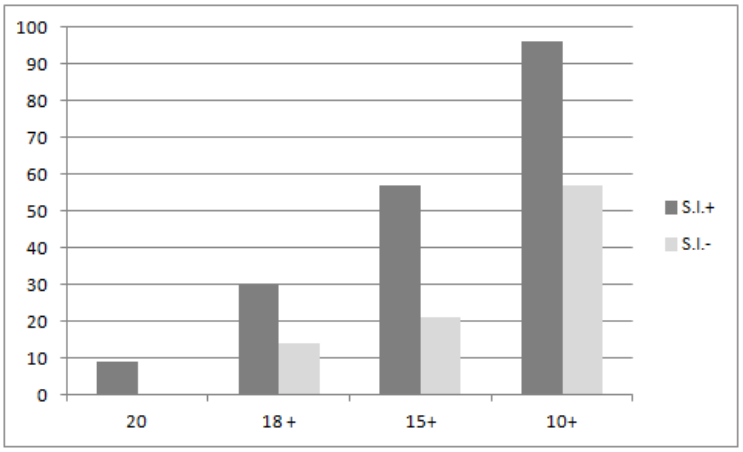

Figure 3.3: The cumulative right answer from SI+ and SI- groups participants in the BPR-5 SR test.

espacial [1], and the respective document can be found in (SPINDOLA, 2010).

3.2 Mental rotation of tridimensional geometric patterns task

The Spatial Cognition task chosen was the mental rotation and recognition of virtual tridimensional geometric patterns. Tasks like that are innate to the spatial cognition, which is recruited for imagination, rotation and recognition of mental images and geometries. The 3 different possible patterns are shown in figure 3.4, as proposed by Shepard and Metzler in (SHEPARD; METZLER, 1971).

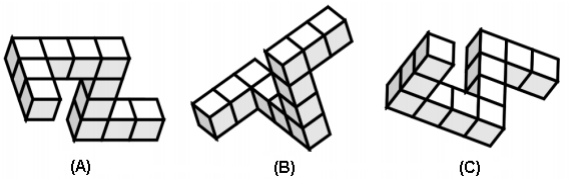

Figure 3.4: The tridimensional patterns shown to the participants.

The participants were exposed to six different combinations of the aforementioned patterns that will be called stimulus from now on, and the differences between each different stimuli is shown in table 3.2. Six different stimuli were used, shown in figure 3.5. During each stimuli the participant should answer by pressing different buttons if the two patterns in the stimuli were the same, or not. The showing of the stimuli was done respecting an interval of 10 seconds between them, and every stimulus was shown for 800 milliseconds.

3.3 EEG recording

The EEG equipment used was described in (CARRA et al., 2007), and used in (CHIARAMONTE et al., 2007a,b; SPINDOLA et al., 2010; SPíNDOLA et al., 2010; SPINDOLA, 2010; MARON; BARONE; RAMOS, 2012, 2013a,b) too. This equipment belongs to the CENT center of CARVI campus of UCS, where the EEG recordings were performed. A

[1] Electroencephalography as spatial cognition evaluating instrument, in a free translation

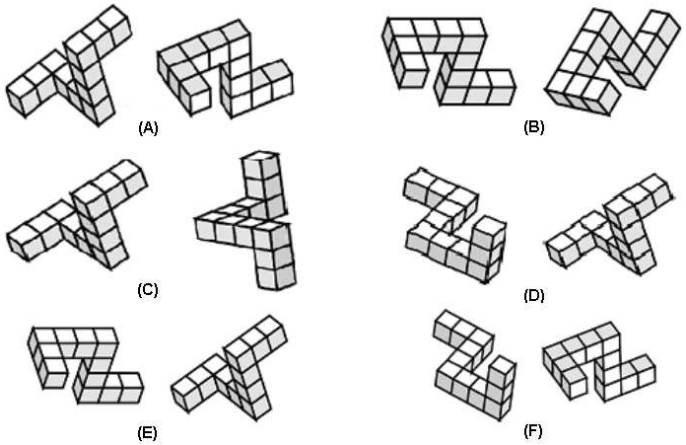

Figure 3.5: The 6 different stimuli the participants were exposed to.

Table 3.2: The differences between the different stimuli

Stimulus	Figure	Patterns	Quadrant(s)
Stimulus 1	3.5 (A)	$different$	$same$
Stimulus 2	3.5 (B)	$same$	$different$
Stimulus 3	3.5 (C)	$same$	$different$
Stimulus 4	3.5 (D)	$different$	$different$
Stimulus 5	3.5 (E)	$different$	$same$
Stimulus 6	3.5 (F)	$same$	$different$

stretch cap with electrodes placed in the 10/20 international system (Jasper - (JASPER, 1958)), shown in figure 3.6, was attached to the subjects head for the EEG recording. 8 different electrodes (channels) were recorded: **FP1, FP2, F3, F4, T3, T4, P3, P4**; followed the guidelines of Luck - use a few electrodes if the interest points are already identified in the scalp (LUCK, 2005).

A water-based conductive gel was used to adjust the electrodes in the participant scalp, resulting in a better quality signal recording. A reference signal was used - a predefined electric current - with the same intensity in each electrode, thus all outputs are guaranteed to have the same amplitude. It is good to say that every channel has the same reference point (System Ground) in both inputs, which are configured to monopolar mode. For minimizing external interference the experiment was conducted in an ambient that follows the concept of a Faraday Cage (JR.; BUCK, 2003).

In figure 3.8 the scheme of the used electroencephalograph can be viewed. Every channel has a set of amplifier circuits with a total gain of 2000 times. Separately: an instrumentation amplifier with a 12.4x gain, an adjustable amplifier that ranges between 2x and 101x gain and the filter gain, adjusted to 84x. Each channel has a high-pass filter with a cutoff frequency of $0.01Hz$, which eliminates the DC signal (constant $0Hz$ frequency signal). A National Instruments 16 bit PCI-MIO-16E-1 board was used to make the Analog to digital conversion. National Instruments LabView (INSTRUMENTS, 2013) is the software used for signal capture, digitizing, digital-filtering, signal processing, programming and visualization. The sampling frequency used was $1.000Hz$ ($1KHz$) (CARRA et al., 2007).

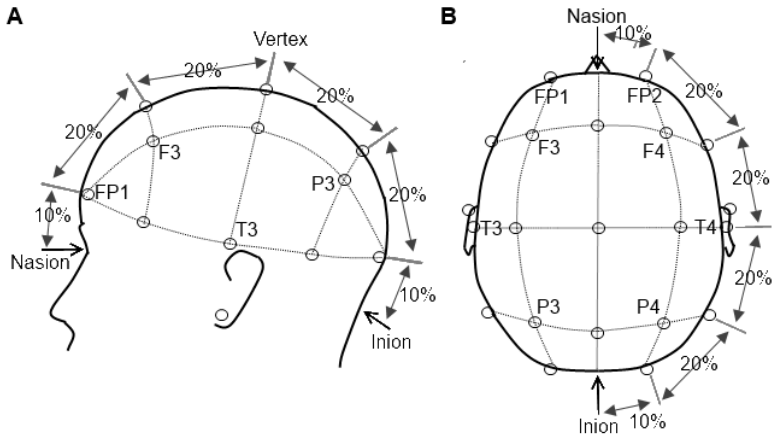

Figure 3.6: International 10/20 electrode placement system. **A)** side view, and **B)** top view.

3.4 Non-linear analysis

Non-linear techniques tend to describe some of the signals generated by biological systems (as the EEG) in a more effective way. The reason behind it: EEG signals are highly non-Gaussian, non-stationary and have a non-linear nature (SUBHA et al., 2010). The choice of using the LLE was done based in the fact that the LLE is related to the energy that is dissipated by the system, if the system is dissipative, like the brain is. More dissipated energy means more brain effort in performing the task. The energy spent is a good measure of the effort done by the brain to perform the task (SPINDOLA, 2010; MARON; BARONE; RAMOS, 2012).

There are statistically relevant differences between the brain effort to perform spatial cognition related tasks between people with an alleged higher degree of development of their spatial cognition and those without any alleged special development of their spatial cognition, as shown in (MARON; BARONE; RAMOS, 2012). Maron, Barone and Ramos used the LLE as the brain effort indicator, and found differences ranging from 5% to 25% (average 15%) between the LLE means of 2 groups: i) a group with an alleged higher degree of development of their spatial cognition, and ii) a control group without any alleged special development of their spatial cognition (MARON; BARONE; RAMOS, 2012).

Figure 3.9 shows a scheme of the process of non-linear analysis used in the present work. In this scheme the TISEAN tools used are discriminate, and its inputs and outputs as well. To calculate the LLEs the raw EEG data need to be prepared, and some channels needed to be discarded (see more details in section 3.5). From every channel the first 150 samples were removed. This follows a general recommendation given in (ABARBANEL et al., 1993), for the right handling of non-linear systems, discarding some of the initial points, where the system did not converge to its attractor. Other reason for removing these samples is that they are located before the P1 (P100) component, which is the moment when your brain receives the sensory input. The measurements done before the P1 (P100) component, which happens after about 100 milliseconds of the sensorial stimuli, represents something not related with the sensorial processing.

The first step to calculate the LLES is the estimation of the time-delay, in samples. The *mutual* program in TISEAN suite (HEGGER; KANTZ; SCHREIBER, 2007) was used for this estimation. After this the embedding dimension was estimated by using the

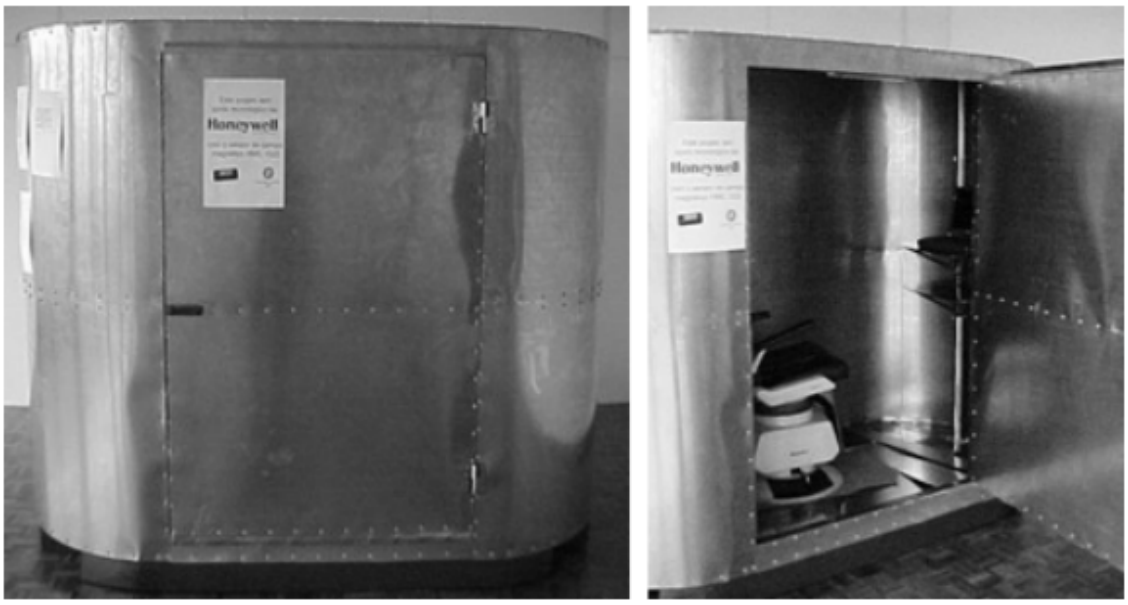

Figure 3.7: The ambient where the EEG measurements were performed, following the concept of a Faraday Cage.

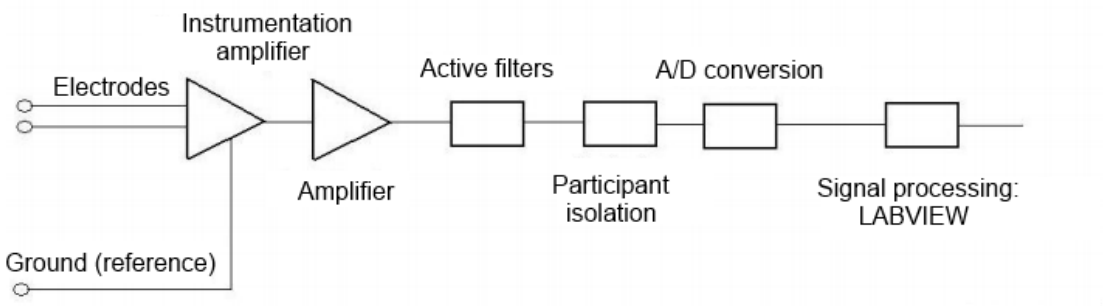

Figure 3.8: Scheme of the electroencephalograph used in this work.

false_nearest program, also from TISEAN suite. The vectors were then reconstructed using *delay*. The last step was calculating the LLEs, which was handed by using the *liap_k*0 program, also from TISEAN suite, with the line fitting (last step required for this program) performed using MATLAB13 curve fitting toolbox (MATHWORKS, 2013).

3.5 Data Sets

There are 3 different relevant data sets in this method, as shown in figure 3.1: i) the EEG data, ii) the calculated LLEs, and iii) the group in which every participant was designed to. These data sets were generated in one of the steps of the experiment, and used as input of the next step.

The raw EEG data was recorded from the participants during the performance of the aforementioned tasks (see section 3.2), in the step names EEG Recording in figure 3.1. The raw EEG data is composed of 8-channel EEGs recordings, sampled at $1.000Hz$ and with 800 data points in each channel. Every stimulus shown to a participant generated an EEG recording.

The LLEs data set was generated by the step named Non-Linear Analysis in figure 3.1.

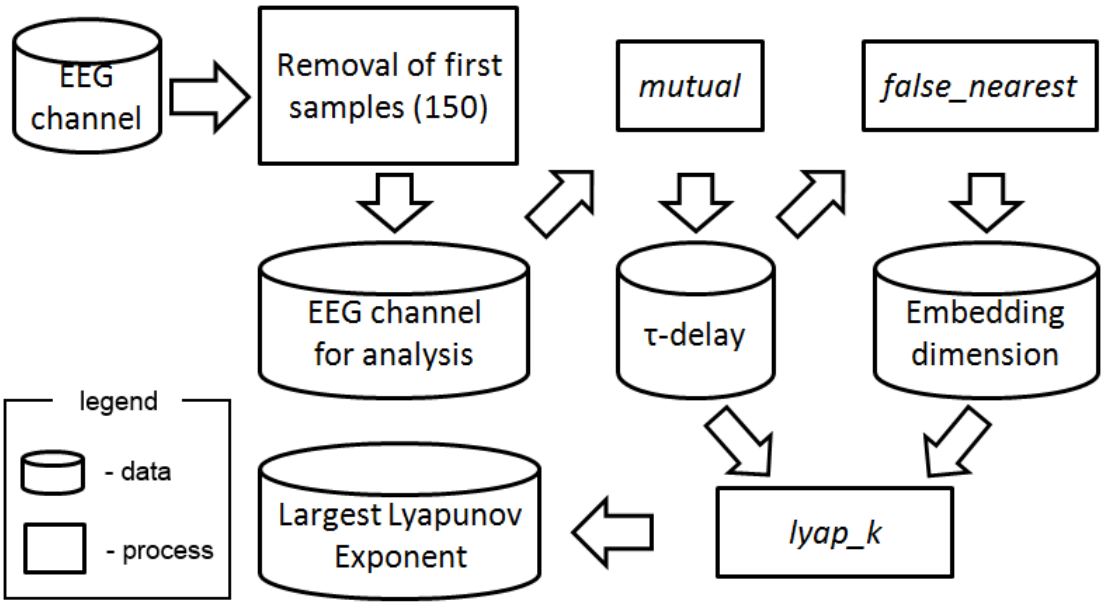

Figure 3.9: Scheme of the non-linear analysis process used, showing the TISEAN tools chain.

Using the EEG data as input for a TISEAN (HEGGER; KANTZ; SCHREIBER, 2007) tools chain the LLEs were calculated as described in section 3.4. Every channel had its LLE calculated, so every LLE data tuple is composed of 8 different LLEs. The table 3.3 shows the statistics for the calculated LLEs, while figure 3.10 shows the histogram and the normal curve which approximates them. In the appendix 1 the histograms and the normal curves which approximate each electrodes are shown.

Table 3.3: Statistics of the LLE data

channel	max.	min.	mean	std. dev.	missing
F3	0.06110	0.00110	0.02831	0.00727	2(0.9%)
F4	0.07250	0.00140	0.02988	0.01082	1(0.45%)
FP1	0.06870	0.00760	0.03178	0.00998	1(0.45%)
FP2	0.06890	0.00630	0.03076	0.01091	3(1.3%)
P3	0.07240	0.00850	0.03309	0.01301	7(3.1%)
P4	0.06330	0.00080	0.02993	0.00740	7(3.1%)
T3	0.06790	0.00750	0.03180	0.01007	1(0.45%)
T4	0.07210	0.01370	0.04760	0.01625	7(3.1%)

In some cases the EEG channel was discarded after the artifact removal, after being visually analyzed by a specialist, so there were some missing data, as shown in table 3.3. The first idea to overcome this problem was using the means of the channel LLE to complete the missing value, but this looks poor, variance would be diminished. Thus a more sophisticated technique was adopted, which awarded more variability in the missing val-

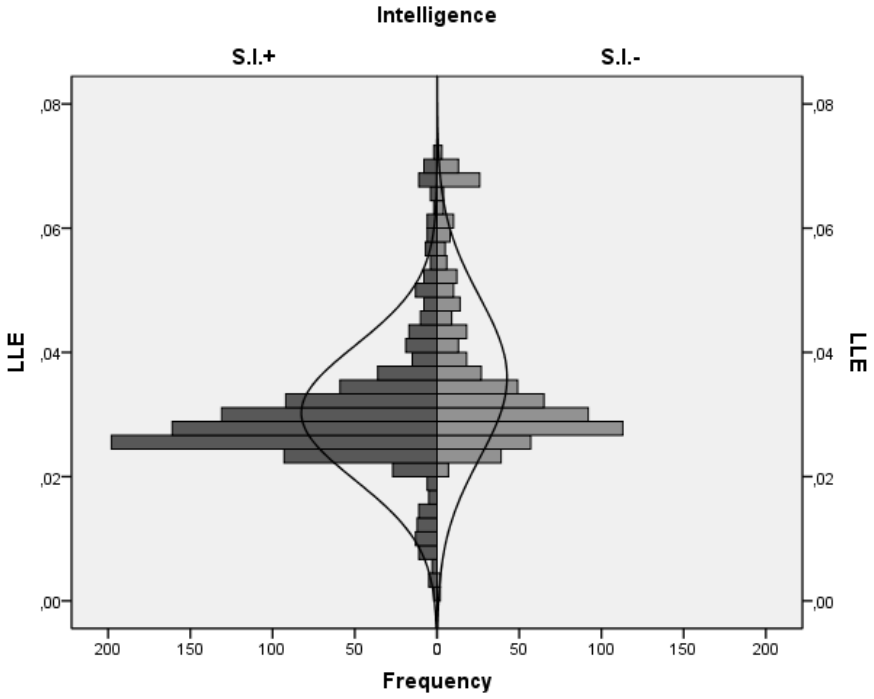

Figure 3.10: Histogram of the calculated LLEs, approximated by a normal curve.

ues: using XMeans clustering, and using the winning centroid value to replace the missing channel LLE value. To chose how many clusters will be used the data was clustered using the XMeans clustering found in WEKA (WAIKATO, 2013). Values ranging from 2 clusters (the minimum possible number of clusters) to 11 clusters $\left(\approx \sqrt{data\ points/2}\right)$, as recommended by Mardia et al. in (MARDIA; KENT; BIBBY, 2008) were tested. In the figure 3.11 we can see the plotting of Bayesian Information Criterion (BIC), and Distortion according to the number of clusters used in the classification. A visual analysis made us decide to use 6 clusters, the number of clusters before a behavior change in BIC value, somehow following the guidelines of Thorndike (THORNDIKE, 1953). Using 6-means XMeans clustering have awarded a BIC of 61.25284 and a distortion of 270.76116.

The step called Group Selection in figure 3.1 used as input the college courses of the participants, and the result of the BPR-5 SR test to validate this classification hypothesis, as described in section 3.1. Then every participant was designed to a group based in his college course: i) SI+ for engineering students, with an alleged higher degree of development of their spatial cognition, and ii) SI- for human and social science students, with no alleged special development of their spatial cognition.

The data tuples used to train and validate the classifiers were composed of a mix of the LLE data and the Group data, as is shown in the step named Classification figure 3.1. The feature values were the LLE and the target value, or class, was the group in which the participant was designed to. The detail of the classification are described in section 3.6.

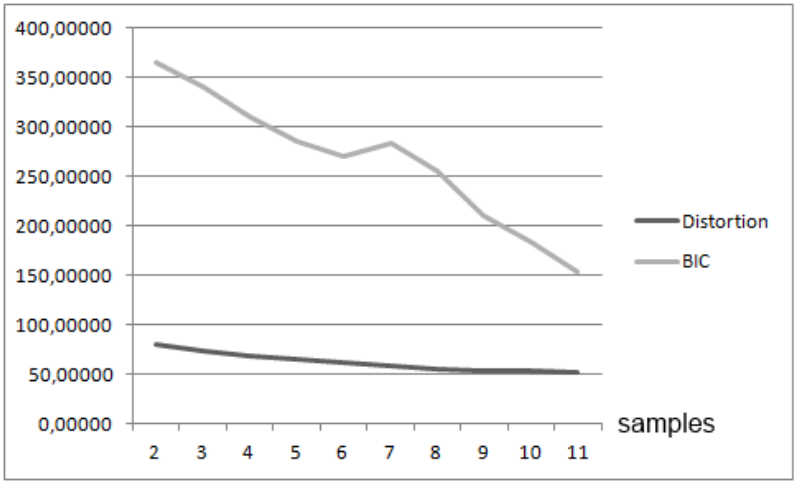

Figure 3.11: BIC and Distortion values for 2 to 11 clusters using the XMeans clustering implementation found in WEKA.

3.6 Classification

3.6.1 Performance metrics

Several different performance metrics were used to evaluate the classifiers tested in the present work. All these metrics are based in the Confusion Matrix, which is a matrix that shows the right and wrong outputs of each classifier. Figure 3.12 shows how a confusion matrix for a 2-classes problem, like the one in the present work, looks like.

Confusion Matrix			
	Class 1	Class 2	
Class 1	T_p	F_p	
Class 2	F_n	T_n	
			accuracy

Figure 3.12: Model for a Confusion Matrix for a 2-classes problem.

The number in T_p cell indicates the True Positive outputs, the outputs that rightly classified instances in the class of interest (Class 1). F_p cell indicates the False Positives, instances that were classified in the interest class (Class 1), but should have been classified in Class 2. F_n cell indicates the False Negatives, the instances that were classified in Class 2, but should have been classified on Class1. Finally there is T_n cell, which indicates the True Negatives, the instances rightly classified in Class 2. From a Confusion Matrix several metrics can be calculated. The ones used in the present work are:

- **Accuracy:** $\frac{(T_p+T_n)}{T}$
- **Error:** $\frac{(F_p+F_n)}{T}$
- **Recall:** $\frac{T_p}{(T_p+F_n)}$

- **Specificity:** $\frac{T_n}{(F_p+T_n)}$
- **Precision:** $\frac{T_p}{(T_p+F_p)}$
- **F1:** $\frac{(2*recall)}{(recall+precision)}$

The standard process for validating the classifiers was the k-times fold validation, with k = 10, usually called 10-times fold validation. The confusion matrix is generated by this process, which consists in creating k copies of the data-set, splitting each copy in k equal partitions. For each set i the i-th partition ($i = [1 \cdots N]$) is taken as test samples. Then the other partitions are used to train the classifier. In figure 3.13 there is an example of a k-times fold validation with k = 3.

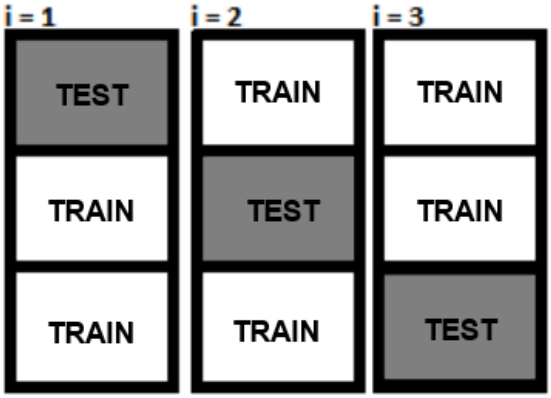

Figure 3.13: An example of a k-times fold validation with k = 3.

3.6.2 Classification process

The last step in the proposed method is the use of a classifier to guess the spatial cognition degree of development of a participant, classifying him in one of the 2 possible groups: **SI+** or **SI-**. In figure 3.14 a simplified view of the proposed classification method can be seen. Starting with the EEGs, the LLEs are calculated, and then used as input for a classifier. The classifier's output is the class of spatial cognition degree of development in which the participant should be in, more detail about the classes can be found in section 3.1. Several different classification algorithms were tested, all the algorithms implemented in WEKA (WAIKATO, 2013) that are capable of handling $\Re$ inputs, but only the best-performing ones are listed, they are:

- Multi-layer perceptron artificial neural network **(MLP)**
- Radial base function artificial neural network **(RBF)**
- Voted perceptron artificial neural network **(VP)**
- Support vector machines **(SVM)**
- K-Nearest Neighbors **(KNN)**

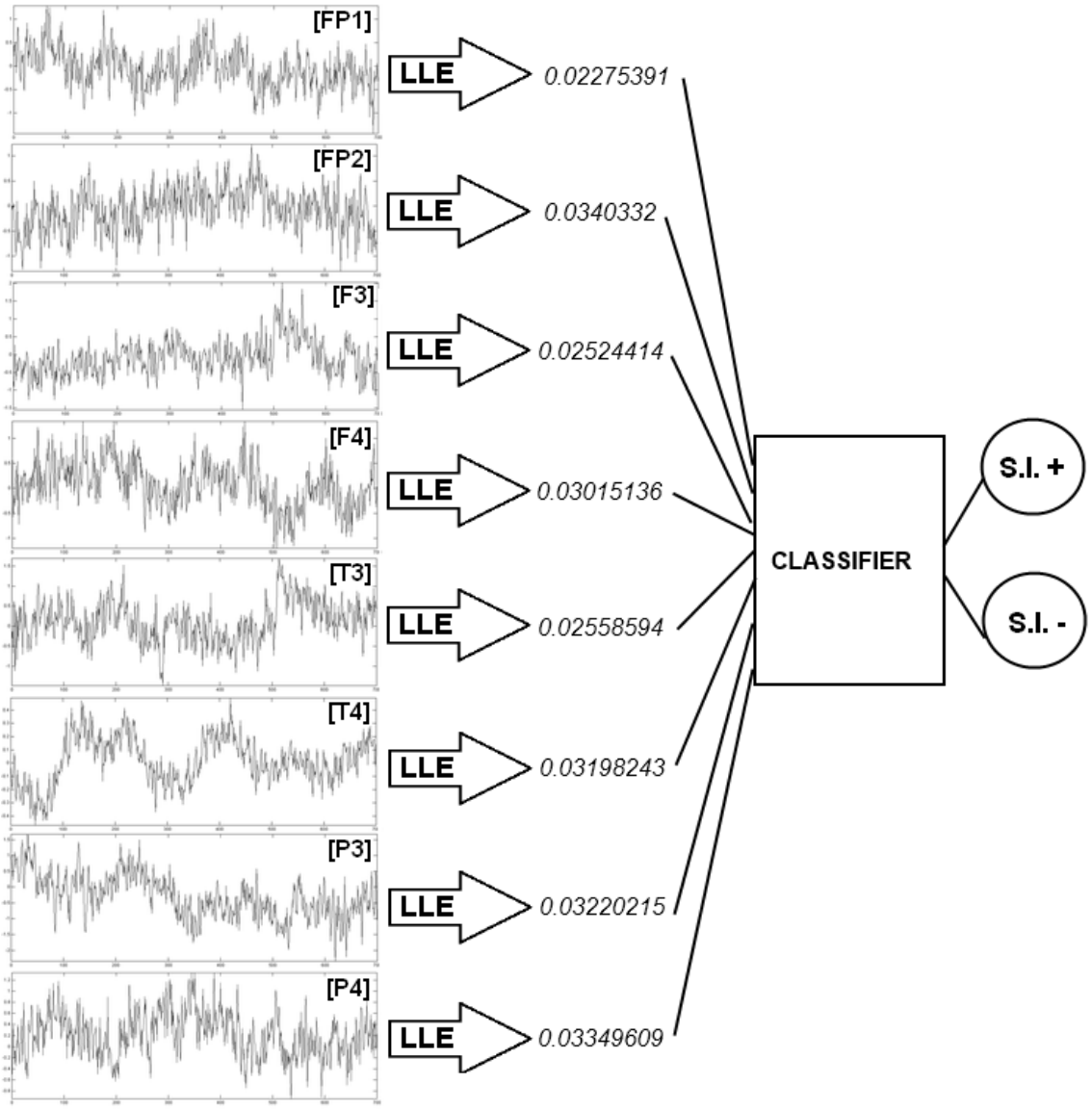

Figure 3.14: A simplified view of the proposed classification method.

The base classifier chosen in the proposed method is the multi-layer perceptron artificial neural network, the most used artificial neural network topology in EEG classification literature (ROBERT; GAUDY; LIMOGE, 2002; SUBHA et al., 2010), and in classification problems using artificial neural networks as well (HAYKIN, 1998). The results obtained using MLP classification were then compared with other classification techniques, regarding their accuracy, sensitivity, F1 measure, and AUC [2]. Accuracy and sensitivity were measured at the point with the highest F1 measure in the ROC curve, the average values for this attributes are also considered. All classification models were validated using 10-times fold validation.

Two different transfer functions were tested: the Hyperbolic Tangent (tansig) [3] , which tend to accelerate the convergence of the MLP (VOGL et al., 1988), and the Logistic Function (logsig) [4] . MLPs using tansig transfer function are not available in WEKA

[2]**AUC:** Area Under (ROC) Curve.
[3]**tansig:** $n = \frac{2}{1+exp(-2*n)} - 1$.
[4]**logsig:** $n = \frac{1}{(1+exp(-n))}$.

(WAIKATO, 2013), but a compatible solution was developed in the scope of this work, and is available at (MARON; BARONE; RAMOS, 2013b). For tuning the MLP parameters different numbers of hidden layers and hidden neurons in each layer were tested, ranging from 1 to 3 hidden layers and from 2 to 18 neurons in each hidden layer. The number of training epochs tested was: 100, 300, 500, and 1000. The learning rates tested was $1 * 10^{-3}$, $1 * 10^{-4}$, and $1 * 10^{-5}$. A t test succeeded to reveal a statistically reliable difference between the accuracy of logsig and tansig MLPs: logsig (M = 69.01%, s = 8.04%) and tansig (M = 70.20%, s = 6.32%), t(408) = 1.667, p = 0.006, α = 0.05.

For testing the K-Nearest Neighbors (K-NN) algorithm the 3 available weight ponderation options were tested: $1 - distance$ (similarity), $1/distance$ (inverse), and no weighting at all. The percentages of nearest neighbors (k) to evaluate were: 1%, 5%, 10%, 20%, 25%, 33%, 50%, 75%, 80%, 90% and 100%. The distance function used was the Euclidean distance. Only the best result achieved for each classifier was kept. The K-NN implementation, named IBK, available in WEKA (WAIKATO, 2013) suite was used. A t test succeeded to reveal a statistically reliable difference between the precision of these classifiers: MLP-tansig (M = 70.20%, s = 6.32%) and K-NN (M = 73.42%, s = 8.87%), t(408) = -4.330, p = 0.000, α = 0.05.

Support Vector Machine (SVM) was tested too. Different Kernel functions and different parameters were evaluated. The kernel functions evaluated were: Polynomial kernel [5] , with p parameter ranging from 1 to 10; Normalized Polynomial kernel [6] with p parameter ranging from 1 to ten; Radial Base Function (RBF) kernel [7] , withγ assuming the values: 0.001, 0.01, 0.1, 0.5, and 1.0; and PUK kernel [8] , with ω and σ assuming the values: 0.001, 0.01, 0.1, 0.5, and 1.0 to 10.0 (1.0 increment). The parameter c - the complexity - assumed the values: 0.1, 0.5, 1.0 to 10.0. The parameter ϵ - for round-off error - as recommended by WEKA, was not altered. Only the best result achieved was kept. The SVM implementation present in WEKA (WAIKATO, 2013) was used; it is called SMO within the WEKA suite. A t test succeeded to reveal a statistically reliable difference between the precision of these classifiers: MLP-tansig (M = 70.20%, s = 6.32%) and SVM (M = 42.59%, s = 36.95%), t(207) = 6.565, p = 0.000, α = 0.05.

For testing Radial Base Function artificial neural network (RBF) the number of clusters ranged from 2 to 11, using the same criteria used in the clustering performed to complete the missing data (section 3.5). The parameter ridge ranged from $1 * 10{-}^{1}$ to $1 * 10^{-10}$. The minimum standard deviation within a cluster was varied from 0.1 to 0.001. Only the best results achieved with each number of clusters were kept. The RBF network available in WEKA (WAIKATO, 2013) suite was used.

Finally Voted Perceptron (VP) was tested. The values for the exponent of the Polynomial kernel ranged between 1 and 10 (1 increment), then from 10 to 100 (10 increment), and then from 100 to 500 (100 increment). The maximum number of alterations in a perceptron ranged from 100 to 1000 (100 increment), and from 1000 to 10,000 (1000 increment). The number of iterations ranged from 1000 to 10,000 (1000 increment), and from 10,000 to 100,000 (10,000 increment). The VP implementation present in WEKA was used. For more information about VP (FREUND; SCHAPIRE, 1998, 1999) are recommended. A t test succeeded to reveal a statistically reliable difference between the

[5] **Polynomial kernel:** $K(x, y) =< x, y >^p$ or $K(x, y) = (< x, y > +1)^p$; p is an arbitrary exponent.

[6] **Normalized Polynomial kernel:** $K(x, y) = \frac{<x,y>}{\sqrt{<x,x><y,y>}}$, where $< x, y >= PolynomialKernel(x, y)$.

[7] **RBF kernel:** $K(x, y) = e^{-(\gamma * <x-y, x-y>^2)}$, γ is an arbitrary parameter.

[8] **PUK kernel:** for references see (UESTUEN; MELSSEN; BUYDENS, 2006).

precision of the 2 best performing Artificial Neural Networks tested: MLP-tansig (M = 70.20%, s = 6.32%) and RBF (M = 72.13%, s = 9.67%), t(430) = 2,446, p = 0.000, α = 0.05.

Figure 3.15 shows the pipeline for testing the MLP, SVM and KNN classifiers. Figure 3.16 shows the pipeline for testing the MLP, RBF and VP ANNs. Both images are of the Weka Knowledge Flow interface (WAIKATO, 2013).

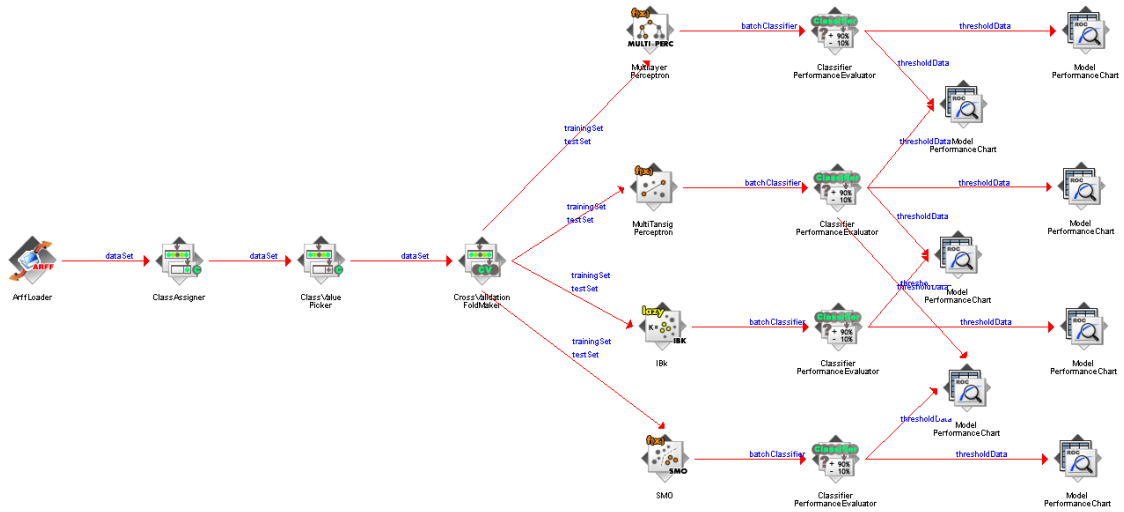

Figure 3.15: Pipeline (Knowledge flow) for testing MLP, SVM, and KNN classifiers.

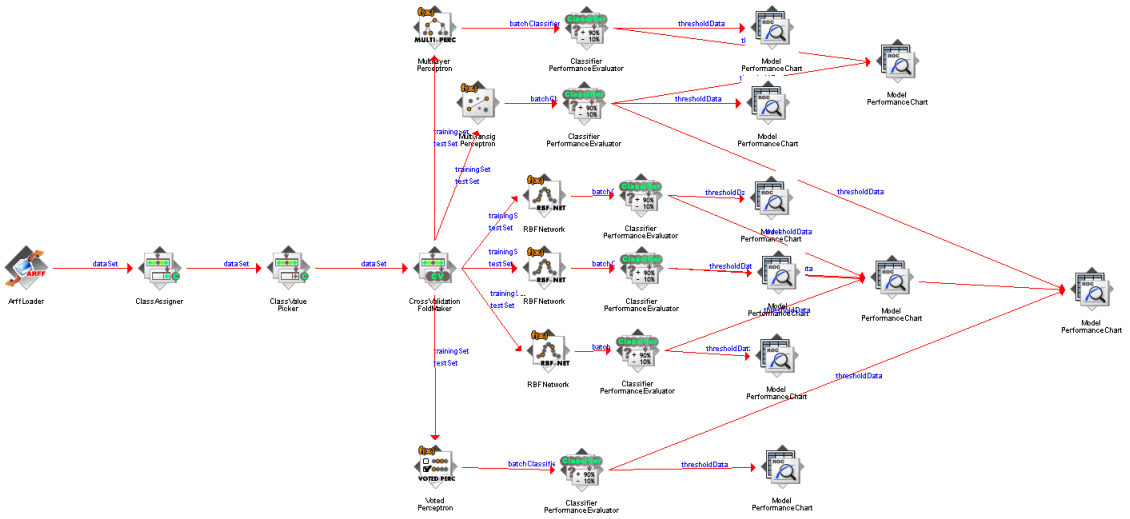

Figure 3.16: Pipeline (Knowledge flow) for testing MLP, RBF, and VP classifiers.

4 RESULTS

4.1 Statistical analysis

The fundamental question in the present dissertation is: It is possible to tell the spatial cognition degree of development from one individual by analyzing his EEG during a mental image rotation task? But before asking the question related more with classification algorithms a more basic question should be done: Is there a statistically relevant difference in the brain activity of individuals with an higher spatial cognition degree of development compared with individuals without any alleged spatial cognition development?

A first analysis that can be seen figure 4.1, which shows the LLE means for all subjects, one can clearly see that the subsets of subjects **S.I.+**, **M+** and **F+** have lower LLE means. This shows that their reconstructed attractors have slower expansion, dissipating less energy, implying less brain effort to perform the aforementioned task.

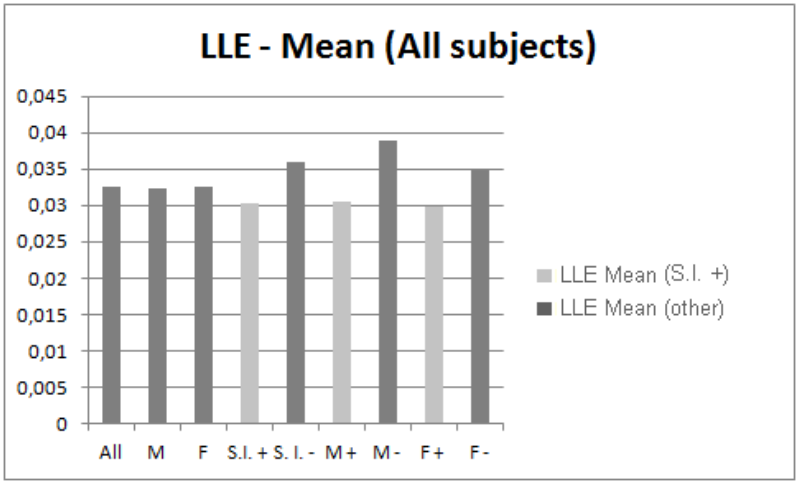

Figure 4.1: LLE means for all subjects.

Figure 4.2 shows a box-plot for the LLE data, allowing an easier visualization of the data general tendencies. One can easily see that SI- group have some sparse low LLE results, but the bulk of the group tends towards being above the mean, while SI+ group tend to be more equilibrated.

The affirmations done in the previous paragraphs are sustained by the evidences showed in figures: 4.1 and 4.3, and in table 4.1. One can easily see an equal qualitative behavior of the LLE means comparing **S.I.+** and **S.I-** groups with the other ones. And in every single electrode there is an akin behavior of lower LLE means in the **S.I.+** group and higher LLE means in the **S.I.-** group. Therefore there is the same qualitative behavior both in the

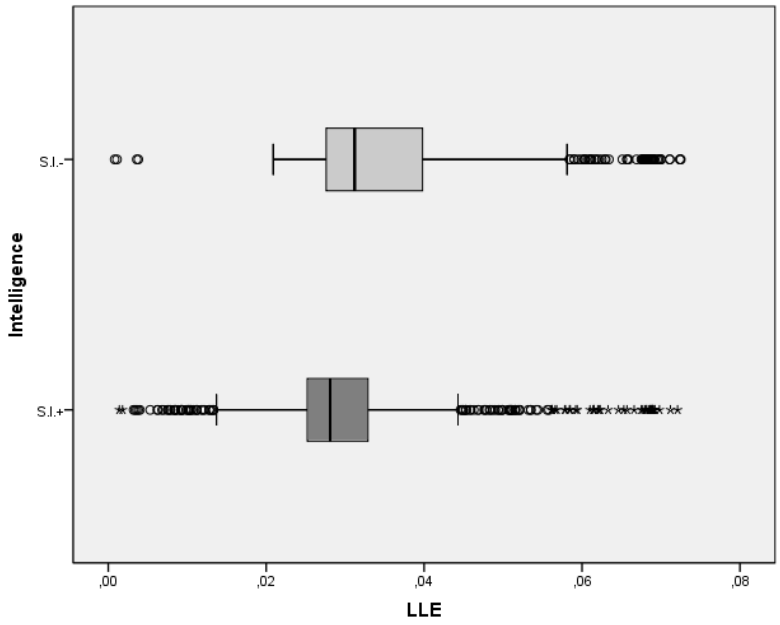

Figure 4.2: A boxplot showing the LLE means in both classes.

system as a whole and locally, in each one of the electrodes.

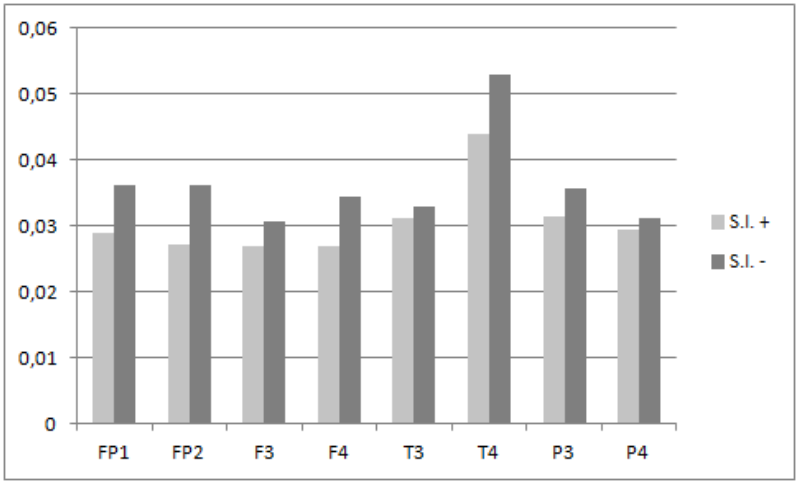

Figure 4.3: **S.I.+** and **S.I.**subjects LLE means in each electrode.

A t test was performed over the data, testing if the subsets of subjects have the same variance in their LLE. Thus the null hypothesis is that both subsets have the same variance. A 95% confidence level was used to perform the Variance Test. The table showed in figure 4.4 proves the initial hypothesis in a statistical fashion: the rejection of the hypothesis of non-statistically relevant differences at a 95% confidence level, therefore, both groups LLEs distributions are statistically relevantly different.

Then an ANOVA Variance Analysis was performed over the LLE means of both groups, using IBM SPSS Statistics software (SOFTWARE, 2013). The result is showed in figure 4.5, a Sigma (Sig. column) of 0.000, which is lower than the critical 0.05. Therefore we have statistical relevant differences between the intellects at a 95% confidence level.

Table 4.1: Differences between **S.I. +** and **S.I. -** LLE means

	S.I. + LLE-M	**S.I. -** LLE-M	difference	$\approx \%$
ALL	0.03027	0.03595	0.00568	16%
FP1	0.02901	0.03621	0.00720	20%
FP2	0.02727	0.03629	0.00902	25%
F3	0.02684	0.03069	0.00385	13%
F4	0.02699	0.03451	0.00752	22%
T3	0.03110	0.03291	0.00181	5%
T4	0.04396	0.05301	0.00905	17%
P3	0.03130	0.03572	0.00442	12%
P4	0.02931	0.03105	0.00174	6%

[ALL] S.I + x S.I. -			
Descriptive Statistics			
VAR			
Mean	0,030228486	0,035952556	
Variance	0,00011496	0,000170022	
Standard Deviation	0,010721955	0,013039241	
Mean Standard Error	0,000338382	0,000521153	
Summary			
F	1,478960628	F Critical value (5%)	1,124876651
p-level 1-tailed	1,90E-08	p-level 2-tailed	3,79E-08
H0 (5%)?	**rejected**		

Figure 4.4: t-test over the LLE means of SI+ and SI- groups. The null-hypothesis of both groups having the same variance is rejected.

The right hemisphere of the brain is said to have a greater activation during 3-D image processing (ROBERTS; BELL, 2000, 2003; GAZZANIGA; IVRY; MANGUN, 2006). Figure 4.6 shows it: there is greater activation of the Temporal (T3, T4) electrodes, being T4 the most activated in the process. The sum of the LLE means in the left hemisphere of the brain is 0.12497, the sum in the right hemisphere is 0.13816, a difference of 0.01319, approximately 9.55%. Figure 4.7 shows the hemisphere differences between S.I.+ group A): with left hemisphere sum of LLE means equal to 0.11827, right hemisphere equal to 0.12755, a difference of 0.00928, approximately 7.85%. Figure 4.7 (B) shows the hemisphere differences for S.I.- group: with left hemisphere sum of LLE means equal to 0.13555, right hemisphere equal to 0.15487, a difference of 0.01932, approximately 15.25%. The summarization of these result are shown in figure 4.8, where the electrodes, and the LLE means difference between SI+ and SI- groups are displayed.

4.2 Classification

With the showing of the statistically relevant differences in brain activity between **SI+** and **SI-** groups while performing a mental image rotation task now is possible to return to the main question of this dissertation: It is possible to tell the spatial cognition degree of development from one individual by analyzing his EEG during a mental image rotation task?

ANOVA

LLE

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	,012	1	,012	90,854	,000
Within Groups	,223	1629	,000		
Total	,236	1630			

Figure 4.5: ANOVA analysis: S.I+ and S.I.- groups. Sigma is lower then 0.05 (confidence level), therefore there is a statistical relevant difference between both groups.

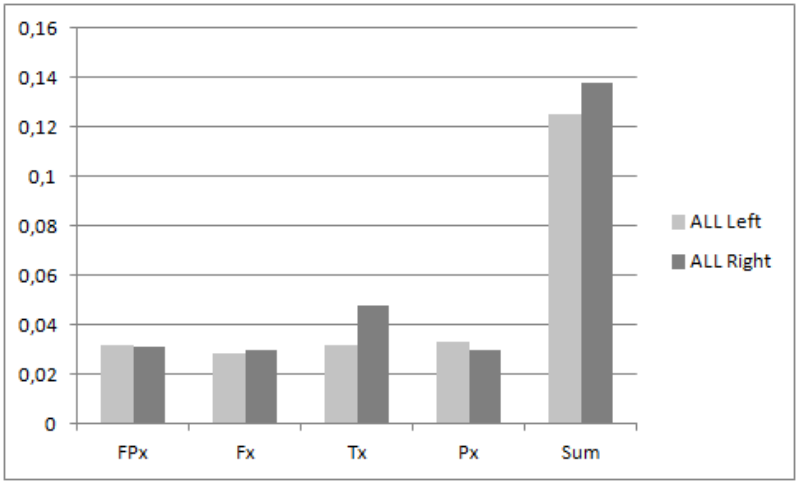

Figure 4.6: Differences between left and right scalp-side electrodes LLEs.

In section 3.6 in chapter 3 the chosen classifiers are shown:

- **Multi-layer perceptron (LLE):** the base classifier, based on literature of EEG analysis.
- **Radial Base Function ANN (RBF):** the best performing classifier in the present problem.
- **Voted Perceptron ANN (VP):** very fast to train.
- **Support Vector Machine (SVM):** an algorithm that finds the vectors supporting a decision hyper-plane, usually deformed by applying kernel functions over these.
- **K-Nearest Neighbors (KNN):** an algorithm that uses the K-nearest data points features distance to classify points in different classes.

As said before, for the sake of the comparison that are going to be shown in the present chapter, only the best results were kept. These results were obtained through testing a large array of different parameters configuration within each classifier, as detailed in section 3.6 in chapter 3. The base classifier chosen in the proposed method is the multi-layer perceptron artificial neural network (MLP), the most used artificial neural network topology in EEG classification literature (ROBERT; GAUDY; LIMOGE, 2002; SUBHA et al., 2010), and in classification problems using artificial neural networks as well (HAYKIN, 1998).

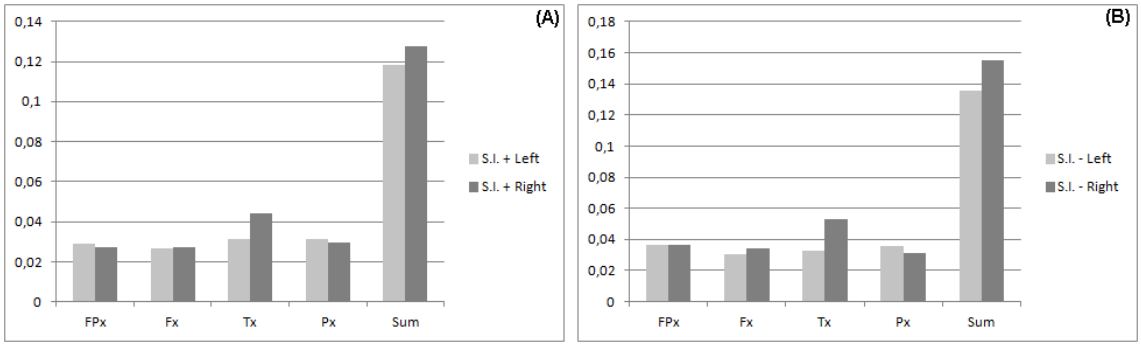

Figure 4.7: Differences between left and right scalp-side electrodes in **(A)** S.I.+ group and **(B)**: S.I.- group.

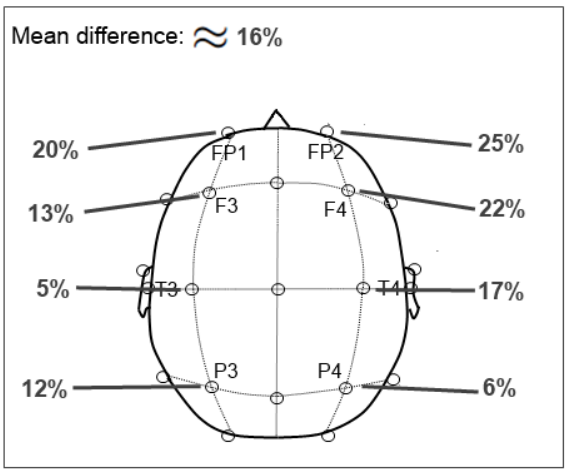

Figure 4.8: The approximated differences between the LLE-means of **SI+** and **SI-** groups in each electrode.

So starting with the base classifier, the MLP, in previous studies, some good results were obtained by using a MLP in MATLAB (MATHWORKS, 2013). When a decision was made, use the classifiers present in WEKA, to ease the porting of the entire method in a single Java Tool, because WEKA is an open-source JAVA software (WAIKATO, 2013). But a problem arose from this decision; the results were not as good as the ones in MATLAB. Some of this issue happens because JAVA and MATLAB represent numbers using different notations and precision. Other issue is that WEKA only gave one choice for the transfer function in the MLP neurons, the logsig function. To tackle with this problem a new classifier was developed in the scope of the present work, and is available in (MARON, 2013). So the classification process was performed using 2 different MLPs: i) using logsig transfer function and ii) using tansig transfer function. The ROC curves showing the precision of the MLP classifiers are displayed in figure 4.9. In both MLPs the best results were achieved with the following parameters:

- **Hidden layers:** 1
- **Hidden neurons:** 18
- **Learning Rate**: 0.001
- **Momentum:** 0.0005

- **Training Epochs:** 500

A t test succeeded to reveal a statistically reliable difference between the precision of logsig and tansig MLPs: logsig (M = 69.01%, s = 8.04%) and tansig (M = 70.20%, s = 6.32%), t(408) = 1.667, p = 0.006, α = 0.05.

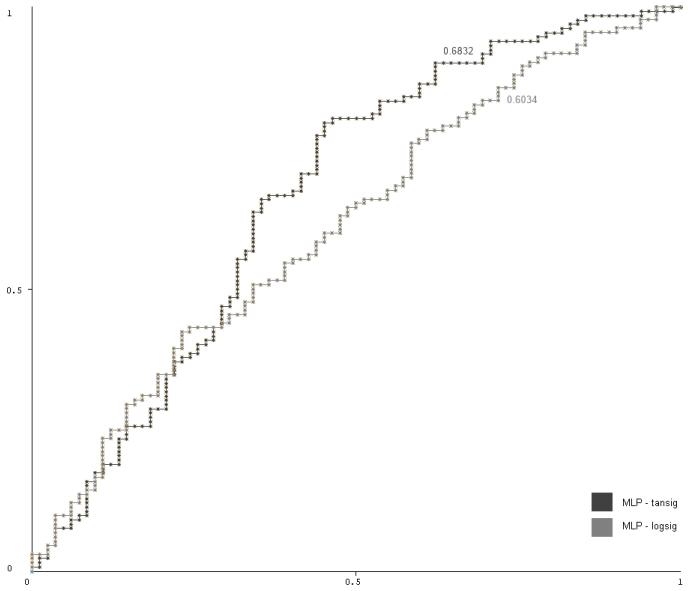

Figure 4.9: ROC curves: MLP - tansig (AUC = 0.6832) and MLP - logsig (AUC = 0.6034).

The second set of results to be shown are the ones obtained by using a K-Nearest Neighbors Algorithm (KNN). Its implementation in WEKA is called IBK. The ROC curves showing the performance of the MLP-tansig classifier (base) and the K-NN classifier are displayed in figure 4.10. The best results were obtained using the following parameters:

- **K-neighbors:** 20%
- **Distance-weighting:** $1 - distance$ (similarity)
- **Neighborhood Search Algorithm:** $LinearNNSearch$ (but all results were very close)
- **Distance function:** Euclidean Distance

A t test succeeded to reveal a statistically reliable difference between the precision of these classifiers: MLP-tansig (M = 70.20%, s = 6.32%) and K-NN (M = 73.42%, s = 8.87%), t(408) = -4.330, p = 0.000, α = 0.05.

Other classifier that showed good results in the previous studies, thus is been showed here, were the Support Vector Machines (SVM). In WEKA its implementation is called SMO. The ROC curves showing the performance of the MLP-tansig classifier and the SVM classifier are displayed in figure 4.11. The best results using SVM were obtained using the following parameters:

- **Kernel Function:** Puk Kernel

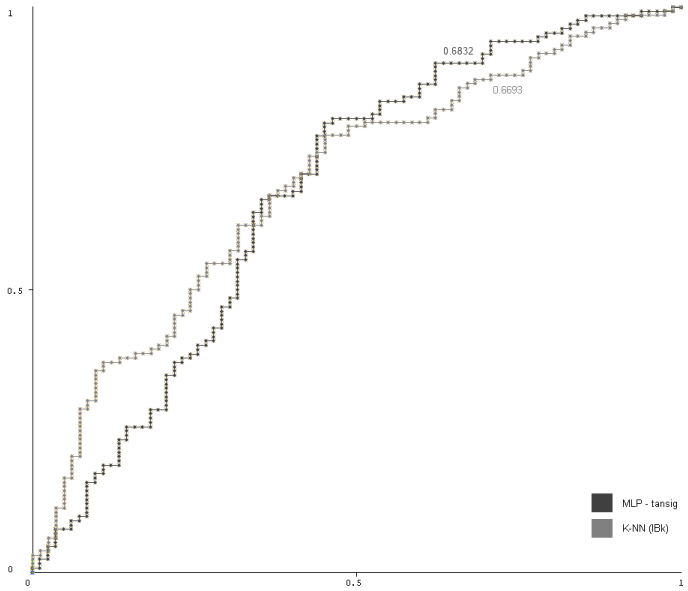

Figure 4.10: ROC curves: MLP - tansig (AUC = 0.6832) and K-NN algorithm (AUC = 0.6693).

  - ω**:** 0.5
  - σ**:** 0.1

- **Complexity:** 3
- ϵ**:** 0.001

A t test succeeded to reveal a statistically reliable difference between the precision of these classifiers: MLP-tansig (M = 70.20%, s = 6.32%) and SVM (M = 42.59%, s = 36.95%), t(207) = 6.565, p = 0.000, α = 0.05.

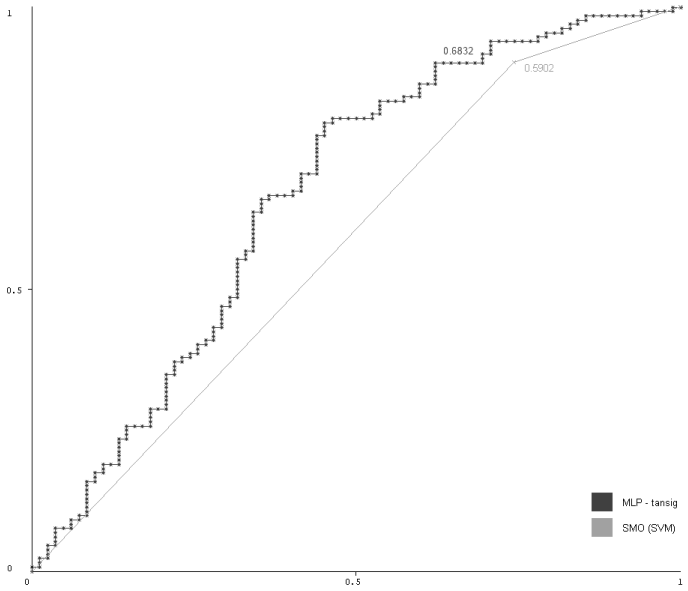

Figure 4.11: ROC curves: MLP - tansig (AUC = 0.6832) and SVM algorithm (AUC = 0.5902).

Radial Base Function Artificial Neural Networks (RBF) were also widely used in EEG analysis works (ROBERT; GAUDY; LIMOGE, 2002; SUBHA et al., 2010). For choosing the number of clusters to use the same criteria used to handle the missing data, as can be seen in section 3.5 from chapter 3, thus, the tested values ranged from 2 to 11; for the sake of clarity only the best 3 results with different number of clusters (2, 4, and 6 clusters) are shown in figure 4.12. The best results (with 4 clusters, it varied using deferent clusters numbers) were achieved using the following parameters:

- **Clusters number:** 4
- **Ridge:** $1 * 10{-}3$
- **Minimum Standard Deviation (for the clusters):** 0.1

The ROC curves showing the performance of the 3 best cluster numbers - 2, 4, and 6 - for the RBF are displayed in figure 4.12.

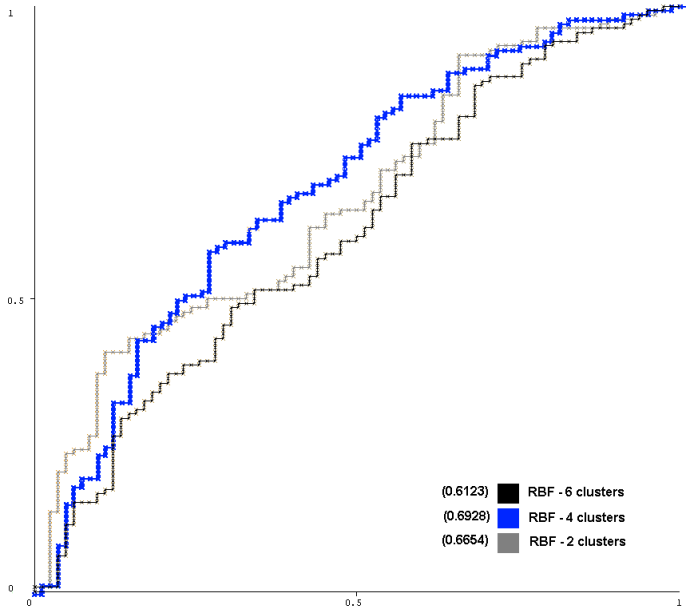

Figure 4.12: ROC curves: RBF networks - 2 (AUC = 0.6654), 4 (AUC = 0.6928) and 6 (AUC = 0.6123) clusters.

Voted Perceptron (VP) was the last artificial neural network classifier tested. Its training is very fast, and it seemed to be promising to deal with real-time classification. In truth its AUC was higher than the SVM, with the same strange 0% of specificity in the point with the higher F1-measure, but the F1 measures for VP were by far the lower ones. The best results were achieved by using the following parameters:

- **Exponent (for polynomial kernel):** 5
- **Maximum iterations:** 10.000
- **MaxK (iterations per neuron):** 500

The ROC curves showing the performance of the MLP-tansig classifier and the best RBF network model (4 clusters), and the best VP model are displayed in figure 4.13. A

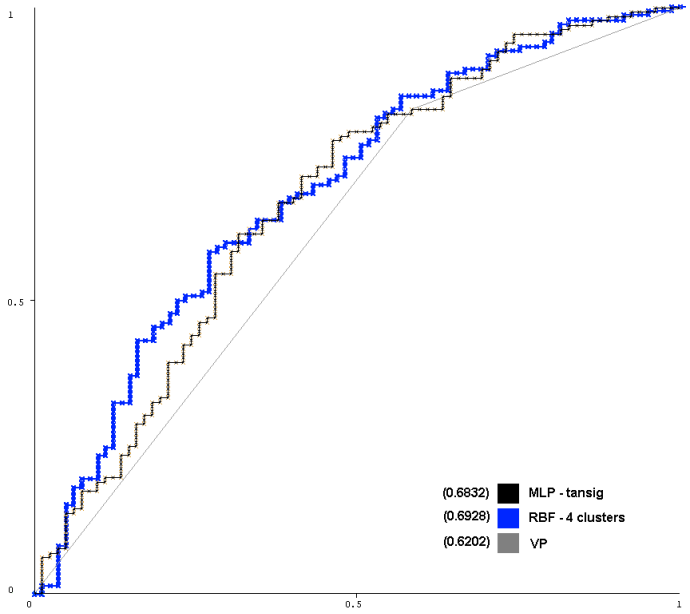

Figure 4.13: ROC curves: MLP - tansig (AUC = 0.6832), RBF with 4 clusters (AUC = 0.6928), and VP (AUC = 0.6202).

t test succeeded to reveal a statistically reliable difference between the precision of the 2 best performing ANN classifiers: MLP-tansig (M = 70.20%, s = 6.32%) and RBF (M = 72.13%, s = 9.67%), t(430) = 2,446, p = 0.000, $\alpha = 0.05$.

The hypothesis of classifying every individual in **SI+** group, the largest one, will achieve 61.68% accuracy, 0% specificity and a F1 of 0.7630. The MLP (base) and RBF (the best performing algorithm) results were superior in all 4 criterions: accuracy, specificity, F1 measure and AUC then the other classifier tests. The hypothesis of using MLP artificial neural networks was a good choice, but using RBF was even better. The results achieved by the classifiers tested in our method are show in table 4.2, the accuracy and specificity were measured at the point with the highest F1 value in the ROC curve; also an average is shown too. The classifiers in table 4.2, are ordered from the best performing to the worst performin. A t test succeeded to reveal a statistically reliable difference between the precision of the 2 best performing Artificial Neural Networks tested: MLP-tansig (M = 70.20%, s = 6.32%) and RBF (M = 72.13%, s = 9.67%), t(430) = 2,446, p = 0.000, $\alpha = 0.05$.

Table 4.2: Classification results, values after the / are averages.

classifier	**accuracy%**	**specificity%**	**F1**	**AUC**
RBF - 4 cl.	67.28 / 59.07	18.29 / 61.83	0.786 / 0.581	0.692
MLP-tansig	69.15 / 57.91	29.26 / 60.32	0.789 / 0.567	0.683
MLP-logsig	63.21 / 54.83	36.58 / 50.81	0.739 / 0.515	0.603
K-NN(IBK)	64.01 / 58.29	08.53 / 60.82	0.771 / 0.576	0.669
SVM(SMO)	61.68 / 55.14	00.00 / 41.87	0.763 / 0.508	0.590
VP	27.07 / 22.29	00.00 / 52.84	0.015 / 0.013	0.620

In this chapter the results obtained in the present work were shown. The chapter starts with a statistical analysis that proves the existence of statistical relevant differences between the brain activity (measured by the LLEs) of **SI+** and **SI-** groups while performing a mental image rotation task. After this the performance of different classifiers: RBF, VP and MLP (using tansig and logsig transfer functions) artificial neural networks, SVM and K-NN algorithm are shown regarding their accuracy, specificity. F1 measure and AUC; combined with statistical relevancy test between their precision.

Next chapter will show the conclusions, discussions and future possibilities that have arose from the results obtained in this work.

5 CONCLUSIONS AND FUTURE WORKS

This work proposed, developed and tested a method for classifying the spatial cognition degree of development of individuals. And also a statistical analysis of the differences in brain activity between both interest groups during a mental tridimensional image rotation task was performed.

The theoretical basis for the method are found in chapter 2, where an overview of non-linear analysis techniques used, classification algorithms and spatial cognition is shown.

In chapter 3 the proposed, developed and tested method was presented in details, describing each of its parts:

- The participants of the experiment.
- The groups of interest in which they were classified (**SI+** and **SI-**).
- The BPR-5 SR psychological test applied to them to strengthen the classification hypothesis.
- The mental tridimensional image rotation task the participants performed.
- The EEG recording equipment used.
- The non-linear analysis techniques applied over the EEGs.
- The data sets generated.
- The classification process: the choosing of classifiers and parameters.

Chapter 4 showed the results achieved, in performing the statistical analysis of the brain activity differences between both groups while performing mental tridimensional image rotation tasks. Also this chapter shows the results obtained in the classifications of the spatial cognition degree of development from the participants, and the parameters used to obtain them using each classifier.

The results found in the statistical analysis of the LLE means confirmed the findings in (ROBERTS; BELL, 2000, 2003; GAZZANIGA; IVRY; MANGUN, 2006), that the right hemisphere of the brain is more activated during tridimensional mental image rotation tasks. The relevant difference in the brain effort among the participants does not vary based on their gender. It is possible to see in figure 4.1 (section 4.1 from chapter 4) that the difference in LLE means between male and female individuals are small (and not even statistical relevant). The difference is between the interest groups: **SI+** and **SI-**, is qualitatively the same while considering the male individuals in **SI+** (**M+**) versus the male

individuals in **SI-** (**M-**) or the female individuals in both groups: **F+** and **F-**, respectively. A deeper investigation in this aspect is out of the scope of the present work.

The brain activity differences found are akin in the system as a whole, and locally, in each electrode, with a mean of approximately 16% difference between both groups. The larger differences were found in FP1 and FP2 electrodes, but the largest LLE mean was found in P4 electrode.

The most challenging point in the task of classifying the spatial cognition degree of development of an individual is the big number of false-positives. Specificities above 50% rarely were obtained in other configurations than the best ones using each classifier. While ANNs and KNN were somehow able to deal with this challenge, the SVM had its results diminished by the high number of false-positives, reaching only 41.87% mean specificity, while all other algorithms surpassed the 50% mark.

The hypothesis of classifying every individual in **SI+** group, the largest one, will achieve 61.68% accuracy, 0% specificity and a F1 of 0.7630. The MLP (base classifier) and RBF (the best performing algorithm) results were superior in all 4 criterions: accuracy, specificity, AUC, and F1 measure then the other classifier tests. The hypothesis of using MLP artificial neural networks was a good choice, but using RBF was even better.

To achieve the best accuracy a 1 hidden-layer MLP, with 18 neurons in its hidden layer was employed. But the best specificity was achieved by a 3 hidden-layers MLP, with 18 neurons in each hidden layer. The time spent to train the 3-layered MLP does not compensate this small gain of specificity, hindered even more for the loss of accuracy that happens. But, in mean, the best accuracy was achieved by using a RBF with 4 clusters. A t test succeeded to reveal a statistically reliable difference between the precision of the 2 best performing classifiers tested: MLP-tansig (M = 70.20%, s = 6.32%) and RBF (M = 72.13%, s = 9.67%), t(430) = 2,446, p = 0.000, $\alpha = 0.05$.

Other contribution of the present work was the development of a new classifier for WEKA open-source suite (WAIKATO, 2013). In WEKA there was only the option to use logsig as transfer function in the MLP classifier. In the scope of the present work a MLP using logsig was implemented for WEKA, and can be found in (MARON, 2013).

5.1 Future works

For future works there are some possibilities that look very promising:

- To investigate the results of using subsets of the electrodes, instead of all of them at the same time
- To investigate the classes distribution and how it affects the results
- To use different non-linear features, like:
  - Fractal Dimension
  - Correlation Dimension
  - Hurst Exponent
  - Entropies
- To try to use more features as input, simultaneously, to enhance the classification rates, linear and non-linear ones.

- To resample the EEGs and check the differences in the calculated features.
- To use a classification committee, to tackle with the better aspect of each algorithm, or with the ones that deal better with each non-linear feature input.
- To use different techniques like, for example, wavelets.
- To port the entire method to a single JAVA tool, making it available for the cognitive investigation researchers.
- To apply variations of the proposed method with different tasks and cognitive profiles.

REFERENCES

ABARBANEL, H. D. et al. The analysis of observed chaotic data in physical systems. **Rev.Mod.Phys.**, [S.l.], v.65, p.1331–1392, 1993.

ACHARYA, R. et al. Non-linear analysis of EEG signals at various sleep stages. **Computer methods and Programs in Biomedicine**, [S.l.], v.80, p.37–45, 2005.

ACHARYA, U. R. et al. Heart rate Variability: a review. **Med. Biol. Eng. Comput.**, [S.l.], v.44, p.1031–1051, 2006.

AHA, D. W.; KIBLER, D.; ALBERT, M. K. Instance-based learning algorithms. In: MACHINE LEARNING, 1991. **...** [S.l.: s.n.], 1991. p.37–66.

AHLSTRÖM, C. et al. Chaotic dynamics of respiratory sounds. **Chaos, Solitons & Fractals**, [S.l.], v.29, n.5, p.1054–1062, 2006.

BACKER, E.; JAIN, A. K. A Clustering Performance Measure Based on Fuzzy Set Decomposition. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, [S.l.], v.PAMI-3, n.1, p.66–75, 1981.

BAKER, J. **50 Physics Ideas You Really Need to Know**. [S.l.]: Quercus, 2007.

BASAR, E. The theory of the whole-brain-work. **International Journal of Psychophysiology**, [S.l.], v.60, n.2, p.133 – 138, 2006.

BISHOP, C. M. **Neural Networks for Pattern Recognition**. New York, NY, USA: Oxford University Press, Inc., 1995.

BURGES, C. J. A Tutorial on Support Vector Machines for Pattern Recognition. http://research.microsoft.com/pubs/67119/svmtutorial.pdf. **Data Mining and Knowledge Discovery**, http://research.microsoft.com/pubs/67119/svmtutorial.pdf, v.2, p.121–167, 1998.

CAMBEL, A. B. **Applied Chaos Theory**: a paradigm for complexity. [S.l.]: Academic Press, 1992.

CARRA, M. et al. Desenvolvimento de um Protótipo EEG como Ferramenta para Caracterização de Sinais Cerebrais em Atividades Relacionadas a Raciocínio Lógico. In: ENCONTRO NACIONAL DE BIOMECâNICA, 2., 2007. **Actas...** IST Press, 2007. p.387–392.

CERUTTI, S. et al. Non-linear algorithms for processing biological signals. **Computer Methods and Programs in Biomedicine**, [S.l.], v.51, n.1 - 2, p.51 – 73, 1996.

CFTCS. **Prisma**: à luz da física. avaiable in: http://cftc.cii.fc.ul.pt/PRISMA/ , visited in: 12/2013, 2012.

CHIARAMONTE, M. S. et al. Em busca de um padrão cognitivo na Engenharia. In: ICECE 2007 INTERNATIONAL CONFERENCE ON ENGINEERING AND COMPUTER EDUCATION, 2007. **Proceedings...** [S.l.: s.n.], 2007. v.1, p.942–946.

CHIARAMONTE, M. S. et al. Em busca de um padrão cognitivo na Engenharia. In: SIMPOSIO IBEROAMERICANO DE EDUCACIóN, CIBERNéTICA E INFORMáTICA (SIECI 2007) AND 6TH CONFERENCIA IBEROAMERICANA EN SISTEMAS, CIBERNéTICA E INFORMáTICA (CISCI 2007), 4., 2007. **...** [S.l.: s.n.], 2007. p.170–175.

CHUNG, S.; KENNEDY, R. Forward-backward non-linear filtering technique for extracting small biological signals from noise. **Journal of Neuroscience Methods**, [S.l.], v.40, n.1, p.71 – 86, 1991.

CRILLY, T. **50 Mathematical Ideas You Really Need to Know**. [S.l.]: Quercus, 2008.

CUNHA, J. A. **Psicodiagnóstico-V**. [S.l.]: Artmed, 2003.

CUSTODIO, R. F. **Análise não-linear no reconhecimento de padrões sonoros** : estudo de caso para sons pulmonares. 1999. Thesis (PhD on Computer Science) — PPGC-UFRGS.

CYBENKO, G. Approximations by superpositions of sigmoidal functions. **Mathematics of Control, Signals, and Systems**, [S.l.], v.2(4), p.303 to 314, 1989.

DEHAENE, S.; BRANNON, E. (Ed.). **Space, Time and Number in the Brain. Searching for the Foundations of Mathematical Thought**. [S.l.]: Academic Press, 2011.

DIAS-TOSTA, E. et al. Decrease of non-linear structure in the EEG of Alzheimer patients compared to healthy controls. **Clinical Neurophysiology**, [S.l.], v.110, p.1159–1167, 1999.

FAURE, P.; KORN, H. Is there chaos in the brain? I. Concepts of nonlinear dynamics and methods of investigation. **Comptes Rendus de lÁcadémie des Sciences - Series III - Sciences de la Vie**, [S.l.], v.324, n.9, p.773 – 793, 2001.

FRASER, A. M.; SWINNEY, H. L. Independent coordinates for strange attractors from mutual information. **Phys. Rev. A**, [S.l.], v.33, p.1134–1140, Feb 1986.

FREUND, Y.; SCHAPIRE, R. E. Large margin classification using the perceptron algorithm. In: ANNU. CONF. ON COMPUT. LEARNING THEORY, 11., 1998. **Proceedings...** [S.l.: s.n.], 1998.

FREUND, Y.; SCHAPIRE, R. E. Large Margin Classification Using the Perceptron Algorithm. **Machine Learning**, [S.l.], v.37, p.277–296, 1999.

GARDNER, H. **Frames Of Mind**: the theory of multiple intelligences. [S.l.]: Basic Books, 1983.

GARDNER, H. **Intelligence Reframed**: multiple intelligences for the 21st century. [S.l.]: Basic Books, 1999.

GARDNER, H. **Mentes que mudam**: a arte e a ciência demudar as nossas idéias e as dos outros. [S.l.]: Bookman, 2005.

GAZZANIGA, M.; IVRY, R.; MANGUN, G. R. **Neurociência Cognitiva**: a biologia da mente. [S.l.]: Artmed, 2006.

GEVINS, A.; SMITH, M. E. Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. **Cerebral Cortex**, [S.l.], v.10, p.829–839, 2000.

GLASS, A.; RIDING, R. J. EEG differences and cognitive style. **Biological Psychology**, [S.l.], v.51, p.23–41, 1999.

GLEICK, J. **Caos**: a criação de uma nova ciência. [S.l.]: Campus, 1989.

HARTIGAN, J. **Clustering Algorithms**. [S.l.]: Books on Demand, 1975. (Wiley Series in Probability and Mathematical Statistics).

HAYKIN, S. **Neural Networks**: a comprehensive foundation. [S.l.]: Prentice Hall, 1998.

HEGGER, R.; KANTZ, H.; SCHREIBER, T. **TISEAN**: nonlinear time series analysis. avaiable in: http://www.mpipks-dresden.mpg.de/ tisean/ , visited in: 12/2013, 2007.

HELMBOLD, D. P.; WARMUTH, M. K. On weak learning. **Journal of Computer and System Science**, [S.l.], v.50, p.551–573, 1995.

HUDSON, C. G. **AT THE EDGE OF CHAOS**: a new paradigm for social work? avaiable in: http://website.cswe.org/publications/members-only/00-2Chaos.htm , visited in: 12/2013, 2000.

INSTRUMENTS, N. **LabView**. avaiable in: http://www.ni.com/labview/whatis/ , visited in: 12/2013, 2013.

JASPER, J. J. The 10/20 international electrode system. **EEG Clinical Neurophysiology**, [S.l.], v.10, p.371–375, 1958.

JIANG, J. J.; ZHANG, Y.; MCGILLIGAN, C. Chaos in Voice, From Modeling to Measurement. **Journal of Voice**, [S.l.], v.20, n.1, p.2 – 17, 2006.

JOSEPH, P.; KANNATHAL, N.; ACHARYA, U. R. Complex Encephalogram Dynamics during Meditation. **Journal of Chinese clinical medicine**, [S.l.], p.220–230, 2007.

JR., W. H. H.; BUCK, J. A. **Eletromagnetismo**. [S.l.]: McGraw Hill Brasil, 2003.

KANDEL, E. R.; SCHWARTZ, J. H. **Principles of Neural Sciences**. [S.l.]: Elsevier, 1985.

KANTZ, H. A robust method to estimate the maximal Lyapunov exponent of a time series. **Physics Letters A**, [S.l.], v.185, p.77–87, 1994.

KENNEL, M. B.; BROWN, R.; ABARBANEL, H. D. I. Determining embedding dimension for phase-space reconstruction using a geometrical construction. **Phys. Rev. A**, [S.l.], v.45, p.3403–3411, Mar 1992.

KORN, H. Is there chaos in the brain? II. Experimental evidence and related models. **C. R. Biol**, [S.l.], v.326, p.787–840, 2003.

LORENZ, E. N. Deterministic Nonperiodic Flow. **J. Atmos. Sci.**, [S.l.], v.20, p.130–141, 1963.

LUCK, S. J. **An Introduction to the Event-Related Potential Technique (Cognitive Neuroscience)**. [S.l.]: A Bradford Book, 2005.

LY, Y.; O'BOYLE, M. W. How sex, native language, and college major relate to the cognitive strategies used during 3-d mental rotation. **The Psychological Record**, [S.l.], v.58, p.287–300, 2008.

MARDIA, K. V.; KENT, J.; BIBBY, J. **Multivariate analysis**. [S.l.]: Academic Press, 2008.

MARON, G. **WEKA - using tansig in a MultiLayer Percetron**. avaiable in: http://bit.ly/MaronGui , visited in: 12/2013, 2013.

MARON, G.; BARONE, D. A. C.; RAMOS, E. A. Measuring the differences between Spatial Intelligence in different individuals using Lyapunov Exponents. In: MDA 2012), 2012. **Proceedings...** [S.l.: s.n.], 2012.

MARON, G.; BARONE, D. A. C.; RAMOS, E. A. Spatial cognition degree of development classification Using multi-layer perceptron and largest Lyapunov exponents. In: PROCEEDINGS OF THE IEEE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN, AND CYBERNETICS (IEEE SMS 2013), 2013. **...** [S.l.: s.n.], 2013.

MARON, G.; BARONE, D. A. C.; RAMOS, E. A. Spatial cognition degree of development classification Using artificial neural networks and largest Lyapunov exponents. In: BRICS COUNTRIES CONGRESS (BRICS-CCI) AND 11TH BRAZILIAN CONGRESS (CBIC) ON COMPUTATIONAL INTELLIGENCE (BRICS-CCI 2013), 1., 2013. **Proceedings...** [S.l.: s.n.], 2013.

MARTINS, L. C. avaiable in: http://www.mundofisico.joinville.udesc.br/index.php? idSecao=9&idSubSecao=&idTexto=187, visited in: 12/2013, 2012.

MATHWORKS. **MATLAB - The Language Of Technical Computing**. avaiable in: http://www.mathworks.com/products/matlab/ , visited in: 12/2013, 2013.

MOREIRA, M. A. **Mapas Conceituais e Diagramas V**. avaiable in: http://www.if.ufrgs.br/ moreira/Livro_Mapas_conceituais_e_Diagramas_V_COMPLETO.pdf , visited in: 12/2013, 2006.

NOVAK, J. D.; nAS, A. J. C. **The Theory Underlying Concept Maps and How to Construct and Use Them**. avaiable in: http://cmap.ihmc.us/publications/researchpapers/theorycmaps/theoryunderlyingconceptmaps.htm , visited in: 12/2013, 2008.

PALUS, M. **Nonlinear Dynamics in the EEG Analysis**: disappointments and perspectives. 1998.

PATIL, S. T.; BORMANE, D. S. Electroencephalograph Signal Analysis During Bramari. In: INTERNATIONAL CONFERENCE ON INFORMATION TECHNOLOGY (ICIT 2006), 9., 2006. **...** [S.l.: s.n.], 2006. v.1, p.26–32.

PELLEG, D.; MOORE, A. X-means: extending k-means with efficient estimation of the number of clusters. In: SEVENTEENTH INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 2000, San Francisco. **Proceedings...** Morgan Kaufmann, 2000. p.727–734.

PETRY, A. **Reconhecimento Automático de Locutor utilizando medidas de Invariantes Dinâmicas Não-Lineares**. 2002. Thesis (PhD on Computer Science) — PPGC UFRGS.

PETRY, A.; BARONE, D. A. C. Fractal dimension applied to speaker identification. In: ICASSP 01), 2001. **...** [S.l.: s.n.], 2001. v.1, p.405–408.

PETRY, A.; BARONE, D. A. C. Text-dependent speaker verification using lyapunov exponents. In: INTERSPEECH, 2002. **...** ISCA, 2002.

PETRY, A.; BARONE, D. A. C. Speaker identification using nonlinear dynamical features. **Chaos, Solitons & Fractals**, [S.l.], v.13, n.2, p.221 – 231, 2002.

PETRY, A.; BARONE, D. A. C. Preliminary experiments in speaker verification using time-dependent largest Lyapunov exponents. **Computer Speech & Language**, [S.l.], v.17, n.4, p.403 – 413, 2003.

PEZARD, L.; NANDRINO, J. Dynamic paradigm in psychopathology: "chaos theory", from physics to psychiatry. **Encephale**, [S.l.], v.27, n.3, p.260–8, 2001.

PINKER, S. **Como a Mente Funciona**. [S.l.]: Companhia das Letras, 1999.

PINKER, S. **Tábula Rasa**. [S.l.]: Companhia das Letras, 2000.

PRIMI, R.; ALMEIDA, L. S. Estudo de validaçã da bateria de provas de raciocínio (BPR-5). **Psicologia**: teoria e pesquisa, [S.l.], v.16, p.165–173, 08 2000.

ROBERT, C.; GAUDY, J.-F.; LIMOGE, A. Electroencephalogram processing using neural networks. **Clinical Neurophysiology**, [S.l.], v.113, n.5, p.694 – 701, 2002.

ROBERTS, J. E.; BELL, M. A. Sex differences on a mental rotation task: variations in electroencephalogram hemispheric activation between children and college students. **Dev Neuropsychol.**, [S.l.], v.17, p.199–223, 2000.

ROBERTS, J. E.; BELL, M. A. Two- and three-dimensional mental rotation tasks lead to different parietal laterality for men and women. **International Journal of Psychophysiology**, [S.l.], v.50, p.235–246, 2003.

ROSENBLAT, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological Review**, [S.l.], v.65, p.386–407, 1958.

ROSENBLATT, F. **Principles of Neurodynamics**. [S.l.]: Spartan, 1962.

SHEPARD, R. N.; METZLER, J. Mental Rotation of Three-Dimensional Objects. **Science**, [S.l.], v.171, p.701–703, 1971.

SOFTWARE, I. **SPSS Statistics**. avaiable in: http://www-01.ibm.com/software/analytics/spss/products/statistics/ , visited in: 12/2013, 2013.

SPINDOLA, M. et al. Cognitive Measure on Different Profiles. In: BRAIN INSPIRED COGNITIVE SYSTEMS 2008, 2010. **...** Springer Science, 2010. p.365–378. (Advances in Experimental Medicine and Biology, v.657).

SPINDOLA, M. M. **Habilidade cognitiva espacial**: medida com eletroencefalografia. 2010. Thesis (PhD on Computer Science) — PPGIE - UFRGS.

SPíNDOLA, M. M. et al. A study of brain reaction to spatial stimuli in students with different background knowledge. **Ciências e Cogniçâo / Science and Cognition**, [S.l.], v.15, n.3, 2010.

SUBHA, D. et al. EEG Signal Analysis: a survey. **Journal of Medical Systems**, [S.l.], v.34, p.195–212, 2010. 10.1007/s10916-008-9231-z.

SWIDERSKI, B.; OSOWSKI, S.; RYSZ, A. Lyapunov Exponent of EEG Signal for Epileptic Seizure Characterization. In: EUROPEAN CONFERENCE ON CIRCUIT THEORY AND DESIGN, 2005., 2005. **Proceedings...** [S.l.: s.n.], 2005.

TAKENS, F. Detecting Strange Attractors in Turbulence. **Dynamical Systems and Turbulence - Lecture Notes in Mathematics**, [S.l.], v.898, p.366–381, 1981.

THEILER, J. et al. Testing for nonlinearity in time series: the method of surrogate data. **Physica D**, [S.l.], v.58, p.77–94, 1992.

THIELO, M. R. **Análise e classificação de séries temporais não estacionárias utilizando métodos não-lineares**. 2000. Dissertation (Masters on Computer Science) — PPGC-UFRGS.

THIELO, M. R.; BARONE, D. A. C. Treatment of Local Minima in a Chaotic Data Clustering Task with an Extremely simple Genetic Algorithm. In: EUROPEAN INTERDISCIPLINARY SCHOOL ON NONLINEAR DYNAMICS FOR SYSTEM AND SIGNAL ANALYSIS (EUROATTRACTOR 2000), 1., 2000. **Proceedings...** [S.l.: s.n.], 2000.

THORNDIKE, R. L. Who Belong in the Family? **Psychometrika**, [S.l.], v.18, 1953.

UESTUEN, B.; MELSSEN, W.; BUYDENS, L. Facilitating the application of Support Vector Regression by using a universal Pearson VII function based kernel. **Chemometrics and Intelligent Laboratory Systems.**, [S.l.], v.81, p.29–40, 2006.

VOGL, T. P. et al. Accelerating the convergence of the backpropagation method. **Biological Cybernetics**, [S.l.], v.59, p.257–263, 1988.

WAIKATO, M. L. G. at University of. **Weka 3**: data mining software in java. avaiable in: http://www.cs.waikato.ac.nz/ml/weka/ , visited in: 12/2013, 2013.

WOLF, A. et al. Determining Lyapunov exponents from a time series. **Physica. D**, [S.l.], v.16, p.285–317, 1985.

WOLFRAMRESEARCH. **MATHEMATICA**: neural networks. avaiable in: http://media.wolfram.com/documents/NeuralNetworksDocumentation.pdf , visited in: 12/2013, 2005.

XU, G. et al. Event-related potential studies of attention to shape under different stimuli tasks. In: CONF PROC IEEE ENG MED BIOL SOC., 2006. **...** [S.l.: s.n.], 2006.

XU, R.; WUNSCH D., I. Survey of clustering algorithms. **Neural Networks, IEEE Transactions on**, [S.l.], v.16, n.3, p.645–678, 2005.

APPENDIX A HISTOGRAMS AND NORMAL CURVES

The next figures will show the histograms, and the normal curves which approximate the LLE distributions for **SI+** and **SI-** groups, repectively, in each analyzed electrode: **FP1** (figure A.1), **FP2** (figure A.2), **F3** (figure A.3), **F4** (figurae A.4), **T3** (figure A.5), **T4** (figure A.6), **P3** (figure A.7), and **P4** (figure A.8).

FP1 electrode, where the difference between the LLE means of both groups was about 20%, it can be seen in 4.1, in chapter 4.

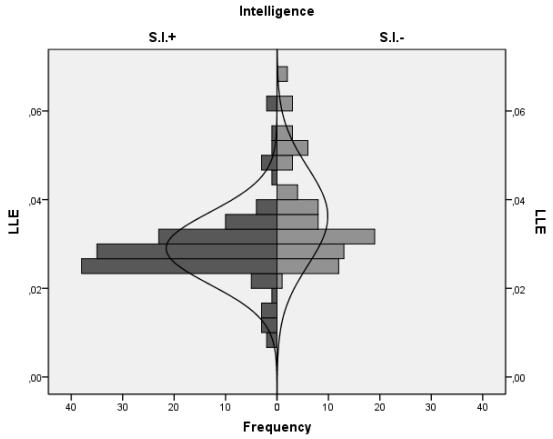

Figure A.1: Histogram and Normal curve which approximates the LLEs distribution for **S.I.+** and **S.I.-** groups, respectively, on **FP1** electrode.

FP2 electrode, where the difference between the LLE means of both groups was about 25%, it can be seen in 4.1, in chapter 4. It was the electrode where the largest difference in the LLE means of both groups was found.

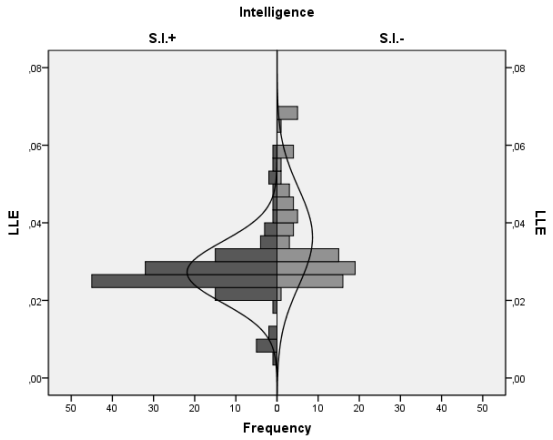

Figure A.2: Histogram and Normal curve which approximates the LLEs distribution for **S.I.+** and **S.I.-** groups, respectively, on **FP2** electrode.

F3 electrode, where the difference between the LLE means of both groups was about 13%, it can be seen in 4.1, in chapter 4.

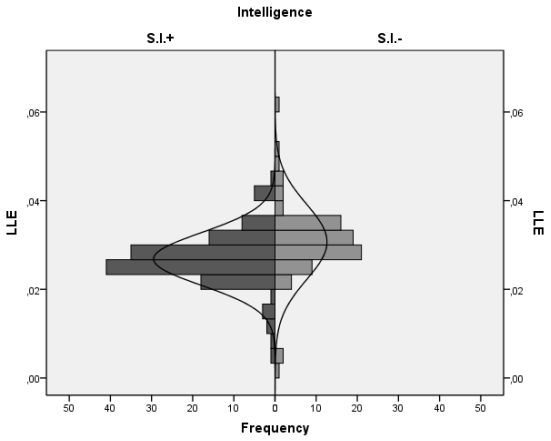

Figure A.3: Histogram and Normal curve which approximates the LLEs distribution for **S.I.+** and **S.I.-** groups, respectively, on **F3** electrode.

F4 electrode, where the difference between the LLE means of both groups was about 22%, it can be seen in 4.1, in chapter 4. The difference in the LLE means for both groups in this electrode was very close from the largest one.

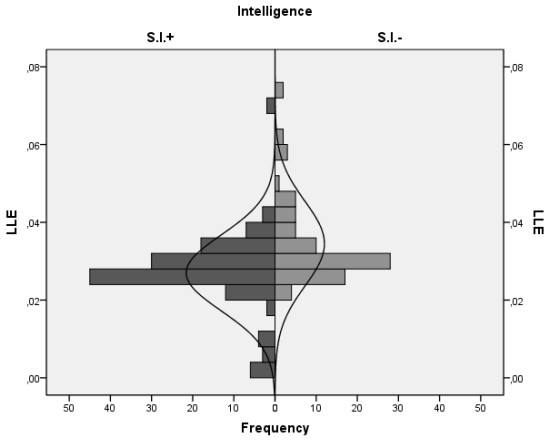

Figure A.4: Histogram and Normal curve which approximates the LLEs distribution for **S.I.+** and **S.I.-** groups, respectively, on **F4** electrode.

T3 electrode, where the difference between the LLE means of both groups was about 5%, it can be seen in 4.1, in chapter 4. It was the electrode where the smaller difference in the LLE means of both groups was found.

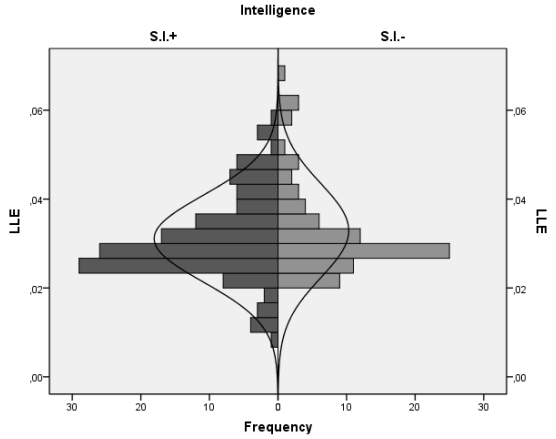

Figure A.5: Histogram and Normal curve which approximates the LLEs distribution for **S.I.+** and **S.I.-** groups, respectively, on **T3** electrode.

T4 electrode, where the difference between the LLE means of both groups was about 17%, it can be seen in 4.1, in chapter 4. This was the electrode where the larger absolute LLE means were found.

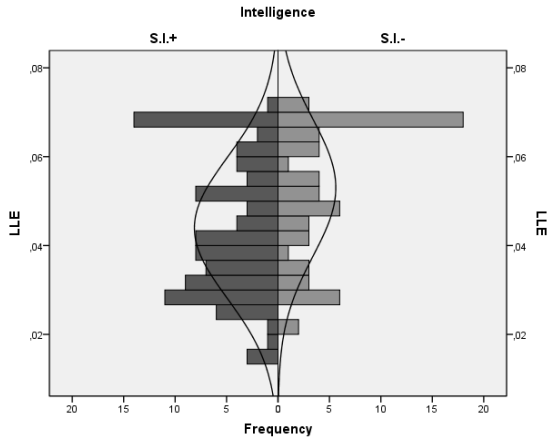

Figure A.6: Histogram and Normal curve which approximates the LLEs distribution for **S.I.+** and **S.I.-** groups, respectively, on **T4** electrode.

P3 electrode, where the difference between the LLE means of both groups was about 12%, it can be seen in 4.1, in chapter 4.

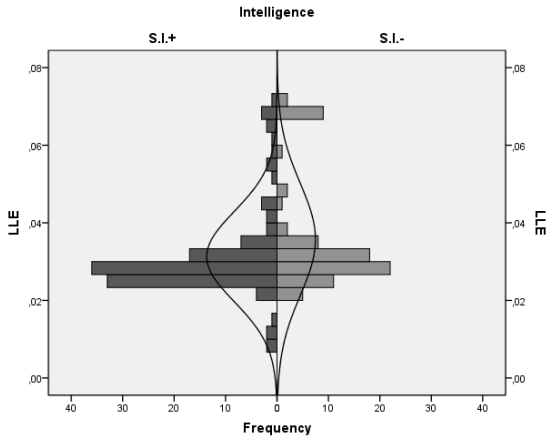

Figure A.7: Histogram and Normal curve which approximates the LLEs distribution for **S.I.+** and **S.I.-** groups, respectively, on **P3** electrode.

P4 electrode, where the difference between the LLE means of both groups was about 6%, it can be seen in 4.1, in chapter 4. The difference the LLE means for both groups in this electrode was very close from the smallest one.

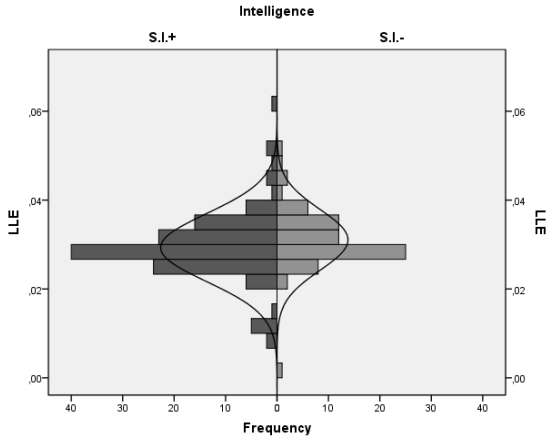

Figure A.8: Histogram and Normal curve which approximates the LLEs distribution for **S.I.+** and **S.I.-** groups, respectively, on **P4** electrode.

APPENDIX B RESUMO EM PORTUGUÊS

B.1 Introdução

O Mapa Conceitual apresentado na figura B.1 sumariza os conceitos, objetivos e hipóteses que guiaram a presente Dissertação.

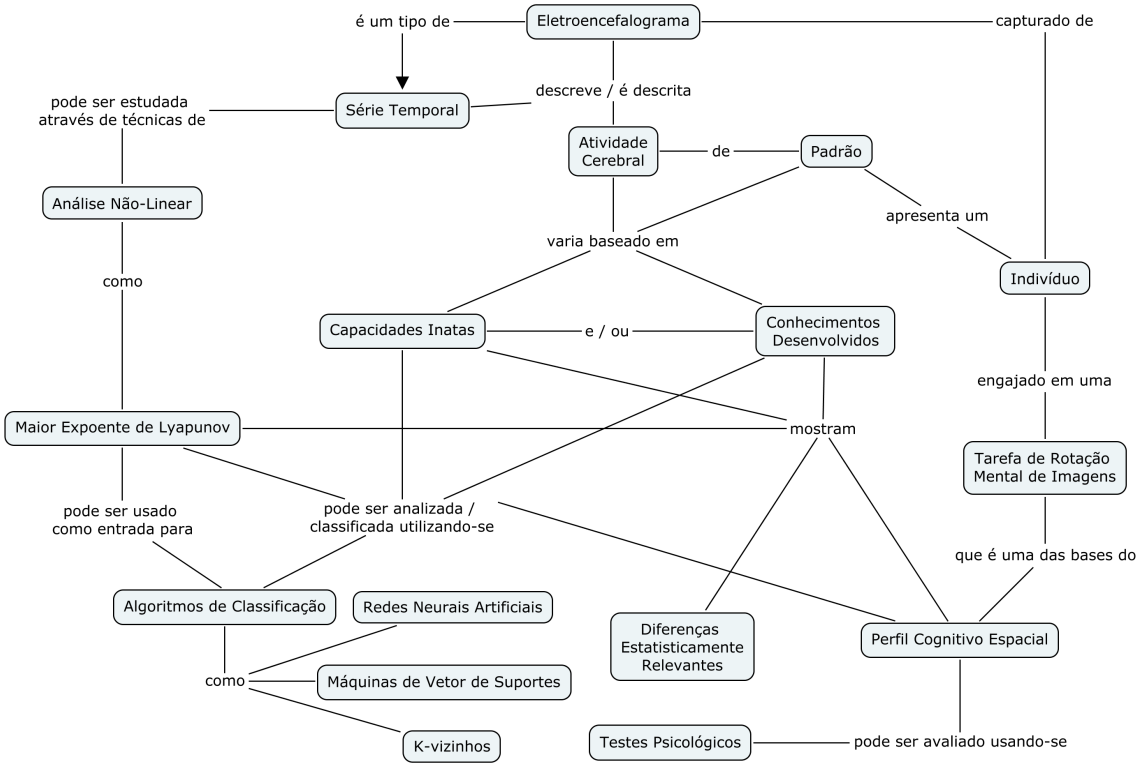

Figure B.1: Mapa Conceitual que sumariza os conceitos, objetivos e hipóteses que guiaram a presente Dissertação.

Os eletroencefalogramas (EEGs) capturados a partir dos escalpos dos participantes do experimento são um tipo de série temporal, que descreve a atividade cerebral de um indivíduo no momento da captura. Os EEGs desse experimento foram capturados durante a realização de tarefas de rotação mental de imagens tridimensionais, como a tarefa descrita em (SHEPARD; METZLER, 1971) - que é uma das bases da Capacidade Cognitiva Espacial - que é o aspecto da inteligência responsável pela imaginação e manipulação mental de geometrias e do espaço em si. A Capacidade Cognitiva Espacial de um indivíduo pode ser avaliada usando-se testes psicológicos, como, por exemplo, o BPR-5 RE (PRIMI; ALMEIDA, 2000). Sabe-se que há diferenças observáveis nos sistemas cerebrais que dão suporte a memória de trabalho e ao controle de atenção. Essas diferenças resultam em diferenças gerais na capacidade cognitiva de diferentes indivíduos (GEVINS; SMITH, 2000). Os padrões de atividade cerebral de um indivíduo variam, baseados em seus conhecimentos desenvolvidos e/ou capacidades inatas, e essas diferenças são estatisticamente relevantes (CHIARAMONTE et al., 2007a,b; SPINDOLA et al., 2010; SPíNDOLA et al., 2010; SPINDOLA, 2010; MARON; BARONE; RAMOS, 2012). Como dito anteriormente, o EEG é um tipo de série temporal, mais precisamente uma série temporal biológica. Sinais complexos, assim como o EEG, são mais bem analisados utilizando-se técnicas de análise não-linear (CHUNG; KENNEDY, 1991; ABARBANEL et al., 1993; CERUTTI et al., 1996; SUBHA et al., 2010). A técnica de análise não-linear que será utilizada no presente trabalho é o cálculo do Maior Expoente de Lyapunov (LLE), que será

calculado a partir de cada EEG. Os LLEs serão utilizados como tuplas de entrada para diferentes algoritmos de classificação: i) Redes Neurais Artificiais, ii) Máquinas de Vetor de Suporte, e iii) K-vizinhos; visando classificar o grau de desenvolvimento cognitivo espacial de cada indivíduo.

B.2 Principais Contribuições

- Realização da análise estatística das diferenças de atividade cerebral entre indivíduos dos grupos de interesse do experimento durante a realização de tarefas de rotação mental de imagens tridimensionais, um resumo dessa análise pode ser vista na figura B.2.
- Proposta, desenvolvimento e teste de um método para a classificação do grau de desenvolvimento da capacidade cognitiva espacial, cujo esquema pode ser visto na figura B.3.
- Desenvolvimento de um novo classificador para a ferramenta open-source WEKA (WAIKATO, 2013), uma MLP que utiliza a tansig como função de transferência. Essa implementação pode ser encontrada em (MARON, 2013).

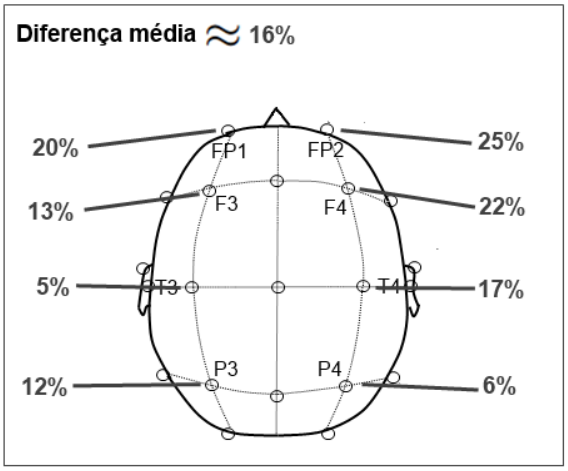

Figure B.2: As diferenças aproximadas entre as médias dos LLEs em cada eletrodo entre os grupos **SI+** e **SI-**.

B.3 Conclusões

O presente trabalho propôs, desenvolveu e testou um método para a classificação do grau de desenvolvimento do perfil cognitivo espacial de diferentes indivíduos. Também foi realizada uma análise estatística das diferenças de atividade cerebral entre os integrantes dos grupos de interesse durante a realização de uma tarefa de rotação mental de imagens tridimensionais.

As bases teóricas do presente trabalho podem ser encontradas no capítulo 2, onde uma revisão sobre as técnicas de análise não-linear utilizadas, algoritmos de classificação e sobre capacidade cognitiva espacial é feita.

No capítulo 3 o método proposto, desenvolvido e testado, cujo esquema pode ser visto na figura B.3, é descrito em detalhes:

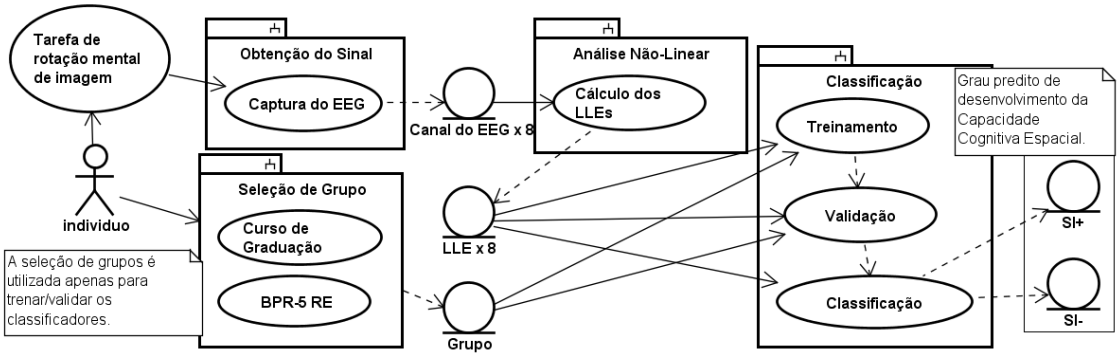

Figure B.3: Esquema do método proposto.

- Os participantes do experimento.
- Os grupos de interesse nos quais eles foram classificados (**SI+** e **SI-**).
- O teste psicológico BPR-5 RE aplicado aos participantes para fortalecer a hipótese de classificação.
- A tarefa de rotação mental de imagens tridimensionais realizadas pelos participantes.
- O equipamento utilizado para a captura dos EEGs.
- As técnicas de análise não-linear aplicadas sobre os EEGs.
- Os conjuntos de dados gerados.
- O processo de classificação: a escolha dos classificadores e dos parâmetros.

O capítulo 4 mostra os resultados obtidos, na análise estatística das diferenças de atividade cerebral entre os dois grupos durante a realização de uma tarefa de rotação mental de imagens tridimensionais. Esse capítulo também mostra os resultados obtidos para a classificação do grau de desenvolvimento da capacidade cognitiva espacial dos participantes, e os parâmetros utilizados para obtê-los usando cada classificador.

Os resultados encontrados na análise estatística das médias dos LLEs confirmam as descobertas realizadas em (ROBERTS; BELL, 2000, 2003; GAZZANIGA; IVRY; MANGUN, 2006), de que o hemisfério direito do cérebro é mais ativado durante tarefas de rotação mental de imagens tridimensionais. A diferença relevante no esforço cerebral entre os participantes é independente de gênero. Pode-se ver na imagem 4.1 (seção 4.1 do capítulo 4) que as diferenças nas médias dos LLEs entre participantes do gênero masculino e do gênero feminino são pequenas (nem mesmo estatisticamente relevantes). A diferença ocorre entre os grupos de interesse: **SI+** e **SI-**, e é qualitativamente a mesma considerando-se os indivíduos do gênero masculino no grupo **SI+** (**M+**) contra os indivíduos do gênero masculino no grupo **SI-** (**M-**) ou os indivíduos do gênero feminino em ambos os grupos: **F+** e **F-**, respectivamente. Uma investigação mais profunda sobre esse aspecto está fora do escopo do presente trabalho.

As diferenças de atividade cerebral são semelhantes tanto no sistema como um todo como localmente, em cada eletrodo. Uma média de aproximadamente 16% de diferença

foi encontrada entre os dois grupos. As maiores diferenças foram encontradas nos eletrodos FP1 e FP2, mas as maiores médias nos LLEs foram encontradas no eletrodo P4.

O maior desafio na tarefa de classificar o grau de desenvolvimento da capacidade cognitiva espacial de um indivíduo é o grande número de falso-positivos. Especificidades acima de 50% raramente foram obtidas em configurações que não as melhores em cada classificador. Enquanto Redes Neurais Artificiais e K-vizinhos foram capazes de lidar com esse desafio as Máquinas de Vetor de Suporte tiveram seus resultados piorados pelo grande número de falso-positivos, alcançando apenas 41,87% de especificidade média, enquanto os outros algoritmos passaram da marca dos 50%.

A hipótese de classificar todos os indivíduos no grupo SI+, o maior deles, obteria 61,68% de acurácia, 0% de especificidade e uma medida F1 de 0,7630. O Perceptron de múltiplas camadas (classificador base) e a Rede Neural Artificial de Funções de base Radial (o algoritmo com o melhor desempenho) apresentaram resultados melhores nos 4 critérios analisados: acurácia, especificidade, área sob a curva ROC e medida F1, do que os outros classificadores analisados. A hipótese de utilizar-se a MLP apresentou-se como uma boa escolha, entretanto o uso da RBF mostrou-se ainda melhor.

Para alcançar a melhor acurácia uma MLP com 1 camada oculta de 18 neurônios foi utilizada. Entretanto a melhor especificidade foi alcançada utilizando-se uma MLP com 3 camadas ocultas, cada uma com 18 neurônios. O tempo utilizado para treinar a MLP de 3 camadas não compensa o pequeno ganho de especificidade, ainda mais que isso acarreta em uma perda de acurácia. Mas, em média, a melhor acurácia foi alcançada utilizando-se uma RBF com 4 clusters. Um teste-t obteve sucesso em revelar uma diferença estatisticamente relevante entre a precisão dos 2 classificadores que apresentaram os melhores desempenhos: MLP - tansig (M = 70.20%, s = 6.32%) e RBF (M = 72.13%, s = 9.67%), t(430) = 2,446, p = 0.000, $\alpha = 0.05$

Outra contribuição do presente trabalho foi o desenvolvimento de um novo classificados para a ferramenta open-source WEKA (WAIKATO, 2013). No WEKA havia apenas a possibilidade de utilizar-se a Função Logística (logsig) como função de transferência na MLP. No escopo do presente trabalho uma MLP utilizando a Tangente Hiperbólica (tansig) foi implementada para o WEKA, e pode ser encontrada em (MARON, 2013).