

Universidade Federal de Campina Grande  
Centro de Engenharia Elétrica e Informática  
Coordenação de Pós-Graduação em Ciência da Computação

Análise de similaridade entre  
classes de padrões de ativação neuronal

Eugênio de Carvalho Saraiva

Dissertação submetida à Coordenação do Curso de Pós-Graduação em  
Ciência da Computação da Universidade Federal de Campina Grande –  
Campus I como parte dos requisitos necessários para obtenção do grau  
de Mestre em Ciência da Computação.

Área de Concentração: Ciência da Computação  
Linha de Pesquisa: Neurociência Computacional

Herman Martins Gomes  
(Orientador)

Sidarta Tollendal Gomes Ribeiro  
(Orientador)

Campina Grande, Paraíba, Brasil

© Eugênio de Carvalho Saraiva, 30/07/2014

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

S243a Saraiva, Eugênio de Carvalho.  
Análise de similaridade entre classes de padrões de ativação neuronal /  
Eugênio de Carvalho Saraiva. – Campina Grande, 2014.  
147 f. : il. color.

Dissertação (Mestrado em Ciência da Computação) – Universidade  
Federal de Campina Grande, Centro de Engenharia Elétrica e Informática,  
2014.

"Orientação: Prof. Dr. Herman Martins Gomes, Prof. Dr. Sidarta  
Tollendal Gomes Ribeiro".

Referências.

1. Abordagem Não Supervisionada. 2. Abordagem Semi-  
Supervisionada. 3. Padrões de Ativação Neuronal. I. Gomes, Herman  
Martins Gomes. II. Ribeiro, Sidarta Tollendal Gomes. III. Título.

CDU 004(043)

## Resumo

Há um número crescente de tecnologias que fazem uso de algoritmos de classificação para a automação de tarefas. Em particular, em Neurociências, algoritmos de classificação foram usados para testar hipóteses sobre o funcionamento do sistema nervoso central. No entanto, a relação entre as classes de padrões de ativação neuronal de áreas específicas do cérebro, como resultado de experiências sensoriais tem recebido pouca atenção. No contexto da Neurociência Computacional, este trabalho apresenta uma análise do nível de similaridade entre classes de padrões de ativação neuronal, com o uso das abordagens de aprendizagem não supervisionada e semi-supervisionada, em áreas específicas do cérebro de ratos em contato com objetos, obtidos durante um experimento envolvendo exploração livre de objetos pelos animais. As classes foram definidas de acordo com determinados tratamentos construídos com níveis específicos de um conjunto de 8 fatores (Animal, Região do Cérebro, Objeto ou Par de Objeto, Algoritmo de Agrupamento, Métrica, *Bin*, Janela e Intervalo de Contato). No total foram analisados 327.680 tratamentos. Foram definidas hipóteses quanto à relação de cada um dos fatores para com o nível de similaridade existente entre os tratamentos. As hipóteses foram verificadas por meio de testes estatísticos entre as distribuições que representavam cada uma das classes. Foram realizados testes de normalidade (*Shapiro-Wilk*, *QQ-plot*), análise de variância e um teste para diferenças entre tendência central (*Kruskal-Wallis*). Com base nos resultados encontrados nos estudos utilizando abordagem não supervisionada, foi inferido que os processos de aquisição e de definição dos padrões de ativação por um observador foram sujeitos a uma quantidade não significativa de ruídos causados por motivos não controláveis. Pela abordagem semi-supervisionada, foi observado que nem todos os graus de similaridade entre pares de classes de objetos são iguais a um dado tratamento, o que indicou que a similaridade entre classes de padrões de ativação neuronal é sensível a todos os fatores analisados e fornece evidências da complexidade na codificação neuronal.

## Abstract

There are a growing number of technologies that make use of classification algorithms for automating tasks. In particular, in Neuroscience, classification algorithms were used to test hypotheses about the functioning of the central nervous system. However, the relationship between the classes of patterns of neuronal activation in specific brain areas as a result of sensorial experience has received little attention. In the context of Computational Neuroscience, this paper presents an analysis of the level of similarity between classes of patterns of neuronal activation with the use of learning approaches unsupervised and semi-supervised in specific areas of rat brain in contact with objects, obtained during an experiment involving free exploration of objects by animals. The classes were defined according to certain treatments constructed with specific levels with set of 8 factors (Animal, Brain Region, Object or Pair of Objects, Clustering Algorithm, Metric, Bin, Window and Interval Contact). In total 327.680 treatments were analyzed. Hypotheses regarding the relationship of each of the factors with the existing level of similarity between treatments were defined. The hypotheses were tested through between statistical distributions representing each class tests. The tests applied where the tests for normality (*Shapiro-Wilk*, *QQ-plot*), analysis of variance and a test for differences in central tendency (*Kruskal-Wallis*) were performed. Based on the results found in studies using an unsupervised approach, it was inferred that the process of acquisition and definition of patterns of activation by an observer was not subject to a significant amount of noise caused by uncontrollable reasons. For the semi-supervised approach, it was observed that not all degrees of similarity between pairs of classes of objects are equal to a given treatment, which indicated that the similarity between classes of patterns of neuronal activation is sensitive to all the factors analyzed and provides evidence about the complexity of neuronal coding.

## Agradecimentos

Agradeço à Santíssima Trindade por tudo. À Deus, por ter me dado a vida e esse mundo. À Jesus, pela redenção. E ao Espírito Santo, por todos os Dons e entusiasmo.

À Maria, Nossa Senhora, por ter aceito sua missão, pelo amor e carinho, e, como também, a todos os Anjos e Santos, por sempre me apontarem a Santidade.

Aos meus pais, Francisco Márcio Eugênio Vieira Saraiva e Solange Braga de Carvalho Saraiva, que me criaram em um lar repleto de paz, amor, alegria, e por serem meus melhores amigos e exemplos de pessoas. Por sempre acreditarem nos meus sonhos e potenciais, e por ter me provado que a Família é a maior Bênção que existe.

Aos meus irmãos, Fernanda de Carvalho Saraiva e Márcio de Carvalho Saraiva, por todo o carinho, atenção, cumplicidade e por saber que nenhuma distância nos separará.

Aos meus demais familiares, avós, tios e primos, por todo apoio, carinho e afeto.

À minha namorada, Nathalia Costa Gonzaga, por sempre estar ao meu lado, em todos os momentos. Por todo carinho e preocupações. Por incentivar em mim o amor, a paz, a Fé, a alegria, o conhecimento, e a vontade de sempre buscar dar o meu melhor.

Aos meus colegas e amigos, que me acompanharam em diversas caminhadas. Pela amizade fraterna e companheirismo.

À Igreja Católica Apostólica Romana, que me dá a oportunidade de servir à Deus pelo próximo, nas pastorais, serviços e missões a mim confiados, e me ajuda a crescer na Fé.

À todos os meus professores e instituições de ensino, por todos os ensinamentos, acadêmicos e não acadêmicos, que me ajudam a me tornar um melhor cidadão.

Aos meus orientadores, professor Herman Martins Gomes, por toda atenção, respeito e amizade. Por muitas vezes ir além do “dever”, pelo simples fato de sempre querer o melhor para mim e para o nosso trabalho. E professor Sidarta Tollendal Gomes Ribeiro, pela colaboração, por ter cedido os dados que utilizamos nesse trabalho, e pela toda disposição empregada.

Aos professores Nivaldo Antonio Portela de Vasconcelos e Joseana Macêdo Fachine Régis de Araújo, pela contribuição em nosso trabalho, por participarem da nossa banca de defesa de dissertação e por sempre me receberem muito bem para ajudar na solução de qualquer problema.

# Conteúdo

1	Introdução .....	1
1.1	Contexto.....	1
1.2	Definição do Problema .....	2
1.3	Objetivos.....	2
1.4	Escopo.....	3
1.5	Resultados e Contribuições .....	3
1.6	Estrutura da Dissertação .....	4
2	Revisão de Literatura .....	5
2.1	Histórico.....	5
2.2	Análise de Padrões de Ativação Neuronal .....	7
2.3	Análise de Agrupamentos.....	15
2.4	Considerações sobre o Capítulo .....	22
3	Materiais e Métodos de Avaliação .....	24
3.1	Esquema da estratégia proposta .....	24
3.2	Ferramentas .....	25
3.2.1	Neurojects .....	25
3.2.2	MySQL .....	26
3.2.3	Weka .....	27
3.2.4	R.....	27
3.3	Dados e Experimento Comportamental .....	28
3.4	Abordagem Não Supervisionada.....	35
3.4.1	Questões .....	35
3.4.2	Fatores e Variável de Resposta .....	36
3.4.3	Níveis dos Fatores e Projeto do Experimento.....	37
3.4.4	Hipóteses.....	38
3.4.5	Testes de Hipóteses .....	39
3.5	Abordagem Semi-Supervisionada.....	41

3.5.1	Questões .....	41
3.5.2	Fatores e Variável de Resposta .....	42
3.5.3	Níveis dos Fatores e Projeto do Experimento .....	44
3.5.4	Hipóteses .....	44
3.5.5	Testes de Hipóteses .....	47
3.6	Teste de Robustez das Inferências .....	48
3.7	Considerações sobre o Capítulo .....	49
4	Apresentação e Análise dos Resultados .....	50
4.1	Qualidade dos Dados .....	50
4.2	Abordagem Não Supervisionada.....	52
4.3	Abordagem Semi-Supervisionada.....	52
4.4	Considerações sobre o Capítulo .....	71
5	Considerações Finais e Sugestões para Trabalhos Futuros .....	72
5.1	Sumário da pesquisa .....	72
5.2	Conclusões .....	73
5.3	Proposta para trabalhos futuros .....	74
	Referências Bibliográficas.....	76
	Apêndice A – Amostra de registros analisados na abordagem semi-supervisionada .....	85
	Apêndice B – Fundamentação Teórica .....	98
	B.1 Conceitos da Ciência da Computação .....	98
	B.1.1 Aprendizagem .....	98
	B.1.2 Análise de Agrupamentos de Dados .....	100
	B.1.3 Métricas.....	105
	B.2 Conceitos da Neurociência .....	108
	B.2.1 Regiões de Interesse no Cérebro .....	108
	B.2.2 Neurônios .....	116
	B.2.3 <i>Bin</i> .....	121
	B.2.4 Janela.....	122

B.2.5 Métodos de Aquisição de Dados Neurais .....	123
B.2.6 Matrizes de Microeletrodos .....	124
B.2.7 Padrão de Ativação Neuronal .....	125
B.3 Metodologia Científica .....	125
B.3.1 Tipos de Estudo .....	126
B.3.2 Fatores .....	126
B.3.6 Análise Estatística.....	128



# Lista de Siglas e Abreviações

ANOVA – Análise de variância

AUROC – Área embaixo da curva ROC

CA – Córtex Auditivo Primário

DBScan – *Density-based spatial clustering of applications with noise*

Dens – Agrupamento baseado em Densidade

DWT - Transformada Discreta de *Wavelet*

EEG – Eletroencefalograma

EM - *Expectation-maximization*

FD - Fractal Dimension

Hier – Agrupamento Hierárquico

HP – Hipocampo

i1 – Primeira parte do intervalo

i2 – Segunda parte do intervalo

i3 – Terceira parte do intervalo

ic – Intervalo Completo

LFP - Potenciais de Campo Locais

MAP - Máxima Probabilidade a Posteriori

Pós-EXP – Pós-exposição

Pré-EXP – Pré-exposição

REM - Sono no qual ocorre o movimento rápido dos olhos

RM - Ressonância Magnética

S1 – Córtex Somatossensorial Primário

SDE – *Data Explorer Signal*

SNEO - *Smoothed Nonlinear Energy Operator*

SWS - Sono de Ondas Lentas

TI - Córtex Temporal Inferior

UPGMA - Método não ponderado de Agrupamento de Pares

V1 - Córtex Visual Primário

WK - Vigília

# Lista de Figuras

Figura 1: Esquema da estratégia proposta .....	24
Figura 2: Ilustração das três fases do experimento (PRÉ-EXP, EXPOSIÇÃO, PÓS-EXP). Adaptada de (RIBEIRO <i>et al.</i> , 2007) .....	29
Figura 3: Ilustração do processo de registro da atividade neuronal. (A) Animal com microeletrodos implantados. (B) Visão geral do sistema: Animal no ambiente povoado com objetos, ligado ao programa de processamento dos sinais elétricos de cada neurônio monitorado. Adaptada de (RIBEIRO <i>et al.</i> , 2007).....	30
Figura 4: Exemplos de arquivos disponibilizados para análise. (A) Registros de intervalo de contato de animal com determinados objetos. (B) Registro de ativação de neurônio.....	31
Figura 5: Exemplo de arquivo contendo padrões de ativação neuronal.....	32
Figura 6: Tempo de contato de cada animal com cada objeto.....	33
Figura 7: Número de neurônios monitorados de cada animal em cada região.....	34
Figura 8: Esquema das atividades realizadas na 1ª etapa. ....	39
Figura 9: Esquema das atividades realizadas na 2ª etapa. ....	40
Figura 10: Esquema do processo de construção das distribuições de similaridade.....	48
Figura 11: <i>Boxplots</i> da distribuição $u_{\text{Ge4\_V1\_}[\text{Bola, Escova}]_{\text{EM\_Dunn\_250b\_10j\_ic}}$ com a remoção gradativa da quantidade neurônios utilizados para compor os padrões de ativação neuronal. ....	51
Figura 12: <i>Boxplots</i> da distribuição $u_{\text{Ge4\_V1\_}[\text{Bola, Escova}]_{\text{EM\_Dunn\_250b\_10j\_ic}}$ com a substituição gradativa da atividade de neurônios por um valor aleatório de uma distribuição de Poisson, utilizados para compor os padrões de ativação neuronal. ....	51
Figura 13: <i>Boxplots</i> das distribuições $u_{(\text{Ge4, Ge5, Ge9, Ge12})_{\text{HP\_}[\text{Bola, Escova}]_{\text{K-}}}$ $\text{Means\_Jaccard\_250b\_10j\_ic}$ .....	53
Figura 14: <i>Boxplots</i> das distribuições $u_{\text{Ge4\_}(\text{Assembleia, HP,S1,V1})_{\text{[Bola, Escova]}_{\text{K-}}}$ $\text{Means\_Dunn\_250b\_10j\_ic}$ .....	54
Figura 15: <i>Boxplots</i> das distribuições $u_{\text{Ge4\_HP\_}([\text{Bola, Escova}], [\text{Bola, Comida}], [\text{Bola,}}}$ $\text{Haste}], [\text{Escova, Comida}], [\text{Escova, Haste}], [\text{Comida, Haste}],)_{\text{K-}}}$ $\text{Means\_Jaccard\_250b\_10j\_ic}$ .....	54
Figura 16: <i>Boxplots</i> das distribuições $u_{\text{Ge4\_HP\_}[\text{Bola, Escova}]_{(\text{K-}}}$ $\text{Means,EM,Hier,Dens)}_{\text{Jaccard\_250b\_10j\_ic}}$ .....	55

Figura 17: <i>Boxplots</i> das distribuições u_Ge4_HP_[Bola, Escova]_K-Means_(Jaccard, Dunn, Silhouette) _250b_10j_ic.....	55
Figura 18: <i>Boxplots</i> das distribuições u_Ge4_HP_[Bola, Escova]_K-Means_Jaccard_(50-400)b_10j_ic.....	56
Figura 19: <i>Boxplots</i> das distribuições u_Ge4_HP_[Bola, Escova]_K-Means_Dunn_100b_(5-40)j_ic.....	56
Figura 20: <i>Boxplots</i> das distribuições u_Ge4_HP_[Bola, Escova]_K-Means_Jaccard_100b_10j_i(1,2,3) .....	57
Figura 21: Disposição espacial das medianas dos níveis de similaridade das classes de padrões de ativação neuronal, em relação aos Animais, às Regiões do Cérebro e ao Número de Neurônios monitorados para a composição dos padrões de ativação.....	58
Figura 22: <i>Boxplots</i> das distribuições u_MediaAnimal_(HP,S1,V1)_[Bola, Escova]_EM_Dunn_250b_10j_ic .....	66
Figura 23: <i>Boxplots</i> das distribuições u_MediaAnimal_(HP,S1,V1)_[Bola, Comida]_EM_Dunn_250b_10j_ic .....	67
Figura 24: <i>Boxplots</i> das distribuições u_MediaAnimal_(HP,S1,V1)_[Bola, Haste]_EM_Dunn_250b_10j_ic .....	67
Figura 25: <i>Boxplots</i> das distribuições u_MediaAnimal_(HP,S1,V1)_[Escova, Comida]_EM_Dunn_250b_10j_ic .....	68
Figura 26: <i>Boxplots</i> das distribuições u_MediaAnimal_(HP,S1,V1)_[Escova, Haste]_EM_Dunn_250b_10j_ic.....	68
Figura 27: <i>Boxplots</i> das distribuições u_MediaAnimal_(HP,S1,V1)_[Comida, Haste]_EM_Dunn_250b_10j_ic.....	69
Figura 28: Disposição espacial das médias do índice Dunn pelo valor AUROC das distribuições u_(Ge4,Ge5,Ge9,Ge12)_(HP,V1,S1)_(Bola,Escova,Haste)_EM_Dunn_250b_10j_ic .....	70

# Lista de Tabelas

Tabela 1 – Trabalhos relacionados – Análise de Padrões de Ativação Neuronal. ....	8
Tabela 2 – Trabalhos relacionados - Análise de Agrupamento.....	16
Tabela 3 - Diferença observada entre as distribuições com foco no fator Animal.....	59
Tabela 4 - Razão de similaridade entre as distribuições com foco no fator Animal. ....	59
Tabela 5 - Diferença observada entre as distribuições com foco no fator Região do Cérebro.	59
Tabela 6 - Razão de similaridade entre as distribuições com foco no fator Região do Cérebro. .....	60
Tabela 7 - Diferença observada entre as distribuições com foco no fator Par de Objetos. ....	60
Tabela 8 - Razão de similaridade entre as distribuições com foco no fator Par de Objetos....	61
Tabela 9 - Diferença observada entre as distribuições com foco no fator Algoritmo de Agrupamento. ....	61
Tabela 10 - Razão de similaridade entre as distribuições com foco no fator Algoritmo de Agrupamento. ....	62
Tabela 11 - Diferença observada entre as distribuições com foco no fator Métrica .....	62
Tabela 12 - Razão de similaridade entre as distribuições com foco no fator Métrica.....	62
Tabela 13 - Diferença observada entre as distribuições com foco no fator <i>Bins</i> .....	63
Tabela 14 - Razão de similaridade entre as distribuições com foco no fator <i>Bins</i> . ....	63
Tabela 15 - Diferença observada entre as distribuições com foco no fator Janela.....	64
Tabela 16 - Razão de similaridade entre as distribuições com foco no fator Janela. ....	64
Tabela 17 - Diferença observada entre as distribuições com foco no fator Intervalo de Contato. ....	65
Tabela 18 - Razão de similaridade entre as distribuições com foco no fator Intervalo de Contato. ....	65

# Capítulo 1

## Introdução

Neste capítulo, é apresentado o contexto e a motivação da pesquisa descrita nesta dissertação. Primeiro, é contextualizada a área da Neurociência Computacional e as pesquisas que estão sendo realizadas nessa área. Dado esse contexto, apresentado um dos problemas que a área possui. Apresentamos, também, uma visão geral das soluções existentes e em que elas satisfazem ou não esses requisitos. A investigação desenvolvida nesta dissertação é apresentada e analisada em linhas gerais.

Por fim, o capítulo é encerrado com uma descrição da organização dos demais capítulos que compõem esta dissertação.

### *1.1 Contexto*

O sistema nervoso é formado por um grande número de elementos que interagem entre si. Para entender esse complexo funcionamento, é necessária a construção e análise de modelos computacionais (STERRATT *et al.*, 2011). A neurociência é uma disciplina que visa analisar o sistema nervoso para entender a base biológica do comportamento (SQUIRE, 2012). Em particular, em Ciência da Computação, estudos estão sendo conduzidos com representações da atividade eletrofisiológica de indivíduos, definidos como padrões de ativação neuronal, a fim de extrair recursos e realizar simulações por algoritmos de classificação, para entender melhor a dinâmica do cérebro e para testar hipóteses sobre o funcionamento do sistema nervoso central (VASCONCELOS *et al.*, 2011; WU, 2006).

Várias tecnologias estão fazendo uso de padrões de ativação neuronal para automação de tarefas (HAYKIN, 1999; YANG, 2008). Sistemas, tais como interfaces cérebro-máquina, trabalham diretamente para recuperar padrões de ativação neuronal e traduzem essa

informação em ações desejadas no mundo externo (DARMANJIAN, PRINCIPE, 2009).

## *1.2 Definição do Problema*

Sendo padrões de ativação neuronal tipos de dados capazes de serem classificados de acordo com a origem e determinados fatores controlados, a relação entre as classes de padrões de ativação neuronal tem recebido pouca atenção.

## *1.3 Objetivos*

Buscando colaborar para a solução desse problema, foram elaborados os seguintes objetivos de pesquisa:

**Objetivo geral:** Colaborar para a atividade de avaliar a similaridade entre classes de padrões de ativação neuronal.

**Objetivos específicos:**

- Escolher conjunto de fatores para definir as classes de padrões de ativação neuronal;
- Analisar características internas de classes padrões de ativação neuronal;
- Elaborar uma estratégia para mensurar níveis de similaridade entre classes de padrões de ativação neuronal; e
- Analisar a influência de determinados fatores para o nível de similaridade entre classes de ativação neuronal.



## *1.4 Escopo*

Esta pesquisa trabalhou com classes de padrões neuronais extraídas, com o uso de microeletrodos para a contagem de potenciais de ação, das regiões cerebrais, do hipocampo, do córtex somatossensorial primário e do córtex visual primário, relacionados às funções de processamento da memória de curta duração, do tato e da visão, respectivamente, de ratos da linhagem Long Evans, criados em cativeiro, durante a exploração de um ambiente no escuro povoado com objetos (bola, escova, comida e haste) com os quais os animais só tiveram contato no momento do experimento.

## *1.5 Resultados e Contribuições*

Os resultados identificados na pesquisa foram:

- Melhor identificação da atividade do cérebro na realização da atividade de compreensão de objetos;
- Entendimento de como é processado, em algumas regiões do cérebro, a semelhança entre determinadas classes de objetos no decorrer do tempo;
- Avaliação do processo de aquisição dos padrões de ativação neuronal utilizados nessa pesquisa; e
- Elaboração de uma estratégia para suporte à atividade de avaliação de padrões de ativação neuronais.

Essas informações podem cooperar para o entendimento do funcionamento do cérebro por neurocientistas e por desenvolvedores de aplicações usando padrões de ativação neuronal, uma vez que, a identificação de regiões do cérebro que possuem atividade semelhantes pode ser usada para reduzir a alocação de recursos ou para testes, ou ainda, a identificação da influência de determinados fatores na atividade do cérebro pode ser utilizada no projeto de sistema.

## *1.6 Estrutura da Dissertação*

Esta dissertação está organizada em demais capítulos que além do Capítulo 1, referente à Introdução deste trabalho, os quais são descritos a seguir.

O Capítulo 2 é intitulado Revisão de Literatura. Nele, são discutidos o contexto histórico das pesquisas realizadas na área deste trabalho de dissertação e alguns estudos correlacionados.

Os Materiais e Métodos de Avaliação utilizados na pesquisa estão registrados no Capítulo 3 desta dissertação.

No Capítulo 4, foi feita a Apresentação e Análise dos Resultados encontrados durante a realização do trabalho.

No Capítulo 5, são apresentadas as Considerações Finais e Sugestões para Trabalhos Futuros identificados ao término das atividades desta dissertação.

Por fim, são disponibilizadas as Referências Bibliográficas utilizadas para o embasamento do trabalho e o Apêndice contendo artefatos produzidos durante a pesquisa.

# Capítulo 2

## Revisão de Literatura

Este capítulo objetiva apresentar uma revisão da literatura relacionada à pesquisa realizada. Inicialmente, é apresentada uma breve contextualização da grande área de estudo, a qual é seguida de uma análise mais específica de trabalhos relacionados à análise de padrões de ativação neuronal e de agrupamentos entre tais padrões. Finalmente, são tecidas considerações finais sobre o capítulo.

### *2.1 Histórico*

No início do século XIX, foram realizadas as primeiras investigações modernas a respeito da função e do funcionamento do cérebro, bem como sobre sua composição. Do século XIX ao século XX, foram elaborados importantes conceitos da pesquisa sobre o cérebro, como a importância do neurônio para a codificação de informações, a localização de unidades funcionais no córtex, e o desenvolvimento de técnicas para a realização de experimentos neurofisiológicos em seres vivos (BREIDBACH, 2001).

Dentro desse período, é possível destacar alguns fatos considerados relevantes para o estado da arte em que esta pesquisa está inserida.

No período da Segunda Guerra Mundial, foi realizado um grande desenvolvimento em várias áreas do conhecimento (MENDELSON, SMITH, WEINGART, 1988), tais como a Neurociência.

Os aspectos que colaboram para esse fato foram o emprego de diversos cientistas, engenheiros e profissionais da área da saúde, em projetos interdisciplinares, como o objetivo de desenvolver técnicas no campo bélico, assistencial e comunicações, e a evolução dos computadores, que se tornaram capazes de realizar maior quantidade de cálculos e processamentos de sinais.

Entre as décadas de 50 e 60, foram iniciadas discussões sobre o quanto o cérebro é um órgão ativo e passivo ao mesmo tempo, isto é, ser um componente que age sobre os demais componentes do corpo para a realização de tarefas, e que sofre influência dos demais componentes do corpo para sua tomada de decisão (BREIDBACH, 2001).

Essas reflexões contribuíram para a valorização do contexto em que determinado indivíduo estava inserido, para a análise de seus dados cerebrais e processamento de informação.

Outro fato importante foi a comercialização do aparelho de tomografia, nos anos 70, que colaborou para a popularização do termo Neurociência, definida com a área do conhecimento que analisa informações cerebrais (PERKEL, BULLOCK, 1968).

No início dos anos 90, George Herbert Walker Bush, o então presidente dos Estados Unidos da América (EUA), proclamou a “Década do Cérebro”, com o objetivo de estimular cientistas a pesquisarem a respeito do funcionamento do cérebro (JONES, MENDELL, 1999).

Durante esses 10 anos, diversas empresas e instituições do governo norte-americano aumentaram seus investimentos em pesquisas sobre o sistema nervoso; havia incentivos como redução de impostos, disposição de crédito e aprovação àqueles pesquisadores que utilizavam métodos científicos para o entendimento da atividade do cérebro e da mente.

De maneira similar, em 2012, a União Europeia iniciou o financiamento de pesquisa relacionada ao mapeamento do cérebro, no projeto “Human Brain Project”<sup>1</sup>, que mobilizará centenas de instituições e custará cerca de 1,5 bilhão de euros.

Ainda mais recentemente, em 2013, Barack Hussein Obama II, atual presidente dos EUA, instituiu o projeto “*Brain Research through Advancing Innovative Neurotechnologies*”<sup>2</sup> (BRAIN Initiative), que irá investir nos próximos 10 anos, mais de 300 milhões de dólares por ano em pesquisas sobre o processamento de informações cerebrais.

Com isso, é perceptível que pesquisas sobre o processamento de informações cerebrais têm estado em grande crescimento ao longo da história.

<sup>1</sup> <http://www.nih.gov/science/brain/>

<sup>2</sup> <https://www.humanbrainproject.eu/>

Atualmente, uma das subatividades envolvidas no processamento de informações cerebrais é a análise de padrões de ativação neuronal, a qual foi realizada nesse trabalho de dissertação, e por esse motivo, na Seção 2.2, são apresentados trabalhos relacionados à realização desta subatividade.

## *2.2 Análise de Padrões de Ativação Neuronal*

Padrões de ativação neuronal são adquiridos por meio de ferramentas como microeletrodos, eletroencefalograma (EEG) ou a ressonância magnética (RM), para observar a atividade neuronal de um indivíduo ao realizar uma determinada atividade (CATTANEO, MAFFEI, MORRONE, 1981).

A análise desse tipo de dado tem sido realizada para inferências quanto ao funcionamento de cérebro (ROLLS, TREVES, 2011).

Nesse contexto, foram selecionados alguns trabalhos considerados os mais relacionados com essa pesquisa de dissertação, um para cada ano, ao longo dos últimos sete anos, data escolhida em virtude do ano do trabalho em que os dados utilizados nessa pesquisa de dissertação foram produzidos. Os trabalhos selecionados são sumarizados na Tabela 1, constituída pelas colunas descritas a seguir.

- **REFERÊNCIAS:** Citação de cada um dos artigos selecionados, presentes nas referências desse trabalho de dissertação.
- **POPULAÇÃO:** Registro da população analisada em cada artigo.
- **REGIÕES DO CÉREBRO:** Regiões do cérebro analisadas em cada indivíduo da população de cada artigo.
- **ANÁLISE:** Ferramentas utilizadas nas análises de cada artigo.
- **RESULTADOS:** Artefatos e conclusões produzidos em cada artigo.

Tabela 1 – Trabalhos relacionados – Análise de Padrões de Ativação Neuronal.

<b>REFERÊNCIAS</b>	<b>POPULAÇÃO</b>	<b>REGIÕES DO CÉREBRO</b>	<b>ANÁLISE</b>	<b>RESULTADOS</b>
RIBERIO <i>et al.</i> , 2007	Ratos	Hipocampo, Córtex Visual Primário e Córtex Somatossensorial Primário	AUROC	Análise de interações entre sono SWS e REM, e da propagação de mudanças sinápticas do hipocampo para o córtex.
YANG <i>et al.</i> 2008	Ratos	Córtex Auditivo Primário (CA)	Experimental	Diferenças de 3 ms no tempo de ativação neuronal podem ser usadas para orientar o comportamento.
NUO <i>et al.</i> , 2009	Macacos	Córtex Temporal Inferior (TI)	Taxa de acerto de classificação	A taxas de disparo dos neurônios individuais de TI pode ser bastante sensível a mudanças de imagem.
BOYER, JAMES, 2010	Ratos	Córtex Posterior Parietal e Córtex Perirhinal	ANOVA	Ratos são capazes de utilizar exclusivamente o sentido visual ou tátil para reconhecer objetos.
VASCONCELOS <i>et al.</i> , 2011	Ratos	Hipocampo, Córtex Visual Primário e	AUROC	Análise do nível de colaboração entre

		Córtex Somatossensorial Primário		regiões cerebrais na realização de tarefas
AZAMI, SANEI, 2012	-	-	Taxa de acerto de classificação	Três novos métodos de detecção dos potenciais de ação usando <i>Smoothed Nonlinear Energy Operator</i> (SNEO), <i>Fractal Dimension</i> (FD) e desvio padrão foram propostos.
TAVARES, 2013	Ratos	Hipocampo, Córtex Visual Primário e Córtex Somatossensorial Primário	AUROC	Imprecisões no registro de intervalo de contato de até três segundos não influenciam significativamente na capacidade de reconhecimento dos objetos.
JOACHIMSTHALER <i>et al.</i> , 2014	Ratos	Córtex Auditivo Primário (CA)	Teste de Kruskal-Wallis	Respostas no CA são bastante suscetíveis a influências de anestésicos.

Todos estes trabalhos, assim como esta pesquisa, utilizaram dados neuronais adquiridos a partir de matrizes de microeletrodos.

No trabalho de Ribeiro *et al.* (2007), os padrões de ativação neuronal de certas regiões do cérebro (Hipocampo, Córtex Visual Primário e Córtex Somatossensorial Primário) e suas características de plasticidade de neurônios foram observados ao longo do tempo, por meio da análise de classificadores, utilizando AUROC. Outro aspecto dessa pesquisa foi a contribuição quanto à observação do processamento distribuído do cérebro.

De acordo com o autor, as memórias episódicas e espaciais são construídas no hipocampo durante a aquisição, mas migram para o córtex cerebral ao longo do tempo. Foi proposto em outro de seus trabalhos que a interação entre ondas lentas (SWS) e movimento rápido dos olhos (REM) propaga mudanças sinápticas recentes do hipocampo para o córtex.

Com o intuito de testar esta teoria, no trabalho de Ribeiro *et al.* (2007), foram avaliados em conjunto atividade extracelular neuronal, potenciais de campo locais (LFP), e os níveis de genes imediatos precoce, relacionados à plasticidade (IEG) de arco e ZIF-268 em ratos expostos a uma experiência exploração espaço-tátil.

Os resultados do estudo apresentaram que as taxas dos disparos de pós-experiência foram mais fortes no SWS e permaneceram muito mais tempo no córtex (horas) do que no hipocampo (minutos). Durante o sono REM, as taxas de disparos mostraram forte dependência temporal entre as áreas do cérebro: a ativação cortical durante a experiência previu atividade do hipocampo na primeira hora pós-experiência, enquanto a ativação do hipocampo durante a experiência previu atividade cortical na terceira hora após a experiência. Quatro horas após a experiência, a expressão IEG foi especificamente regulada positivamente durante o sono REM no córtex, mas não no hipocampo. A quantidade do gene Arc no córtex foi proporcional à amplitude LFP na faixa do eixo (10-14 Hz), mas não às taxas de disparo, como se espera de sinais mais relacionados à entrada dendrítica que a saída somática.

O autor concluiu que existem indicadores de que a ativação hipocampal-cortical durante a vigília é seguida por várias ondas de plasticidade cortical de maneira em que os ciclos de sono completos repitam.

O trabalho de Ribeiro *et al.* (2007) possui grande relação com a presente pesquisa de dissertação, uma vez que os dados utilizados nesta pesquisa foram adquiridos por meio dos experimentos realizados naquele trabalho.



Dessa forma, esta pesquisa complementa o trabalho apresentado anteriormente, realizando novas análises dos dados.

A obra de Yang *et al.* (2008) apresenta um estudo sobre o menor intervalo de tempo que é necessário para observar os padrões de ativação neuronal do córtex auditivo para a classificação de atividade. Nesse trabalho foram implantados pares de microeletrodos para estimular os neurônios no córtex auditivo rato diretamente e foi descoberto que os ratos podem explorar as diferenças nos intervalos de tempo de atividade cortical que sejam tão curtos quanto 3 ms para orientar as decisões.

De maneira similar, nesta pesquisa de dissertação, também foi realizada análise de fatores que podem influenciar na classificação de padrões de ativação neuronal, dentre eles o intervalo de tempo em que determinada atividade eletrofisiológica ocorreu.

A análise de influências de fatores à qualidade de classificação de padrões de ativação neuronal também é uma questão investigada no trabalho de Nuo *et al.* (2009).

Em especial, o trabalho de Nuo *et al.* (2009) realiza inferências sobre a atividade de reconhecimento de objetos, entre elas, que a proximidade e a similaridade entre objetos possuem grande influência para a taxa de acertos de classificação. Sendo esse último fator também analisado nessa pesquisa.

De acordo com o autor, primatas podem facilmente identificar objetos visuais em grandes mudanças na posição - propriedade comumente referida como "invariância de posição". Esta capacidade é amplamente assumida como dependente de neurônios no córtex temporal inferior (TI), que pode responder seletivamente a objetos visuais isolados ao longo de grandes intervalos de posição da retina, da mesma forma.

Nuo *et al.* (2009) também afirmam que no mundo real, os objetos raramente aparecem de forma isolada, e da interação entre a posição invariante e a representação de vários objetos (ou seja, a desordem). Nesta questão é intuitivo que as representações de objetos próximos podem interferir uns com os outros e que os grandes campos receptivos necessários para a invariância de posição podem agravar esse problema aumentando o intervalo no qual a interferência atua.

Os autores relatam que a maioria das respostas dos neurônios de TI está fortemente afetada pela presença de desordem, embora os mecanismos externos (como atenção) sejam muitas vezes invocados como uma saída para o problema.

Nesse contexto, o trabalho apresentou (usando dados e simulações neuronais gravadas) que as propriedades intrínsecas das respostas de neurônios de TI, por si só, podem suportar o reconhecimento de objetos com limitada desordem. Além disso, foram realizadas extensas simulações de populações neuronais hipotéticas para identificar os neurônios mais representativos de uma população.

Como resultado, as simulações mostram que a propriedade fundamental neuronal para o reconhecimento na desordem não é a preservação da magnitude da resposta, mas a preservação da preferência de cada neurônio

No trabalho de Boyer e James (2010), é realizada uma análise de como uma abordagem *cross-modal* (correlacionando o comportamento entre áreas do cérebro) pode colaborar para a atividade de reconhecimento de objetos com base neurofisiológica.

Os resultados desse trabalho sugerem que o desempenho espontâneo tátil-visual é guiado por uma representação de objeto que é distribuída entre os córtices parietais perirhinal e posterior.

O estudo da atividade de reconhecimento de objetos também é realizado neste trabalho de dissertação, analisando o nível de distinção dos objetos entre padrões de ativação neuronal adquiridas de assembleias de neurônios de regiões distintas do cérebro. A avaliação da similaridade, também realizada nesse trabalho de dissertação, contribuirá para mensurar o quão são distribuídas essas representações no cérebro de ratos.

Na pesquisa de Vasconcelos (2011), é considerada a possibilidade de classificar comportamentos com base nos mesmos padrões de ativação neuronal utilizados neste estudo. Sua pesquisa teve como objetivo investigar se respostas crossmodais em áreas sensoriais primárias podem representar informações comportamentalmente relevantes. Para isso, foi empregada a mesma base de dados utilizada no trabalho de dissertação, no tocante às gravações neuronais no córtex somatossensorial primário e no córtex visual primário de ratos.

Foi observado que, durante a livre exploração de novos objetos, as respostas do Córtex

Visual Primário (V1) e do Córtex Somatossensorial Primário (S1) apresentaram quantidades comparáveis de informações sobre a identidade do objeto. Durante a execução de uma tarefa de livre discriminação tátil, recrutamento era mais lento e menos robusto em V1 que em S1. No entanto, as respostas de V1 foram significativamente correlacionadas com o desempenho durante as sessões.

Os resultados suportam a noção de que as áreas sensoriais primárias têm preferência por uma determinada modalidade, mas pode envolver-se em processamento de cross-modal significativa, dependendo da demanda de tarefas, isso contribui quanto à observação do comportamento distribuído e de realocação do processamento de informações no cérebro.

No trabalho de Azami e Sanei (2012), foram analisadas formas de reduzir a influência do ruído em dados neuronais para a contabilização de potenciais de ativação. São propostas três novas abordagens baseadas em operador de suavização de energia não-linear, dimensão fractal e desvio padrão para detectar os pontos de dados neuronais ruidosos.

Embora em muitos casos, especialmente quando existem várias fontes de ruído, estes métodos podem não ser aceitáveis como os detectores de pico, os autores usaram filtro Savitzky-Golay e transformada discreta de wavelet (DWT) como etapas de pré-processamento para superar esse problema.

Os resultados da pesquisa mostram que, quando há muito ruído no sinal, o método proposto utilizando o desvio padrão e DWT pode detectar os picos melhor do que os outros métodos. A taxa de detecção média e falsa detecção de picos para o método proposto com base no desvio padrão e DWT são, respectivamente, 100% e 43% para os sinais de semi-reais com  $SNR = -5$  dB.

Na presente pesquisa de dissertação, de forma similar, é proposta uma abordagem que também possui como característica a redução do impacto de ruídos para o reconhecimento de classes de padrões de ativação neuronal.

No trabalho de Tavares (2013) é analisado o impacto de imperfeições na aferição do intervalo de contato de ratos com objetos. Os dados utilizados nesse trabalho também são os mesmos utilizados na presente pesquisa.

Seus resultados apontaram que, a partir de uma abordagem supervisionada, existe uma relação inversamente proporcional entre a qualidade de classificação dos padrões de ativação

neuronal, mensurada pela área embaixo da curva ROC (AUROC), e o tamanho da imprecisão do intervalo de contato, sendo imprecisões inferiores a dois segundos não significativos.

Outro resultado desse trabalho foi a comparação entre classificadores (Naïve Bayes, Árvore de Decisão, Máquina de Vetores de Suporte e Redes de Função de Base Radial). A análise das médias mostrou que o classificador Naïve Bayes apresenta melhores resultados, em comparação aos demais, e que a Árvore de Decisão apresenta resultado inferior. Rede RBF e Máquina de Vetores de Suporte apresentaram resultados próximos, segundos as medidas e os testes realizados.

Nesta pesquisa de mestrado, entre os fatores estudados, também é considerado o tempo de contato em que o animal esteve com determinado objeto para a construção dos padrões de ativação. Com isso, uma vez utilizadas abordagens não-supervisionada e semi-supervisionada, os resultados dessa pesquisa colaboram com as inferências do trabalho apresentado, pois apresentam outras características do mesmo dado utilizado.

No trabalho de Joachimsthaler et al. (2014), são analisadas propriedades de padrões de ativação neuronal, com o intuito de desvendar suas influências para a classificação desses padrões.

Segundo o autor, estudos de respostas neuronais no córtex auditório primário (CA) com o indivíduo acordado são importantes, a fim de estabelecer relações adequadas entre os dados neuronal e percepção auditiva.

Com isso, é apresentado nesse trabalho um estudo comparativo sobre a representação dos parâmetros de resposta neuronal para tons em áreas corticais auditivas primárias e de ordem superior de ratos acordados.

Foram quantificados 12 propriedades neuronais de processamento de tom, com o intuito de calcular as similaridades e diferenças de função entre os campos, e para discutir como função do CA no rato é comparável ao de macacos e ao de gatos acordados.

Foram feitas a partir de 1.400 pequenos grupos de neurônios, registros das camadas corticais III / IV no campo auditivo primário, no campo auditivo anterior, no segundo campo auditivo e no campo dorsoposterior.

Os resultados da pesquisa demonstram relação da atividade espontânea com a especificidade de campo, a correlação entre a atividade espontânea e evocada, a latência de resposta de tons, a nitidez de sintonia de frequência, os padrões de resposta temporais e grau de variação entre a frequência característica e a melhor frequência. Análises comparativas e quantitativas mostraram que as características de resposta medidas foram, em vários graus, suscetíveis a influências de anestésicos.

O trabalho de Joachimsthaler *et al.* (2014) possui relação com nossa pesquisa, uma vez que nossa pesquisa também é realizada análise de fatores que podem influenciar no comportamento da atividade eletrofisiológica no córtex e, assim como na pesquisa citada, nesse trabalho de dissertação é utilizado o Teste de Kruskal-Wallis para análise de distribuições que representam tratamentos, conforme descrito na Seção 2.3.6.

Uma atividade bastante comum na área de Mineração de Dados, em Ciências da Computação, é a análise de agrupamentos. Neste trabalho de dissertação, a análise de agrupamentos foi utilizada para mensurar as relações de similaridade entre classes de padrões de ativação neuronal e, com isso, para o processamento de informações cerebrais.

Na seguinte seção, são apresentados trabalhos relacionados a esta pesquisa que utilizaram a análise de agrupamentos.

### *2.3 Análise de Agrupamentos*

A análise de agrupamento é uma atividade importante no campo da análise exploratória de dados e está sendo aplicada em uma variedade de disciplinas de engenharia e em pesquisas científicas. Muitas vezes, é relacionada com o termo “aprendizagem não supervisionada” na literatura de reconhecimento de padrões e inteligência artificial (JAIN, DUBES, 1988).

Uma aplicação típica desse tipo de análise é a comparação de algoritmos de agrupamento, que pode ser utilizada em diversos domínios, tais como agrupamento de faces, músicas, filmes, entre outros (MOURA, 2012).

Nesse contexto, foram selecionados alguns trabalhos relacionados, realizados ao longo dos últimos sete anos, conforme descrito na seção anterior, que são apresentados na Tabela

2, na mesma estrutura da Tabela 1.

Tabela 2 – Trabalhos relacionados - Análise de Agrupamento.

<b>REFERÊNCIAS</b>	<b>POPULAÇÃO</b>	<b>REGIÕES DO CÉREBRO</b>	<b>ANÁLISE</b>	<b>RESULTADOS</b>
JOSHUA <i>et al.</i> , 2007	Macacos	Globo Pálido, Núcleos da Base, Córtex Motor Primário	Comparação de agrupamento	Medidas quantitativas de isolamento de potencial de ação podem ser obtidos independentemente do método utilizado para a detecção e classificação.
FLETCHER <i>et al.</i> , 2008	-	-	Características de Análise de agrupamento de dados neuronais	Descrição do projeto CARMEN, de sua infraestrutura e de uma ferramenta do projeto para visualização de padrões.
DARMANJIAN, PRINCIPE, 2009	Ratos	Córtex Motor Primário	Comparação de agrupamento	Modelo de agrupamento hierárquico não supervisionado para representação de dados neurais espaço-temporal.
NARIHISA <i>et al.</i> 2010	Macacos	Córtex Temporal Inferior	Comparação de agrupamento	Algoritmo de agrupamento para redução de dimensões na construção de padrões de ativação neuronal.
ROLLS, TREVES,	-	-	Características de	Descrição de análises

2011			Análise de agrupamento de dados neuronais	quantitativas teóricas de codificação neural.
BESTEL, DAUS, THIELEMANN, 2012	-	-	Comparação de agrupamento	Abordagem para avaliação de distribuição de probabilidade das características de potenciais de ação com a finalidade de determinar os candidatos mais adequados para a melhora da classificação.
JAMMALAMADAKA, <i>et al.</i> , 2013	Ratos	Hipocampo	K-função	Análise da distribuição da população de espinhas dendríticas em neurônios dissociados do hipocampo.
SHALCHYAN, FARINA, 2014	Ratos	Hipocampo	Comparação de agrupamento	Abordagem bayesiana de agrupamento que não faz suposições sobre a distribuição dos aglomerados e usa estimativa dos grupos em cada intervalo de tempo como uma prévia para a classificação bayesiana dos dados no intervalo de tempo subsequente.

No trabalho de Joshua *et al.* (2007) é proposta uma abordagem para mensurar a qualidade de dados neuronais, por meio da comparação de agrupamentos de ativação e de ruído.

Segundo os autores, já foram propostas muitas abordagens para o problema de detecção e classificação de potenciais de ação registrados extracelularmente, mas apenas alguns métodos quantificam a qualidade deste processo fundamental. Na maioria dos casos, a avaliação da qualidade é baseada no julgamento subjetivo de observadores humanos e em unidades registradas divididas em grupos denotados como "isolados" e "não isolados".

Para os autores, esta avaliação subjetiva impede avaliação abrangente dos estudos unitários desde o parâmetro mais básico, ou seja, a qualidade de dados não é explicitamente definida.

Nesse contexto, no trabalho são propostas medidas objetivas para avaliar a qualidade dos dados de potenciais de ação com base no momento em que os potenciais foram detectados e a amostragem do sinal analógico de dados corticais e basal-gânglios de alta frequência.

No trabalho também foi utilizado um algoritmo vizinhos mais próximos para estimar a proporção de erros de classificação positiva e falsos negativos.

Os resultados do trabalho mostram que a quantificação da qualidade de gravação pela relação sinal-ruído pode ser enganadora. Esta é melhor avaliada por um escore de isolamento que mede a sobreposição entre o ruído e os grupos de potenciais de ação. Para validar essas medidas de qualidade, foram simulados a detecção de potenciais de ação e os erros de classificação, o que apresentou que as pontuações são bons preditores da frequência de erros.

A confiabilidade da pontuação de isolamento também foi verificada por erros implantados em dados reais e dos gânglios da base, utilizando diferentes algoritmos de classificação.

Os autores concluem que medidas quantitativas de isolamento de pico podem ser obtidas independentemente do método utilizado para a detecção de potenciais de ação e classificação, e recomendam os seus relatórios em qualquer estudo com base na atividade de



neurônios individuais.

O trabalho de Joshua *et al.* (2007) possui relação com a presente pesquisa, também assume que por meio de algoritmos de agrupamento é possível a separação de dados representativos de dados não representativos e que a comparação entre esses dois grupos dá suporte à análise de qualidade dos dados.

No trabalho de Fletcher *et al.* (2008), a análise de agrupamento é utilizada para detecção de potenciais de ação, sendo analisada a eficácia de se utilizar essa atividade para esse fim. Nesse trabalho foi utilizado um arcabouço de desenvolvimento e análise intitulado CARMEN. Esse arcabouço tem como objetivo disponibilizar um laboratório de neurociência virtual: uma infraestrutura para compartilhamento de dados, ferramentas e serviços.

Neste trabalho, também são descritos o projeto CARMEN, bem como a infraestrutura de nó e uma ferramenta cliente associado para visualização de padrões e realização de pesquisas, o Data Explorer Signal (SDE), e novos métodos de detecção de potenciais de ação, que são fundamentais para os serviços prestados por CARMEN.

Da mesma forma que no trabalho de Fletcher *et al.* (2008), características, como tamanho do agrupamento e homogeneidade do agrupamento, também são analisadas nesse trabalho de dissertação. As discussões sobre os métodos do projeto também foram importantes para a construção dos métodos computacionais implementados nesse trabalho de dissertação.

No trabalho de Darmanjian e Principe (2009), análise de agrupamentos foi utilizada para implementar uma interface de comunicação com o cérebro para classificar novos padrões de ativação neuronal em grupos de ações, pré-determinados, de acordo com a ação que o indivíduo gostaria de realizar com um braço, o qual não possui.

Nesse trabalho é empregado um modelo de agrupamento hierárquico não supervisionado visando descobrir tanto as interdependências entre canais neurais e os aglomerados auto-organizados representados nos dados neurais espaço-temporal.

Assim como no trabalho de Darmanjian (2009), nesta pesquisa de dissertação é realizado agrupamento de padrões de ativação neuronal, usando o mesmo algoritmo de agrupamento (K-Means), e são estudadas as distâncias entre os agrupamentos.

Na obra de MATSUMOTO *et al.* (2010), os grupos criados a partir de padrões de ativação neuronal, foram analisados por meio de métricas de avaliação externa, com o objetivo de encontrar o melhor algoritmo de agrupamento de acordo com um agrupamento de controle. Matsumoto comparou o desempenho de um algoritmo de agrupamento hierárquico com um algoritmo de agrupamento hierárquico otimizado.

Nesse trabalho foi utilizado um algoritmo de Bayes para realizar agrupamentos e redução de dimensão simultaneamente. Foram utilizados dados neuronais artificiais e reais para analisar o desempenho e os resultados demonstraram melhores agrupamentos com o algoritmo otimizado do que em um subespaço que foram obtidos pela análise de componentes principais.

Nessa dissertação, depois de agrupar os padrões de ativação neuronal, foram usadas métricas de avaliação interna e também avaliados os desempenhos de algoritmos de agrupamento.

Estas métricas são utilizadas para avaliar o desempenho dos algoritmos de agrupamento, no entanto, as métricas externas consistem em comparar o agrupamento gerado com alguma informação a priori, e as métricas internas consistem em avaliar o ajuste entre as estruturas, utilizando apenas os dados propriamente ditos (JAIN, DUBES, 1988).

A relevância do uso de abordagens não supervisionadas para extração de características intrínsecas dos dados neuronais é um dos temas discutidos no trabalho de Rolls e Treves (2011). Desta forma, esse trabalho comunga das ideias do trabalho de dissertação, usadas para justificar o objetivo de análise dessa pesquisa.

Rolls e Treves (2011) descrevem resultados de análises teóricas de informação quantitativa de codificação neural, particularmente nas áreas visual, olfativo, degustativo, do hipocampo e córtex orbitofrontal. De acordo com os autores, a maior parte das informações é codificada pelas taxas de disparo dos neurônios, e essa codificação é considerada robusta para as representações de taxa de queima de diversos neurônios independente de pequenas populações de neurônios.

Além disso, na obra de Rolls e Treves (2011), é afirmado que a informação pode ser lida rapidamente (em menos de 20 milissegundos) a partir de tal codificação.

Outro ponto discutido é a pouca quantidade encontrada de pesquisas envolvendo

sincronização dependente do estímulo de diferentes neurônios, ou o momento dos potenciais de ação dentro de um *bin* único neurônio. Também é defendido que a decodificação do produto escalar da atividade sináptica de neurônios é biologicamente plausível.

Por fim, os autores concluem que o entendimento do código neural é fundamental para entender não só como o córtex o representa, mas também os processos de construção de informações.

No trabalho de Bestel, Daus, Thielemann (2012), é proposto um novo algoritmo de agrupamento capaz de se adaptar a novas demandas de análise de características reduzindo a perda computacional, já executada. Com o intuito de validar o novo algoritmo, este tem o seu desempenho analisado com base em métricas externas.

Outro exemplo de validação de resultados adquiridos por meio de uma abordagem não supervisionada é o trabalho de Jammalamadaka (2013). Este o faz comparando o número de grupos encontrados pelo algoritmo de agrupamento com a K-função (OKABE, YAMADA, 2001).

Segundo o autor, embora em constante progresso, uma tarefa desafiadora que ainda permanece na área da Neurociência Computacional é o desenvolvimento de algoritmos de ordenação de potenciais de ação eficientes que permitam uma análise precisa do sinal ao nível de uma única célula.

No trabalho, é afirmado que a maioria dos algoritmos de ordenação disponíveis atualmente apenas extrai um tipo de recurso específico, como os principais componentes ou coeficientes Wavelet dos sinais de potenciais de ação medidos, a fim de separar as diferentes formas de pico gerados por diferentes neurônios, embora que, devido à grande variedade nas formas de potenciais de ação obtidos, a derivação de um conjunto de recursos ideal ainda é uma questão muito complexa.

Com o intuito de colaborar para a solução dessa problemática, nesse trabalho é proposto um algoritmo que extrai uma variedade de recursos baseados em componentes geométricos, Wavelet e principais componentes, e deriva automaticamente um subconjunto de características, mais adequado para classificar um conjunto individual de sinais de potenciais de ação.

Os resultados dos testes com arquivos de dados simulados e dados obtidos a partir de

matrizes de microeletrodos demonstraram desempenho superior da abordagem proposta pelos autores.

Por fim, o trabalho de Shalchyan e Farina (2014) propõe um algoritmo de agrupamento baseado no algoritmo Bayseano (GELMAN *et al.*, 2013). No trabalho de Shalchyan e Farina também é realizada a validação dos resultados por meio de métricas externas de avaliação de agrupamento.

De acordo com os autores, potenciais de ação de múltiplos neurônios registrados em um sinal são normalmente separados por agrupamento. Nesse contexto, o trabalho de Shalchyan e Farina (2014) apresenta uma abordagem bayesiana de agrupamento que não faz suposições sobre a distribuição dos aglomerados e usa, baseado em densidade kernel, estimativa de densidade dos agrupamentos em cada intervalo de tempo como uma prévia para a classificação bayesiana dos dados no intervalo de tempo subsequente.

O método proposto foi testado e comparado à abordagem baseada em modelo de Gauss para o rastreamento de grupo usando conjuntos de dados simulados e experimentais de ratos.

Os resultados descritos na pesquisa demonstram que a estimativa não-paramétrica de densidade no uso de algoritmo proposto superou o modelo gaussiano sequencial de encaixe em ambos os testes de dados simulados e experimentais.

Os autores concluem que o uso de agrupamento baseado em densidade kernel não-paramétrica que não faz suposições sobre a distribuição dos aglomerados aumenta a capacidade de rastrear aglomerado não estacionários ao longo do tempo em relação à abordagem de modelagem baseada em conjuntos de Gauss.

## 2.4 Considerações sobre o Capítulo

De acordo com a revisão de literatura realizada e considerando os trabalhos relacionados a esta pesquisa de dissertação, apesar da crescente valorização e quantidade de pesquisas na área de Processamento de Informações Cerebrais, ainda existe uma lacuna na utilização de análise de agrupamentos de padrões de ativação neuronal, no que diz respeito à análise de similaridade entre classes desse tipo de dado cerebral.

Nesse contexto, justifica-se a intenção da presente pesquisa em colaborar para o preenchimento dessa lacuna.

No Capítulo 3 são apresentados os materiais e métodos empregados na realização dessa pesquisa.

# Capítulo 3

## Materiais e Métodos de Avaliação

Neste capítulo é apresentado o esquema da estratégia proposta, as ferramentas, os métodos e os dados utilizados durante o estudo e a realização das abordagens investigadas (não supervisionada e semi-supervisionada) para a solução do problema objeto desta dissertação, o qual consiste na análise de similaridade entre classes de padrões de ativação neuronal, e o teste de robustez das inferências com relação à qualidade dos dados.

### 3.1 Esquema da estratégia proposta

O esquema da estratégia proposta nesse trabalho de dissertação, cujo objetivo é analisar os níveis de similaridade entre classes de padrões de ativação neuronal, é ilustrado na Figura 1.

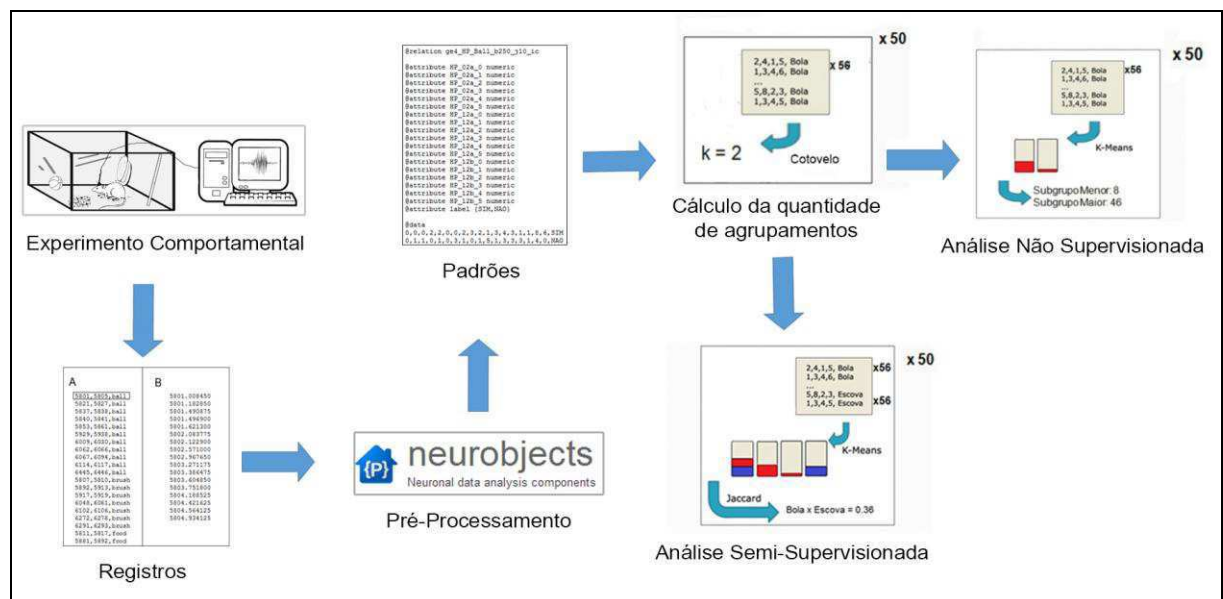


Figura 1: Esquema da estratégia proposta

De acordo com a Figura 1, a estratégia proposta é composta por sete fases: Experimento Comportamental, Registros, Pré-Processamento, Padrões, Cálculo da quantidade de agrupamentos, Análise Semi-Supervisionada e Análise Não Supervisionada.

Na fase do experimento comportamental, foi gerada a base de dados com os registros dos comportamento do experimento. Esses dados foram pré-processados utilizando a ferramenta Neuroobjects para a construção dos padrões. Em seguida foram calculadas as quantidades de agrupamentos presentes em cada classe de padrões de ativação neuronal, após isso, essa informação foi utilizada como entrada na realização das análises não supervisionada e semi-supervisionada.

Os detalhes de implementação de cada fase serão apresentados nas seções a seguir.

## 3.2 *Ferramentas*

As ferramentas utilizadas nesse trabalho de dissertação são apresentadas nas subseções seguintes, bem como a finalidade em que foram empregadas. São elas: Neuroobjects, MySQL, Weka e R.

### 3.2.1 Neuroobjects

O Neuroobjects<sup>3</sup> é uma biblioteca desenvolvida, na linguagem de programa Java, pelo professor Nivaldo Vasconcelos, durante o seu doutorado, com o objetivo de manipular a base de dados neuronais produzida no estudo do professor Ribeiro (2007). Essa ferramenta está disponível na internet para download. Neste trabalho de dissertação, foi utilizado o programa dataset-generator.jar para a criação do banco de dados em MySQL e construção dos padrões de ativação neuronal em arquivos .aff.

Esta biblioteca contém um conjunto de rotinas comumente usado para fazer a análise de dados neuronal. A biblioteca é dividida nos seguintes componentes principais:

<sup>3</sup> <https://code.google.com/p/neuroobjects/> (acessado em 22/07/2014)

- ***SpikeHandler***: Realiza uma representação independente da fonte de um conjunto de treino de potenciais de ação;
- ***SpikeRateMatrix* e *CountMatrix***: Fornece uma representação abstrata de um modelo para estimar a função de taxa de potenciais de ação. Além do *CountMatrix*, outros modelos podem ser implementados, tal como um modelo gaussiano;
- ***BehaviorHandler***: Fornece uma representação independente da fonte dos dados de comportamento de um animal; e
- ***PatternHandler***: Fornece uma abstracção útil de um conjunto de padrões de taxa de potencial de ação, tal como descrito em *CountMatrix*.

A biblioteca também é constituída de componentes para aplicações de técnicas de aprendizado supervisionado e para identificar o comportamento da atividade do neurônio (o pacote `nda.analysis`).

### 3.2.2 MySQL

O MySQL<sup>4</sup> é um sistema de gerenciamento de banco de dados que utiliza a linguagem de programação de banco de dados SQL como interface, desenvolvida por David Axmark, Allan Larsson e Michael "Monty" Widenius, pertencente a empresa Oracle<sup>5</sup>. Esse sistema é utilizado por diversas empresas, como Yahoo! Finance, MP3.com, Motorola, NASA, Silicon Graphics e Texas Instruments em aplicações de missão crítica. Dentre as várias edições disponíveis do sistema, foi utilizada a edição "Community Server", versão 5.7.1. Em nosso trabalho, esta ferramenta é empregada com o intuito de armazenar e viabilizar a realização de consultas diretas aos dados neuronais.

<sup>4</sup> <http://www.mysql.com/> (acessado em 22/07/2014)

<sup>5</sup> <http://www.oracle.com/br/> (acessado em 22/07/2014)



### 3.2.3 Weka

O Weka<sup>6</sup> é um ambiente de desenvolvimento escrito na linguagem de programação Java, construído na Universidade de Waikato, Nova Zelândia entre os anos de 1993 e 2006. É licenciado à General Public License, sendo, portanto, possível estudar e alterar o respectivo código fonte, e agrega algoritmos provenientes de diferentes abordagens da inteligência artificial.

Desse pacote, utilizamos os algoritmos de cálculo de distância entre dois pontos (distância Euclidiana) (JAIN, DUBES, 1988), de agrupamento de dados baseado em centroide (*K-Means*) (MACQUEEN, 1967), de agrupamento de dados baseado em distribuições (*Expectation-Maximization*) (DEMPSTER, LAIRD, RUBIN, 1977), de agrupamento hierárquico de dados (Simple-Linkage) (JAIN, DUBES, 1988), e de agrupamento de dados baseado em densidade (*Density-based spatial clustering of applications with noise*) (ESTER *et al.*, 1996), implementados nas classes `weka.core.EuclideanDistance`, `SimpleKMeans`, `EM`, `HierarchicalClusterer` e `MakeDensityBasedClusterer` do pacote `weka.clusterers`, respectivamente, evocadas diretamente no código escrito para a realização dos experimentos dessa pesquisa de dissertação.

### 3.2.4 R

R<sup>7</sup> é um ambiente de desenvolvimento expansível para cálculos estatísticos, bastante utilizado por estatísticos e desenvolvedores de software que realizam análise de dados (VANCE, 2009; FOX, 2005).

Criado por Ross Ihaka e Robert Gentleman do departamento de Estatística da Universidade de Auckland, Nova Zelândia, o código fonte do R está disponível sob a licença

<sup>6</sup> <http://www.cs.waikato.ac.nz/ml/weka/> (acessado em 22/07/2014)

<sup>7</sup> <http://www.r-project.org/> (acessado em 22/07/2014)

GNU GPL<sup>8</sup> para os sistemas operacionais Windows, Unix/Linux e Macintosh.

Nesta pesquisa, foram utilizados os testes estatísticos de Shapiro-Wilk (1965) para análise de normalidade das distribuições, implementado na função `shapiro.test(x)`, de Bartlett (1937) para análise de variância, implementado na funções `bartlett.test(y,x)` e o de Kruskal-Wallis (1952) para análise de semelhança entre distribuições, implementado nas função `kruskal.test(y,g)` e otimizado para múltiplas comparações na função `kruskalmc(y,g)` do pacote `pgirmess`.

### 3.3 *Dados e Experimento Comportamental*

Os dados analisados neste estudo foram adquiridos durante um experimento envolvendo exploração livre de objetos por ratos (RIBEIRO *et al.*, 2007).

Os ratos que foram utilizados eram da linhagem Long Evans, criados em cativeiro, estes, antes de qualquer contato com objetos do experimento (os quais eram todos desconhecidos para os animais), foram submetidos a uma cirurgia para o implante dos microeletrodos nas regiões do córtex visual primário (V1), do córtex somatossensorial primário (S1) e do hipocampo (HP).

Após a recuperação dos animais, os sinais referentes às atividades elétricas do conjunto de neurônios selecionado eram registrados a partir dos microeletrodos. Os sinais captados eram processados para a contagem dos potenciais de ação de cada neurônio e armazenados.

O conjunto de dados nessa dissertação é formado pelo vetor dos registros de tempo de disparo de cada neurônio e dos registros de atividades as quais cada animal estava realizando em cada momento nas fases do experimento.

O registro de atividade de cada animal foi feito por um observador humano, que manualmente anotou os intervalos de contato, a partir do registro de vídeo feito por uma câmera com luz infravermelha

Em todas as fases do experimento, o animal era monitorado em todos os seus estados

<sup>8</sup> <http://www.gnu.org/copyleft/gpl.html> (acessado em 22/07/2014)

comportamentais: vigília (WK), sono de ondas lentas (SWS), e sono onde ocorre o movimento rápido dos olhos (REM); em um ambiente em ausência de luminosidade.

Na Figura 2 são apresentadas as três fases (PRÉ-EXP, EXPOSIÇÃO, PÓS-EXP) do experimento. Na PRÉ-EXP, os animais foram colocados no ambiente de experimentação duas horas antes do contato com os objetos. Esta fase tinha como objetivo deixar o animal se acostumar com o ambiente, tornando a exposição dos objetos a única novidade.



Figura 2: Ilustração das três fases do experimento (PRÉ-EXP, EXPOSIÇÃO, PÓS-EXP). Adaptada de (RIBEIRO *et al.*, 2007)

Na EXPOSIÇÃO, foram inseridos os objetos (uma bola, uma escova, uma quantidade de ração e uma haste com elementos pontiagudos distribuídos ao longo de sua altura) nos quatro cantos do ambiente.

Nessa fase também foi realizada a observação das ativações neuronais nas regiões S1, V1, HP, enquanto o animal realizava contato com cada um dos objetos, durante 20min.

Na PÓS-EXP, foram removidos os objetos e foi mantido o animal no ambiente, durante 3h a 48h. Esta fase tinha como objetivo estudar vestígios dos estímulos encontrados durante a vigília anterior, especialmente durante o sono (RIBEIRO *et al.*, 2007).

Os dados experimentais abordados nesta dissertação atêm-se exclusivamente à fase EXPOSIÇÃO, na qual há contato do animal com algum dos objetos.

Na Figura 3, há uma ilustração do processo de registro da atividade neuronal.

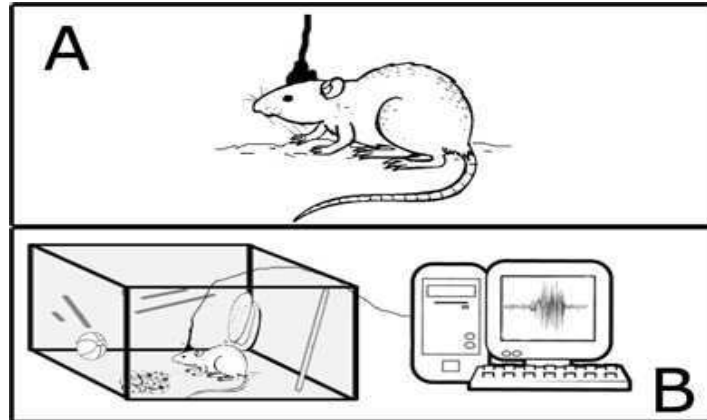


Figura 3: Ilustração do processo de registro da atividade neuronal. (A) Animal com microeletrodos implantados. (B) Visão geral do sistema: Animal no ambiente povoado com objetos, ligado ao programa de processamento dos sinais elétricos de cada neurônio monitorado. Adaptada de (RIBEIRO *et al.*, 2007).

Na Figura 3.A, pode ser visto um animal com um conector captando os sinais dos microeletrodos implantados. Na Figura 3.B, tem-se uma visão geral do experimento comportamental: animal no ambiente povoado com objetos, ligado ao programa de processamento dos sinais elétricos de cada neurônio monitorado.

Esse experimento (RIBEIRO *et al.*, 2007), o qual originou os dados em análise, foi realizado com um total de 4 animais (Ge4, Ge5, Ge9 e Ge12), todos criados em cativeiro, sem nunca terem qualquer contato com os objetos antes do momento do experimento.

Os dados foram disponibilizados pelo Professor Dr. Sidarta Ribeiro da Universidade Federal do Rio Grande do Norte (UFRN) - Brasil, armazenados em arquivos digitais, registros de ativação de cada neurônio e registros de intervalo de contato de cada animal com determinados objetos, no formato texto, os quais são exemplificados na Figura 4.

A	B
5801, 5805, ball	5801.008450
5821, 5827, ball	5801.182850
5837, 5838, ball	5801.490875
5840, 5841, ball	5801.496900
5853, 5861, ball	5801.621300
5929, 5938, ball	5802.083775
6009, 6030, ball	5802.122900
6062, 6066, ball	5802.571000
6067, 6094, ball	5802.967650
6114, 6117, ball	5803.271175
6445, 6446, ball	5803.386475
5807, 5810, brush	5803.604850
5892, 5913, brush	5803.751800
5917, 5919, brush	5804.188525
6048, 6061, brush	5804.421625
6102, 6106, brush	5804.564125
6272, 6278, brush	5804.934125
6291, 6293, brush	
5811, 5817, food	
5881, 5892, food	

Figura 4: Exemplos de arquivos disponibilizados para análise. (A) Registros de intervalo de contato de animal com determinados objetos. (B) Registro de ativação de neurônio.

Na Figura 4.A está destacado um intervalo em que o animal esteve em contato com uma bola. De acordo com a Figura 4.B, o animal teve durante esse intervalo de contato 17 potenciais de ação.

Com base nesses arquivos disponibilizados e utilizando o sistema Neuroobjects (VASCONCELOS et al, 2011), descrito na Seção 3.1.1, foram produzidos novos arquivos contendo padrões de ativação neuronal, os quais foram analisados nesse trabalho de dissertação. Um exemplo desse tipo de arquivo é ilustrado na Figura 5.

```

@relation ge4_HP_Ball_b250_j10_ic

@attribute HP_02a_0 numeric
@attribute HP_02a_1 numeric
@attribute HP_02a_2 numeric
@attribute HP_02a_3 numeric
@attribute HP_02a_4 numeric
@attribute HP_02a_5 numeric
@attribute HP_12a_0 numeric
@attribute HP_12a_1 numeric
@attribute HP_12a_2 numeric
@attribute HP_12a_3 numeric
@attribute HP_12a_4 numeric
@attribute HP_12a_5 numeric
@attribute HP_12b_0 numeric
@attribute HP_12b_1 numeric
@attribute HP_12b_2 numeric
@attribute HP_12b_3 numeric
@attribute HP_12b_4 numeric
@attribute HP_12b_5 numeric
@attribute label {SIM,NAO}

@data
0,0,0,2,2,0,0,2,3,2,1,3,4,3,1,1,8,6,SIM
0,1,1,0,1,0,3,1,0,1,5,1,3,3,3,1,4,0,NAO

```

Figura 5: Exemplo de arquivo contendo padrões de ativação neuronal.

O arquivo está dividido em três blocos (título, atributos, dados). No bloco de Título (*@relation*), é registrada a descrição do tratamento cujos padrões correspondem. No bloco de Atributos (*@attribute*), são registradas as variáveis analisadas cujos valores irão compor os padrões. No bloco de Dados (*@data*), são registrados os padrões.

O registro compreendendo do primeiro ao penúltimo elemento em cada padrão representa as quantidades de ativações que determinado neurônio realizou dentro de um intervalo, e o último elemento é utilizado para sinalizar se nesse determinado intervalo, um específico animal esteve ou não em contato com determinado objeto.

Na Figura 6 são sumarizados os tempos de contato, em segundos, que cada animal esteve com cada objeto.

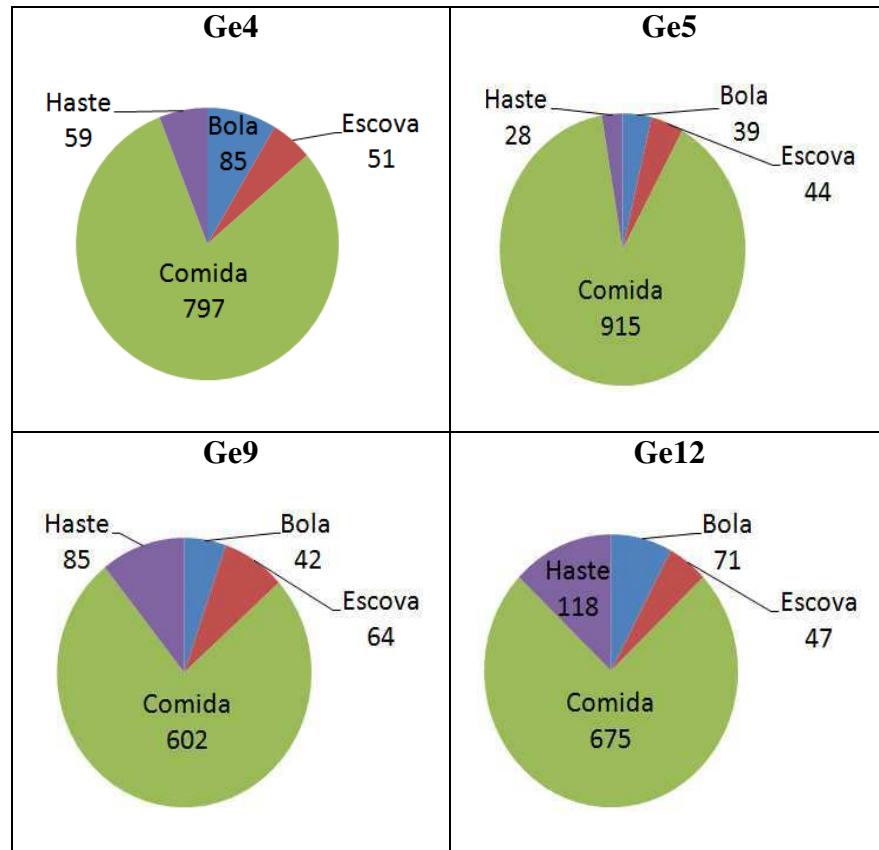


Figura 6: Tempo de contato de cada animal com cada objeto.

Pode-se observar que o objeto “comida” foi aquele com o qual todos os animais tiveram maior tempo de contato.

Na Figura 7 são apresentadas as quantidades de neurônios monitorados em cada animal, em cada região.

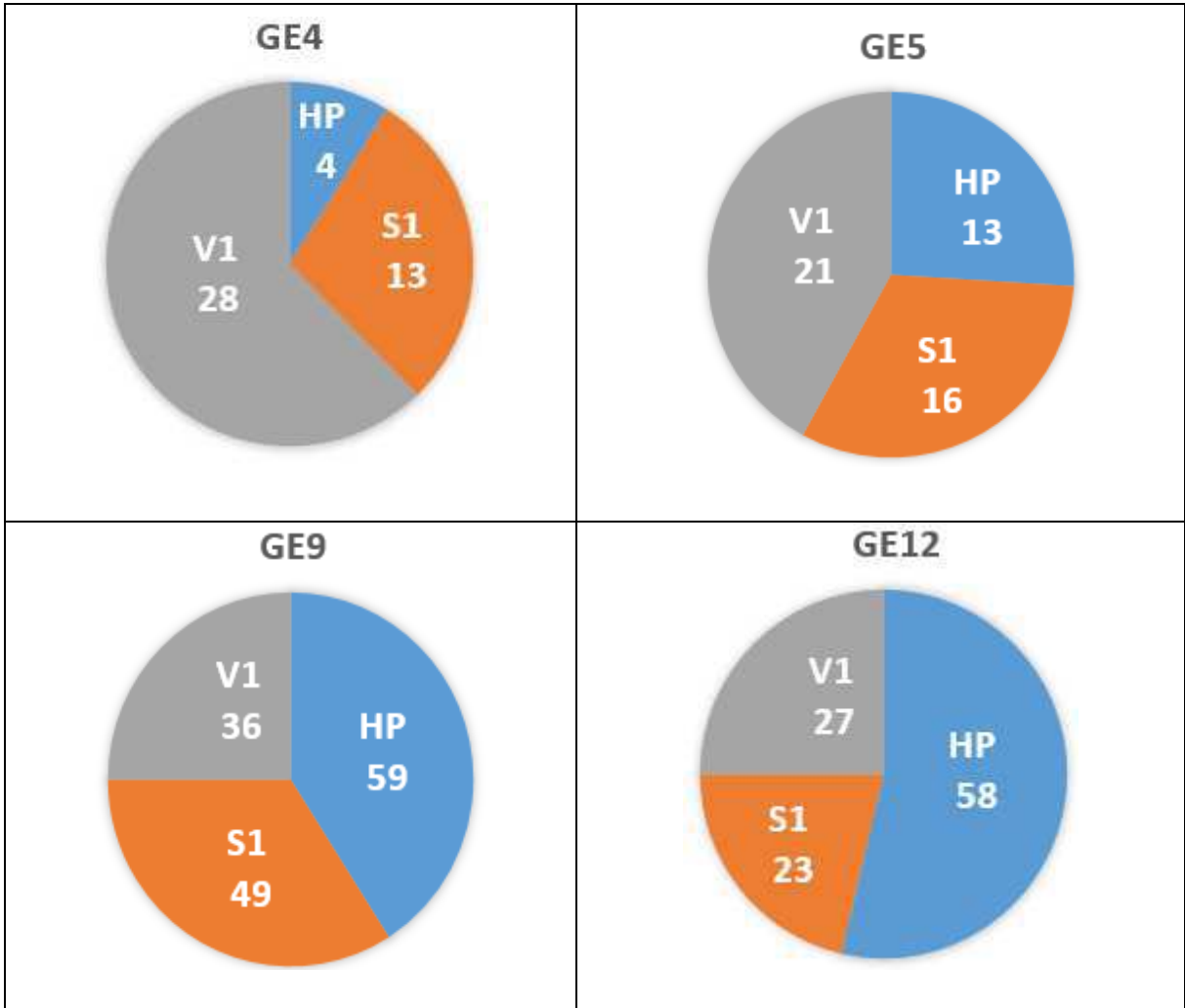


Figura 7: Número de neurônios monitorados de cada animal em cada região.

É possível observar que essas quantidades diferem em cada animal. Em alguns animais, as quantidades de neurônios em algumas áreas são muito diferentes (Ex: HP em Ge4 e em Ge12), o que pode influenciar as relações de similaridade entre as classes de padrões de ativação neuronal quando são analisados os padrões daquelas áreas. Essa diferença na quantidade de neurônios monitorados é justificada por depender de um implante cirúrgico e da anatomia específica de cada animal, o que torna difícil garantir que as mesmas quantidades de neurônios poderão ser monitoradas em todos os eletrodos.

Tais observações podem auxiliar no entendimento de algumas inferências realizadas e casos particulares encontrados na execução das abordagens descritas nas Seções 3.4 e 3.5.



### 3.4 *Abordagem Não Supervisionada*

Uma das abordagens desenvolvidas na pesquisa deste trabalho de dissertação foi a não supervisionada. A adoção deste tipo de abordagem foi motivada pela questão de pesquisa apresentada na próxima seção. Nesta seção, também são discutidos os as questões abordadas, fatores e a variável de resposta, níveis dos fatores e projeto do experimento, hipóteses, e os testes de hipótese.

#### 3.4.1 Questões

Foram elaboradas duas questões com o intuito de guiarem a realização das atividades dessa abordagem, e cujas respostas dessem suporte a inferências quanto à qualidade do processo de aquisição dos dados, que podem ser conferidas no capítulo 5 (Considerações Finais e Trabalhos Futuros).

As questões analisadas na realização dessa abordagem foram:

**Q1:** É possível, automaticamente construir subgrupos, a partir das classes de padrões de ativação neuronal?

**Q1.1:** Existem diferenças nas composições de cada subgrupo das classes de padrões de ativação neuronal?

Para responder essas questões, foram analisados oito fatores e uma variável de resposta, descritos na Seção 3.3.2, apresentada a seguir, e testadas duas hipóteses, descritas na Seção 3.3.3.

### 3.4.2 Fatores e Variável de Resposta

Para investigar as hipóteses da pesquisa (descritas na Seção 3.3.4), foram determinados os seguintes fatores e variável de resposta:

Fatores:

1. Animal (Ge4, Ge5, Ge9, Ge12);
2. Área do cérebro (HP, V1, S1 e Assembleia);
3. Objeto (Bola, Escova, Comida e Haste);
4. Métrica de definição da quantidade de agrupamentos (Cotovelo, Silhueta);
5. Tamanho do Bin (100ms, 150ms, 200ms, 250ms, 300ms, 350ms, 400ms);
6. Tamanho da Janela (5 amostras, 10 amostras, 15 amostras, 20 amostras, 25 amostras, 30 amostras, 35 amostras, 40 amostras); e
7. Intervalo de Contato (ic, i1, i2, i3).

Variável de resposta:

1. Subgrupos de padrões de ativação neuronais de cada classe, de acordo com os valores dos fatores.

O primeiro fator está relacionado com o animal cujos padrões de ativação neuronal são analisados em cada experiência (conforme discutido no Capítulo 2). Os animais são identificados com o acrônimo "Ge" seguido de um número para o animal. Os animais que irão ser utilizados são o Ge4, Ge5, Ge9, Ge12.

O segundo fator refere-se às áreas do cérebro onde os padrões de ativação neuronal são analisados em cada experiência. As áreas do cérebro que serão analisadas são o hipocampo, córtex visual e córtex somatossensorial, identificados como HP, V1 e S1, respectivamente, e a região formada pela maior assembleia de neurônios detectada durante a exploração dos

objetos pelo animal. De acordo com Hebb (1949), as assembleias são subconjuntos de neurônios com comportamento de co-ativação significativa. O algoritmo utilizado nesta pesquisa para detecção das assembleias teve como base a análise de componentes principais (PCA), conforme descrito no trabalho de Lopes-dos-Santos, Ribeiro e Tort (2013).

O terceiro fator refere-se ao objeto que o animal estava em contato no momento em que foram extraídos os padrões de ativação neuronal. Os objetos que os animais tiveram contato foram os seguintes: Bola, Escova, Comida e Haste.

O quarto fator está relacionado à métrica de definição da quantidade de agrupamentos. Nesse trabalho de dissertação foram empregadas as métricas Cotovelo e Silhueta.

O quinto fator é referente ao tamanho em milissegundos do *bin*, o qual se refere ao intervalo de tempo em que a oscilação da atividade elétrica de um neurônio será processada para a contagem dos potenciais de ação. Os tamanhos dos *bins* adotados foram 50ms, 100ms, 150ms, 200ms, 250ms, 300ms, 350ms e 400ms.

O sexto fator está relacionado ao tamanho da janela, esse tamanho indica quantos *bins* irão compor um padrão. Os tamanhos da janela empregados foram 5, 10, 15, 20, 25, 30, 35, 40 amostras.

O sétimo fator é referente ao intervalo de contato em que cada animal interagiu com determinado objeto. Os níveis manipulados foram o intervalo completo (*ic*), a primeira, segunda e terceira parte da divisão do intervalo completo (*i1*, *i2* e *i3*, respectivamente), em subpartes de mesma duração, para analisar a similaridade entre as classes de padrões de ativação durante todo o experimento e em diferentes intervalos, respectivamente.

É importante ressaltar que todos os fatores foram combinados livremente por serem independentes.

### 3.4.3 Níveis dos Fatores e Projeto do Experimento

Os níveis de um fator são os possíveis valores que esse pode assumir. De acordo com a seção anterior, verifica-se que existem quatro níveis para o primeiro fator, quatro níveis para o segundo fator, quatro níveis para o terceiro fator, dois níveis para o quarto fator, oito níveis para o quinto fator, oito níveis para o sexto fator e quatro níveis para o sétimo fator,

totalizando 32768 tratamentos.

Cada tratamento foi representado por uma distribuição estatística. As distribuições são compostas por 50 amostras. Cada amostra contém um valor de uma determinada característica analisada de agrupamentos gerados. Cada agrupamento foi gerado sobre 56 padrões, escolhidos aleatoriamente, de uma determinada classe de objeto no mesmo formato usado no trabalho de Vasconcelos (2011).

Nessa pesquisa foi decidido realizar uma análise fatorial completa, ou seja, foi decidido realizar a análise dos níveis de similaridade entre classes de padrões de ativação neuronal de acordo com todos os tratamentos anteriormente citados, com o intuito de consolidar as inferências a serem realizadas.

#### 3.4.4 Hipóteses

Nesta seção são apresentadas as hipóteses formuladas com base nas questões apresentadas na seção anterior, juntamente com estruturas de representação matemática da hipótese.

**H1-0:** Não é possível, de maneira ótima, agrupar em subgrupos as classes de padrões de ativação neuronais.

$$\exists X \mid X \subset S \rightarrow X = S \quad (14)$$

Em que  $X$  é um grupo de padrões e  $S$  são todos os padrões de uma determinada classe.

**H1.1-0:** Todos os subgrupos distintos de padrões de ativação neuronal possuem a mesma quantidade de elementos.

$$\exists X \& Z \mid X \subset S \text{ AND } Z \subset S \text{ AND } X \neq Z \rightarrow |X| = |Z| \quad (15)$$

Conforme pode ser visualizado, as hipóteses são encadeadas, isto é, apenas se a hipótese H1-0 for confirmada, será realizado o teste da hipótese H1.1-0.

Como descrito na Seção 2.3, para a realização dos testes das hipóteses anteriormente apresentadas, faz-se necessária a manipulação de fatores e de variáveis de respostas, os quais são descritos na Seção 3.3.2.

### 3.4.5 Testes de Hipóteses

Os testes das hipóteses H.1 e H.1.1 foram realizados em duas etapas, a fim de encontrar a quantidade ótima de subgrupos que podem ser formados de padrões de ativação neuronais referentes aos tratamentos descritos na Seção 3.3.3, (primeira etapa), e contabilizar a quantidade de elementos em cada um desses subgrupos (segunda etapa).

Os esquemas das atividades realizadas nas duas etapas são ilustrados nas Figuras 8 e 9.

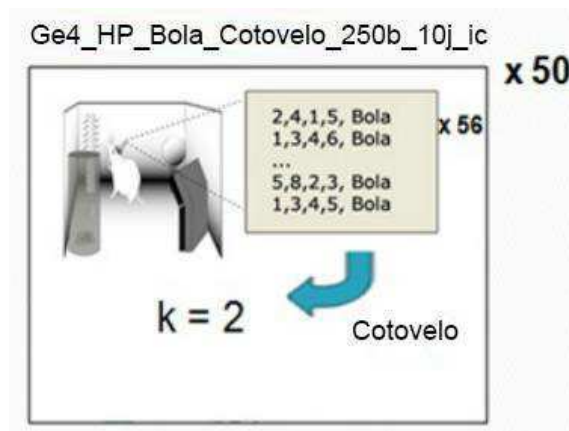


Figura 8: Esquema das atividades realizadas na 1ª etapa.

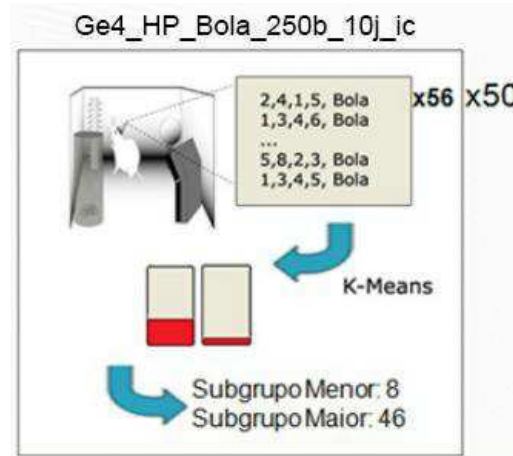


Figura 9: Esquema das atividades realizadas na 2ª etapa.

Na primeira etapa, foram construídos 50 conjuntos de 56 padrões, selecionados ao acaso, para cada tratamento. Em cada grupo, foram aplicados os índices Cotovelo (KETCHEN, SHOOK, 1996) e Silhueta (ROUSSEEUW, 1987), que retornam a maior quantidade de subgrupos que cada grupo pode possuir, e após isto, foi calculada a média da quantidade de subgrupos para cada método.

A construção de 50 grupos é justificada pela tentativa de aumentar a representatividade dos padrões selecionados para a determinação do valor da característica analisada (número de elementos em cada subgrupo), uma vez que a soma de padrões disponíveis é vasta. A quantidade de 56 padrões em cada conjunto se dá por ser a maior quantidade de padrões disponíveis de acordo com todas configurações de *bin* e janela, fornecidos pelo sistema Neurobjects, descrito na Seção 3.2.1.

Na segunda etapa, cada conjunto foi reagrupado de acordo com o algoritmo K-Means (JAIN, DUBES, 1988), utilizando como parâmetro referente à quantidade desejada de agrupamentos a quantidade média de subgrupos encontrada na etapa anterior, e após isto, foi calculada a média de elementos encontrados em cada subgrupo resultante do agrupamento de acordo com o algoritmo K-Means.

### 3.5 *Abordagem Semi-Supervisionada*

Outra abordagem desenvolvida nesse trabalho de dissertação foi a semi-supervisionada, motivada pela questão de pesquisa apresentada na próxima seção. Nesta seção, também são discutidos os fatores e a variável de resposta, níveis dos fatores e projeto do experimento, hipóteses, e os testes de hipótese.

#### 3.5.1 Questões

Foi elaborada uma questão com intuito de guiar a realização das atividades dessa abordagem. A obtenção da resposta dessa questão também foi utilizada para sinalizar o alcance do objetivo da aplicação dessa abordagem.

A questão analisada na realização dessa abordagem foi:

**Q2:** Quais são os níveis de similaridade entre as classes de objetos de padrões de ativação neuronal (Bola, Escova, Comida, Haste) de um determinado animal (Ge4, Ge5, Ge9, Ge12) em certas áreas do cérebro (hipocampo, somatossensorial primário, ou córtex visual primário - HP, S1 e V1, respectivamente) de acordo com uma determinada métrica para avaliar a similaridade (Jaccard, Dunn ou Silhueta) (JAIN, DUBES, 1988) e o algoritmo de agrupamento (K-Means, EM, agrupamento hierárquico ou agrupamento com base na densidade) (JAIN, DUBES, 1988), o tamanho do *bin* e da janela?

Com a resposta dessa questão, foram realizadas inferências quanto ao nível de similaridade entre classes de padrões de ativação neuronal, que podem ser conferidas no Capítulo 4 (Apresentação e Análise dos Resultados).

Para responder essas questões, foram analisados oito fatores, descritos na Seção 3.4.2, e testadas oito hipóteses, descritas na Seção 3.4.3.

### 3.5.2 Fatores e Variável de Resposta

A fim de investigar as hipóteses descritas na Seção 3.5.4, foi necessário determinar um conjunto de fatores e variável de resposta apresentados a seguir:

Fatores:

1. Animal (Ge4, Ge5, Ge9, Ge12);
2. Área do cérebro (HP, V1, S1 e Assembleia);
3. Pares de classes de objetos de padrões de ativação neuronal (Bola e Escova, Bola e Comida, Bola e Haste, Escova e Comida, Escova e Haste, e Comida e Haste);
4. Algoritmo de agrupamento (K-Means, EM, Hier. e Dens.);
5. Métrica de similaridade (Jaccard, Dunn, Silhueta);
6. Tamanho do Bin (50ms, 100ms, 150ms, 200ms, 250ms, 300ms, 350ms, 400ms);
7. Tamanho da Janela (5 amostras, 10 amostras, 15 amostras, 20 amostras, 25 amostras, 30 amostras, 35 amostras, 40 amostras); e
8. Intervalo de Contato ( $i_c$ ,  $i_1$ ,  $i_2$ ,  $i_3$ ).

Variável de resposta:

1. Nível de semelhança entre os padrões de ativação neuronal para objetos diferentes em contato com o animal.

O primeiro fator está relacionado com o animal cujos padrões de ativação neuronal são analisados em cada experiência. Os animais são identificados com o acrônimo "Ge" concatenados com um número para o animal. Os animais utilizados foram o Ge4, Ge5, Ge9, Ge12.

O segundo fator refere-se às áreas do cérebro onde os padrões de ativação neuronal são analisados em cada experiência. As áreas do cérebro que serão analisadas são o hipocampo, o



córtex visual e o córtex somatossensorial, identificados como HP, V1 e S1, respectivamente, e a região formada pela maior assembleia de neurônios detectada durante a exploração dos objetos pelo animal. O algoritmo utilizado nesta pesquisa para detecção das assembleias teve como base a análise dos principais componentes (PCA), conforme descrito no trabalho de Lopes-dos-Santos, Ribeiro e Tort (2013).

O terceiro fator refere-se aos pares de classes de objetos de padrões de ativação neuronal entre os quais serão computados nível de semelhança entre eles. Os pares são combinados, dois a dois, e envolvem as seguintes classes de objetos: Bola, Escova, Comida e Haste.

O quarto fator está relacionado ao algoritmo de agrupamento de padrões de ativação neuronal. Os algoritmos utilizados foram K-Means, EM, agrupamento hierárquico e agrupamento baseado em densidade, identificados como K-Means, EM, Hier. e Dens., respectivamente.

O quinto fator está relacionado à métrica de comparação de semelhança entre os grupos formados. Nesse trabalho de dissertação foram empregadas Jaccard, Silhueta e Dunn.

O sexto fator é referente ao tamanho em milissegundos do *bin*, o qual se refere ao intervalo de tempo em que a oscilação da atividade elétrica de um neurônio será processada para a contagem dos potenciais de ação. Os tamanhos dos *bins* adotados foram 50ms, 100ms, 150ms, 200ms, 250ms, 300ms, 350ms e 400ms.

O sétimo fator está relacionado ao tamanho da janela, esse tamanho indica quantos *bins* irão compor um padrão. Os tamanhos da janela empregados nessa dissertação foram 5, 10, 15, 20, 25, 30, 35 e 40 amostras.

O oitavo fator é referente ao intervalo de contato em que cada animal interagiu com determinado objeto. Os níveis manipulados foram o intervalo completo (ic), a primeira, segunda e terceira parte da divisão do intervalo completo (i1, i2 e i3, respectivamente), em subpartes de mesma duração, para analisar a similaridade entre as classes de padrões de ativação durante todo o experimento e em diferentes intervalos, respectivamente.

É importante ressaltar que todos os fatores foram combinados livremente por serem independentes.

### 3.5.3 Níveis dos Fatores e Projeto do Experimento

De acordo com a seção anterior, temos quatro níveis para o primeiro fator, quatro níveis para o segundo fator, seis níveis para o terceiro fator, quatro níveis para o quarto fator e três níveis para o quinto fator, oito níveis para o sexto fator e oito níveis para o sétimo fator, quatro níveis para o oitavo fator, totalizando 294912 tratamentos.

Cada tratamento foi representado por uma distribuição estatística. As distribuições são compostas por 50 amostras. Cada amostra contém o valor de similaridade mensurado a partir de agrupamentos gerados. Cada agrupamento foi gerado sobre 56 padrões, escolhidos aleatoriamente, de cada classe analisada, mesmo formato usado no trabalho de Vasconcelos (2011).

Nessa pesquisa foi definida a realização de uma análise fatorial completa, e as distribuições foram analisadas em agrupamentos de acordo com as hipóteses descritas na Seção 3.5.4.

### 3.5.4 Hipóteses

Foram assumidos que nas estruturas de representação (denominadas neste trabalho como tratamentos), presentes nas formulações de hipóteses seguintes, a letra 'u' indica a distribuição para um dado tratamento e o sinal '=' indica que o tratamento é estatisticamente equivalente a outro tratamento com nível de significância de 0,05. As letras “b” e “j” sinalizam o valor atribuídos para os fatores *bin* e janela, respectivamente, conforme descritos na Seção 3.4.2.

**H1-0:** Independente dos animais, os níveis de similaridade são os mesmos, por exemplo,

$u_{Ge4\_HP\_ [Bola, Escova] \_K-means\_Jaccard\_250b\_10j\_ic} =$

$u_{Ge5\_HP\_ [Bola, Escova] \_K-means\_Jaccard\_250b\_10j\_ic} =$

$u_{Ge9\_HP\_ [Bola, Escova] \_K-means\_Jaccard\_250b\_10j\_ic} =$

$u_{Ge12\_HP\_ [Bola, Escova] \_K-means\_Jaccard\_250b\_10j\_ic} =$

**H2-0:** Os níveis de similaridade entre as duas classes de objetos de padrões de ativação neuronal são independentes da região do cérebro, por exemplo,

$u_{Ge4\_HP\_ [Bola, Escova] \_K-means\_Jaccard\_250b\_10j\_ic} =$

$u_{Ge4\_S1\_ [Bola, Escova] \_K-means\_Jaccard\_250b\_10j\_ic} =$

$u_{Ge4\_V1\_ [Bola, Escova] \_K-means\_Jaccard\_250b\_10j\_ic} =$

$u_{Ge4\_Assembleia\_ [Bola, Escova] \_K-means\_Jaccard\_250b\_10j\_ic}$

**H3-0:** Independentemente dos pares de objeto, os níveis de similaridade são os mesmos, por exemplo,

$u_{Ge4\_HP\_ [Bola, Escova] \_K-means\_Jaccard\_250b\_10j\_ic} =$

$u_{Ge4\_HP\_ [Bola, Comida] \_K-means\_Jaccard\_250b\_10j\_ic} =$

$u_{Ge4\_HP\_ [Bola, Haste] \_K-means\_Jaccard\_250b\_10j\_ic} =$

$u_{Ge4\_HP\_ [Escova, Comida] \_K-means\_Jaccard\_250b\_10j\_ic} =$

$u_{Ge4\_HP\_ [Escova, Haste] \_K-means\_Jaccard\_250b\_10j\_ic} =$

$u_{Ge4\_HP\_ [Comida, Haste] \_K-means\_Jaccard\_250b\_10j\_ic}$

**H4-0:** Os níveis de similaridade mantêm-se independentemente dos diferentes algoritmos de agrupamento, por exemplo,

$u_{Ge4\_HP\_ [Bola, Escova] \_K-means\_Jaccard\_250b\_10j\_ic} =$

$u_{Ge4\_HP\_ [Bola, Escova] \_EM\_Jaccard\_250b\_10j\_ic} =$

$u_{Ge4\_HP\_ [Bola, Escova] \_Hier\_Jaccard\_250b\_10j\_ic} =$

$u_{Ge4\_HP\_ [Bola, Escova] \_Dens\_Jaccard\_250b\_10j\_ic}$

**H5-0:** Independentemente da métrica de similaridade, os níveis de similaridade são os mesmos, por exemplo,

u\_Ge4\_HP\_[Bola, Escova]\_K-means\_Jaccard\_250b\_10j\_ic =  
 u\_Ge4\_HP\_[Bola, Escova]\_K-means\_Dunn\_250b\_10j\_ic =  
 u\_Ge4\_HP\_[Bola, Escova]\_K-means\_Silhouette\_250b\_10j\_ic

**H6-0:** Independentemente do tamanho do *bin*, os níveis de similaridade são os mesmos, por exemplo,

u\_Ge4\_HP\_[Bola, Escova]\_K-means\_Jaccard\_100b\_10j\_ic =  
 u\_Ge4\_HP\_[Bola, Escova]\_K-means\_Jaccard\_150b\_10j\_ic =  
 u\_Ge4\_HP\_[Bola, Escova]\_K-means\_Jaccard\_200b\_10j\_ic =  
 u\_Ge4\_HP\_[Bola, Escova]\_K-means\_Jaccard\_250b\_10j\_ic =  
 u\_Ge4\_HP\_[Bola, Escova]\_K-means\_Jaccard\_300b\_10j\_ic =  
 u\_Ge4\_HP\_[Bola, Escova]\_K-means\_Jaccard\_350b\_10j\_ic =  
 u\_Ge4\_HP\_[Bola, Escova]\_K-means\_Jaccard\_400b\_10j\_ic

**H7-0:** Independentemente do tamanho da janela, os níveis de similaridade são os mesmos, por exemplo,

u\_Ge4\_HP\_[Bola, Escova]\_K-means\_Jaccard\_250b\_5j\_ic =  
 u\_Ge4\_HP\_[Bola, Escova]\_K-means\_Jaccard\_250b\_10j\_ic =  
 u\_Ge4\_HP\_[Bola, Escova]\_K-means\_Jaccard\_250b\_15j\_ic =  
 u\_Ge4\_HP\_[Bola, Escova]\_K-means\_Jaccard\_250b\_20j\_ic =  
 u\_Ge4\_HP\_[Bola, Escova]\_K-means\_Jaccard\_250b\_25j\_ic =  
 u\_Ge4\_HP\_[Bola, Escova]\_K-means\_Jaccard\_250b\_30j\_ic =  
 u\_Ge4\_HP\_[Bola, Escova]\_K-means\_Jaccard\_250b\_35j\_ic =  
 u\_Ge4\_HP\_[Bola, Escova]\_K-means\_Jaccard\_250b\_40j\_ic

H8-0: Independentemente do intervalo de contato, os níveis de similaridade são os mesmos, por exemplo,

$u_{\text{Ge4\_HP\_ [Bola, Escova] \_K-means\_Jaccard\_250b\_10j\_ic}} =$

$u_{\text{Ge4\_HP\_ [Bola, Escova] \_K-means\_Jaccard\_250b\_10j\_i1}} =$

$u_{\text{Ge4\_HP\_ [Bola, Escova] \_K-means\_Jaccard\_250b\_10j\_i2}} =$

$u_{\text{Ge4\_HP\_ [Bola, Escova] \_K-means\_Jaccard\_250b\_10j\_i3}}$

### 3.5.5 Testes de Hipóteses

Para o teste das hipóteses anteriormente listadas, foram obtidas distribuições de similaridade, cujo processo de construção é descrito a seguir e ilustrado na Figura 10.

Sejam A e B duas classes de padrões de ativação neuronal.

Passos:

1. São selecionados ao acaso 56 padrões de cada classe;
2. Os padrões são agrupados de acordo com um algoritmo de agrupamento;
3. Afere-se a similaridade entre o agrupamento que possui maior quantidade de padrões da classe A e o agrupamento que possui maior quantidade de padrões da classe B; e
4. Os passos anteriores são repetidos até ser formada uma distribuição composta por 50 valores de similaridade. Essa distribuição representará a similaridade entre as classes A e B.

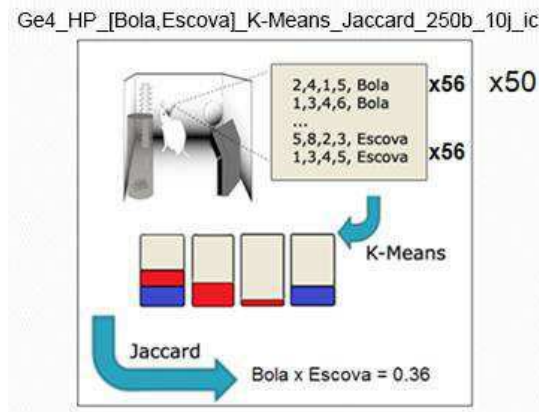


Figura 10: Esquema do processo de construção das distribuições de similaridade

As distribuições não são emparelhadas por causa das diferenças no que diz respeito à natureza e ao número de neurônios entre as regiões do cérebro, o que por sua vez são diferentes entre os animais.

Para cada distribuição, foram feitos testes de normalidade, análise de variância e um teste para diferenças entre tendência central.

A escolha da técnica de análise de agrupamento é justificada pelos tipos e estruturas de relações entre as variáveis a serem estudadas, interdependentes e causa/efeito, respectivamente (HAIR, 2009).

### 3.6 Teste de Robustez das Inferências

Os testes foram realizados sobre a robustez das inferências por meio da utilização das técnicas *Neuron Dropping* (MATSUMOTO et al., 2010), cuja estratégia ajuda a medir o impacto sobre a média de qualidade quando você tem diferentes quantidades de neurônios que formam o conjunto de dados, e *Dados Surrogados* (NICOLELIS, LEBEDEV, 2009) cuja estratégia é semelhante à técnica *Neuro Dropping*, com diferença de que em vez de remover as linhas, essas linhas são substituídas por outras linhas relacionadas, geradas aleatoriamente de acordo com uma distribuição de Poisson, substituindo desta forma a atividade específica de determinado neurônio por informação genérica, mas com mesma distribuição estatística.

Os testes foram aplicados sobre 50 classes de padrões de ativação neuronal, selecionadas ao acaso.

### *3.7 Considerações sobre o Capítulo*

Nesse capítulo, foram apresentados como é composta a estratégia proposta, como as ferramentas Neurobjects, MySQL, Weka e R foram empregadas na realização desse trabalho de pesquisa, as descrições do processo de geração da base de dados e a maneira como esses dados foram registrados, das abordagens realizadas, e dos testes de robustez das inferências.

No Capítulo 4 são apresentados os resultados das atividades realizadas conforme descrito nesse capítulo, e a análise desses resultados.

# Capítulo 4

## Apresentação e Análise dos Resultados

Neste capítulo são apresentados os resultados dos testes descritos no capítulo anterior, bem como uma análise e as possíveis inferências formuladas para justificar tais observações. Essas informações estão divididas nas Seções 4.1, 4.2 e 4.3, relacionadas ao teste de robustez das inferências em relação à qualidade dos dados e às atividades de Análise dos Agrupamentos e de Análise dos Níveis de Similaridade, respectivamente.

### *4.1 Qualidade dos Dados*

Os resultados dos testes de robustez das inferências em relação à qualidade dos dados mostraram que inferências tem um nível aceitável de robustez em relação à qualidade dos dados.

Nas Figura 11 e 12 são ilustrados dois testes dentre o conjunto de testes de robustez realizados, os quais utilizaram as técnicas *Neuro Dropping* e Dados Surrogados, respectivamente, conforme descritos na Seção 3.5.



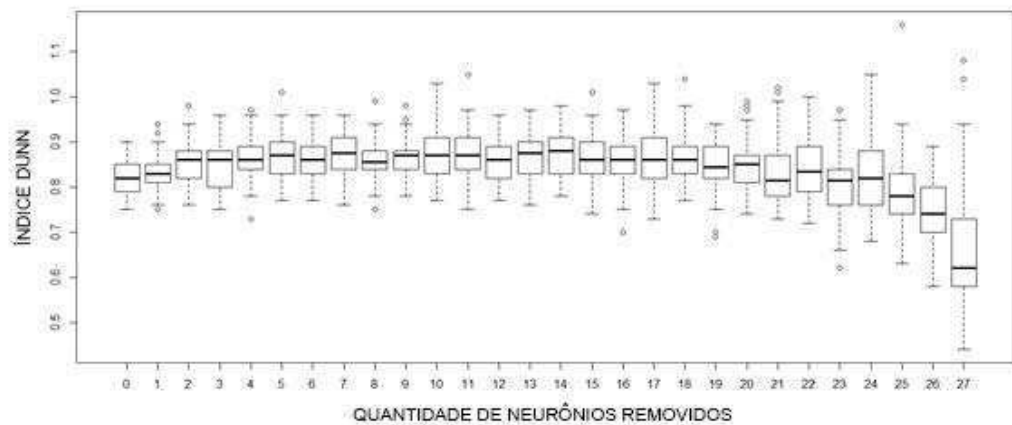


Figura 11: *Boxplots* da distribuição  $u_{\text{Ge4\_V1}}[\text{Bola, Escova}]_{\text{EM\_Dunn\_250b\_10j\_ic}}$  com a remoção gradativa da quantidade de neurônios utilizados para compor os padrões de ativação neuronal.

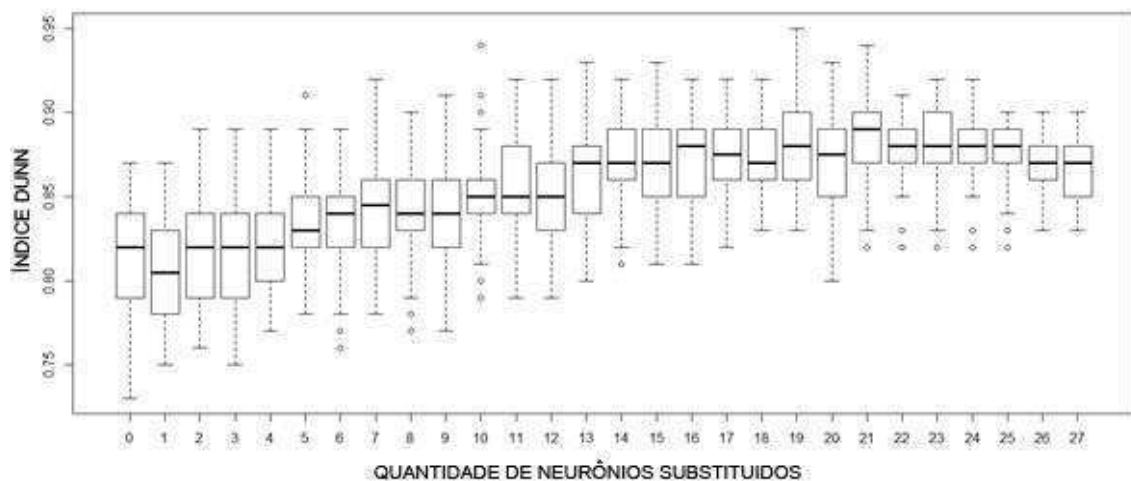


Figura 12: *Boxplots* da distribuição  $u_{\text{Ge4\_V1}}[\text{Bola, Escova}]_{\text{EM\_Dunn\_250b\_10j\_ic}}$  com a substituição gradativa da atividade de neurônios por um valor aleatório de uma distribuição de Poisson, utilizados para compor os padrões de ativação neuronal.

Conforme pode ser visualizado na Figura 11, à medida que são removidos os neurônios utilizados para compor os padrões de ativação neuronal, os valores do index Dunn vão se tornando menores, o que implica que as classes vão se tornando cada vez mais similares.

A Figura 12 indica que a similaridade entre as classes é sensível à atividade de uma certa quantidade de neurônios, esta quantidade julgada adequada.

## 4.2 *Abordagem Não Supervisionada*

Após a definição das classes de padrões de ativação neuronal, de acordo com os fatores descritos na Seção 3.3.2, e a realização da primeira etapa do teste das hipóteses, descrita na Seção 3.3.5, foi encontrado que em média, conforme os índices Cotovelo e Silhueta, a quantidade ótima de subgrupos que cada grupo possuía era dois. Com esse resultado, é rejeitada a hipótese H1-0.

Para a segunda etapa do teste das hipóteses, empregou-se o valor dois para determinar o número de agrupamentos esperados após a aplicação do algoritmo K-Means.

Após isto, foi feita a média de elementos encontrados em cada subgrupo resultante do agrupamento de acordo com o algoritmo K-Means.

Foi observado que em todos os grupos, em média, o maior subgrupo possuía 48 padrões e o menor 6 padrões, com desvio padrão de 5 e 5 padrões respectivamente. Com esse resultado, é rejeitada a hipótese H1.1-0.

Também foi analisado o erro quadrático médio em todos os grupos. Em média, o menor subgrupo apresentou erro quadrático médio 40% maior que o maior grupo.

Com base nesses resultados, foi inferido que as atuais classes de padrões de ativação neuronal podem ser divididas em dois subgrupos, sendo um deles mais representativo e conciso, e outro menos representativo e diversificado, esse último, possivelmente resultante de ruídos no registro dos dados.

## 4.3 *Abordagem Semi-Supervisionada*

Uma vez realizada a análise do *QQ-plot* e do teste de Shapiro Wilk (SHAPIRO, 1965) em cada distribuição, foi observado que a maioria das distribuições não se comporta normalmente.

No teste de Bartlett (BARTLETT, 1937), todas as distribuições apresentaram diferenças significativas ( $\alpha = 0,05$ ) entre suas variâncias.

Com base nestes dois primeiros resultados, utilizou-se o teste não paramétrico de Kruskal-Wallis, o qual demonstrou que há diferença significativa ( $p \leq 0,05$ ) entre as tendências centrais de cada grupo de distribuições em cada teste de hipótese, estas listadas na Seção 3.4.3.

As Figuras 13 a 20 contêm representações gráfica (*Boxplots*) de alguns grupos de distribuições estudados, referindo-se, respectivamente, para o teste de hipóteses H1-0 à H8-0.

Os valores no eixo x são definidos de acordo com as distribuições, referentes a cada nível do fator analisado no teste de uma hipótese. Os valores no eixo y são relacionados aos níveis de similaridade entre as classes de padrões de ativação neuronal de acordo com uma métrica de similaridade para determinado tratamento.

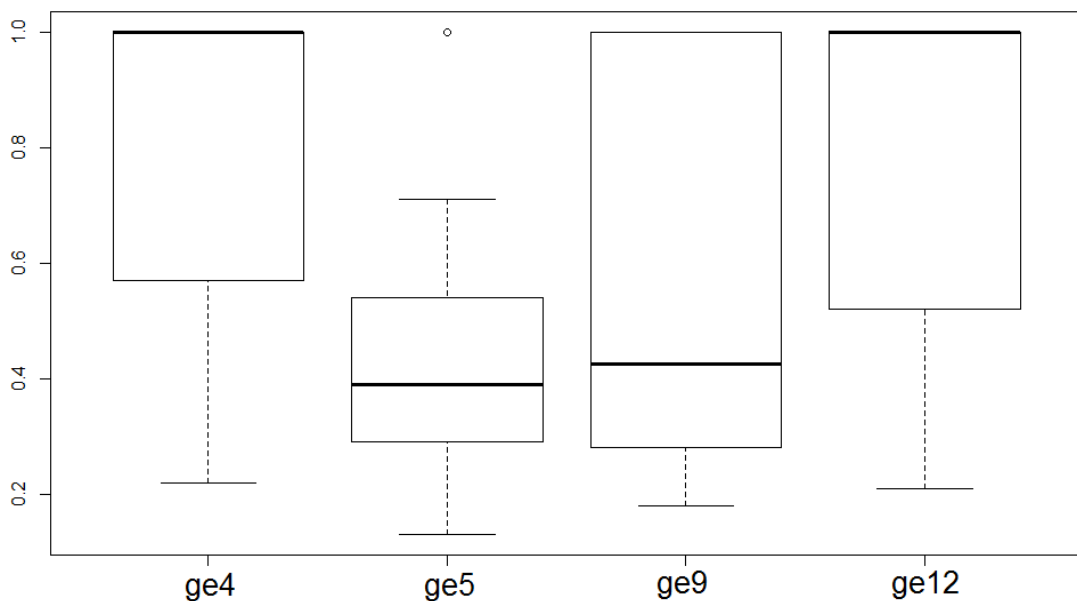


Figura 13: *Boxplots* das distribuições  $u_{(Ge4, Ge5, Ge9, Ge12)}_{HP_{[Bola, Escova]}_K}$ -  
Means\_Jaccard\_250b\_10j\_ic

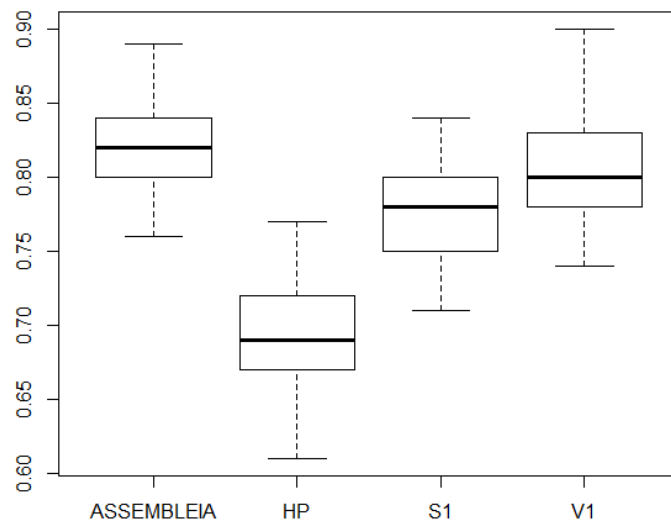


Figura 14: *Boxplots* das distribuições  $u_{Ge4}_{(Assembleia, HP, S1, V1)}_{[Bola, Escova]}_{K-Means\_Dunn\_250b\_10j\_ic}$

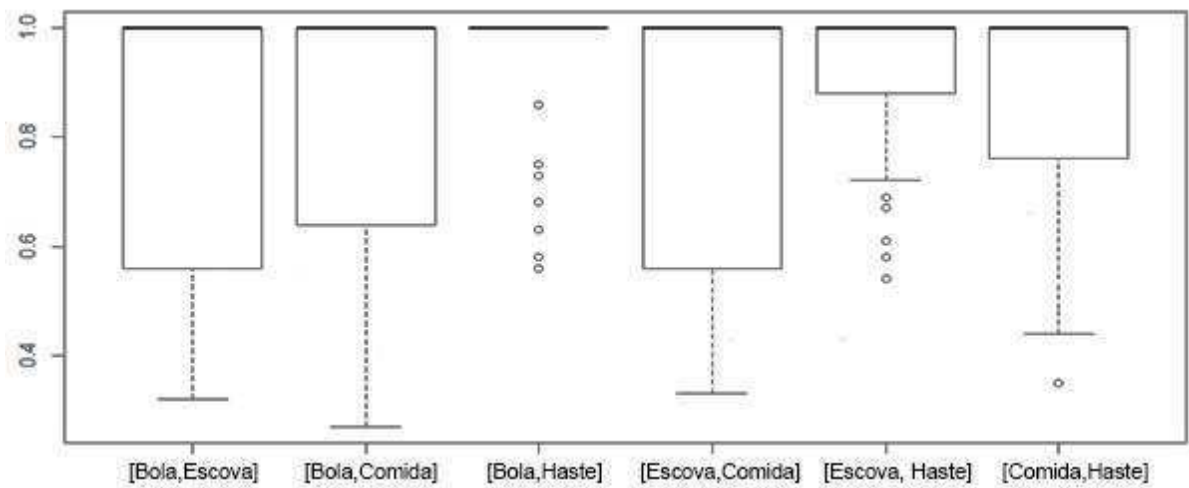


Figura 15: *Boxplots* das distribuições  $u_{Ge4\_HP}_{([Bola, Escova], [Bola, Comida], [Bola, Haste], [Escova, Comida], [Escova, Haste], [Comida, Haste])}_{K-Means\_Jaccard\_250b\_10j\_ic}$

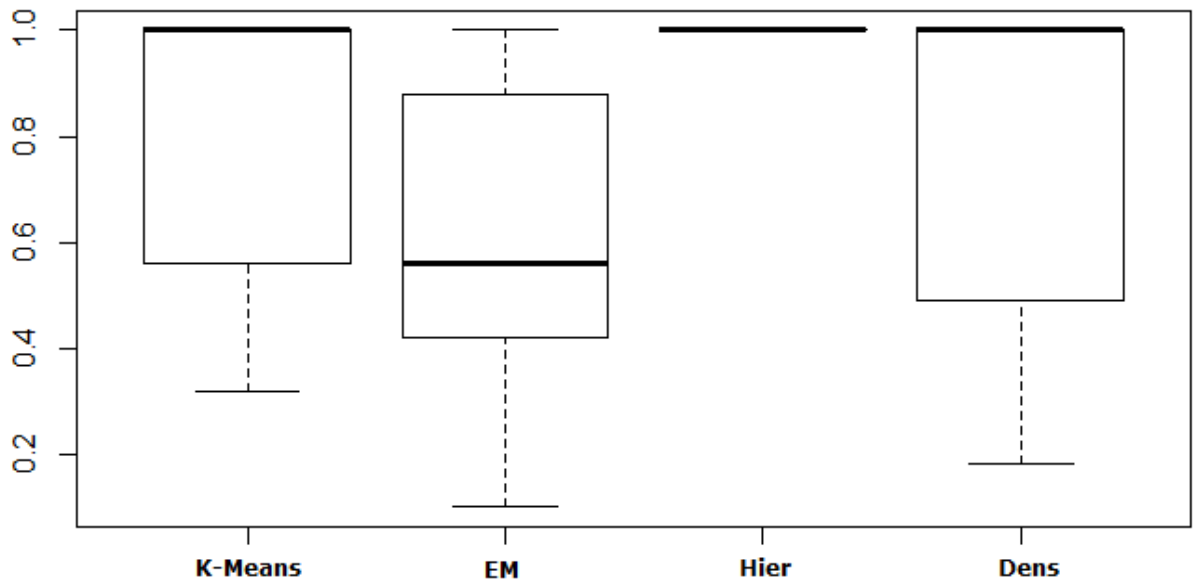


Figura 16: *Boxplots* das distribuições  $u_{\text{Ge4\_HP}}[\text{Bola, Escova}]_{(\text{K-Means, EM, Hier, Dens})\_Jaccard\_250b\_10j\_ic}$

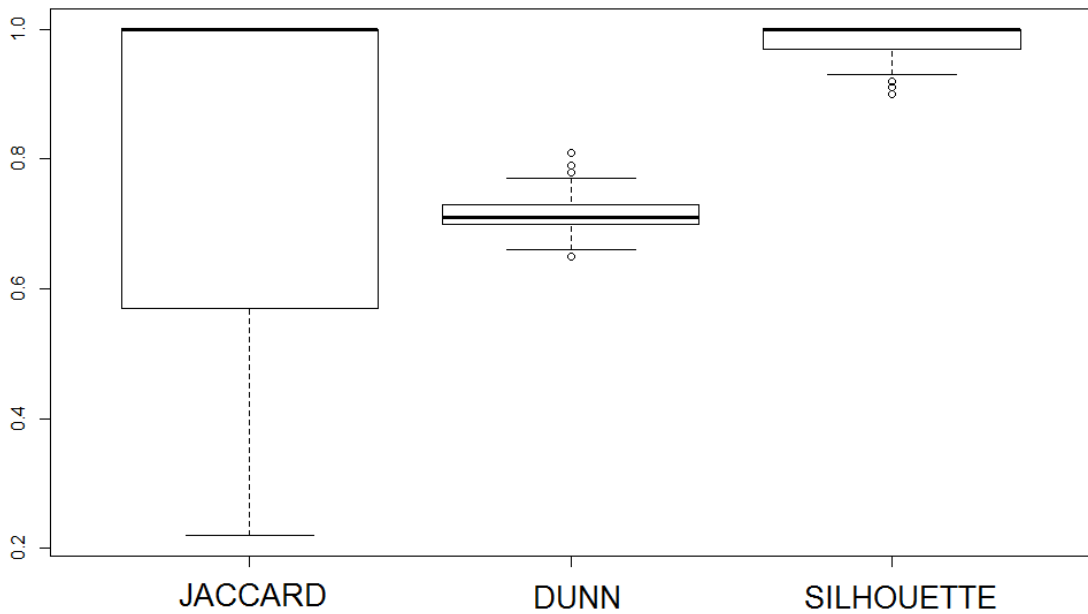


Figura 17: *Boxplots* das distribuições  $u_{\text{Ge4\_HP}}[\text{Bola, Escova}]_{\text{K-Means}}(\text{Jaccard, Dunn, Silhouette})_{250b\_10j\_ic}$

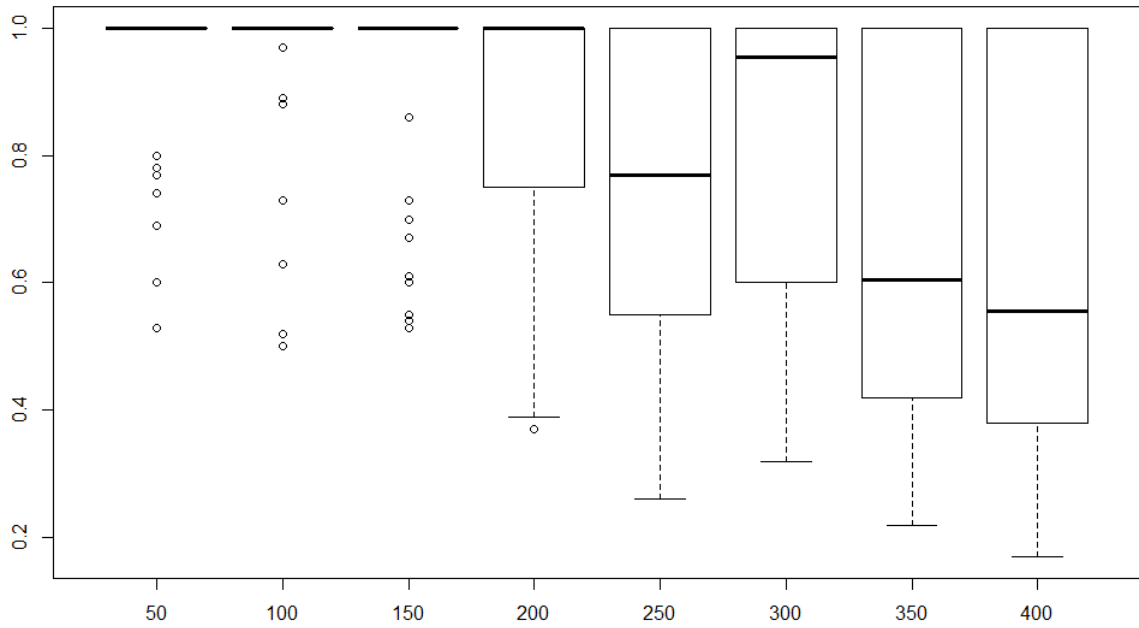


Figura 18: *Boxplots* das distribuições  $u_{\text{Ge4\_HP}}[\text{Bola, Escova}]_{\text{K-Means\_Jaccard}}(50-400)b_{10j\_ic}$

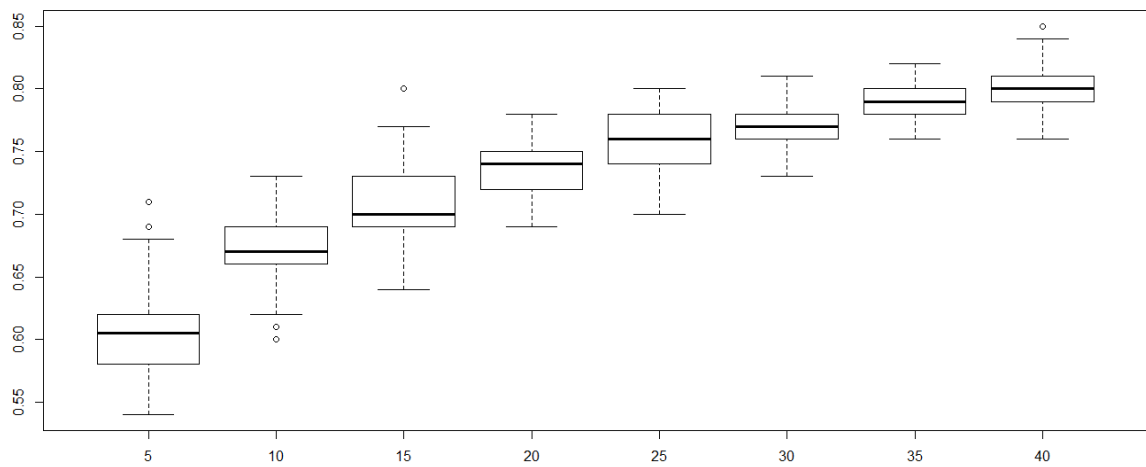


Figura 19: *Boxplots* das distribuições  $u_{\text{Ge4\_HP}}[\text{Bola, Escova}]_{\text{K-Means\_Dunn}}100b_{(5-40)j\_ic}$

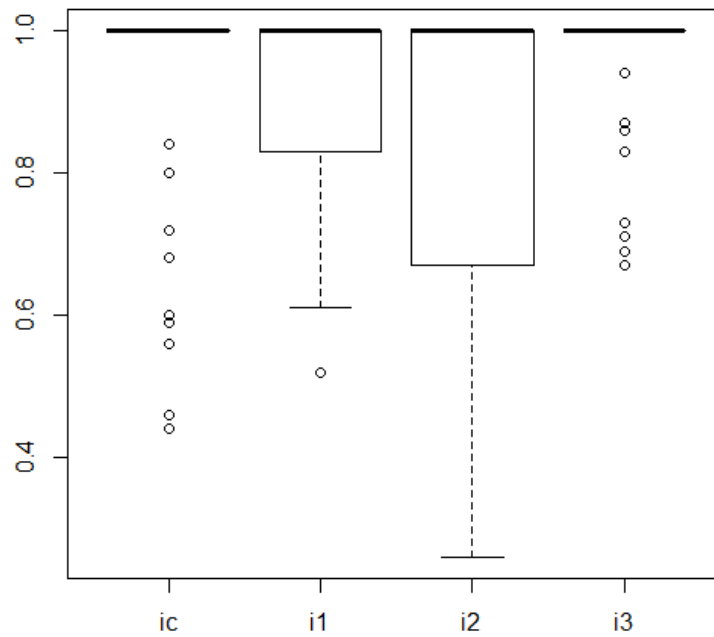


Figura 20: *Boxplots* das distribuições  $u\_Ge4\_HP\_ [Bola, Escova] \_K\text{-Means\_Jaccard\_}100b\_10j\_i(1,2,3)$

Algumas das demais imagens utilizadas para teste das hipóteses podem ser conferidas no Apêndice A.

Também foi realizada uma análise quanto à disposição espacial das medianas dos níveis de similaridade das classes de padrões de ativação neuronal, em relação aos Animais, às Regiões do Cérebro e ao Número de Neurônios monitorados para a composição dos padrões de ativação, conforme pode ser visualizado na Figura 21.

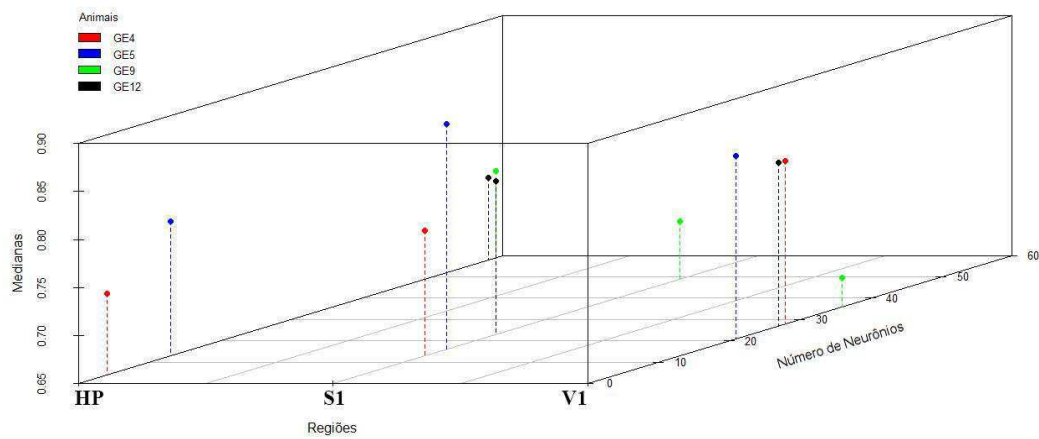


Figura 21: Disposição espacial das medianas dos níveis de similaridade das classes de padrões de ativação neuronal, em relação aos Animais, às Regiões do Cérebro e ao Número de Neurônios monitorados para a composição dos padrões de ativação.

Nas seguintes tabelas é apresentada a diferença observada entre as distribuições com foco em determinado fator, e a razão de similaridade entre as distribuições com foco em determinado fator.

A diferença observada foi calculada por meio da função `kruskalmc()` da ferramenta R, conforme descrita na seção 3.2.4, aplicada a todos os grupos de distribuições, agrupados de acordo com um determinado fator, com o valor arredondado em 2 casas decimais, com o uso da função `round()` da ferramenta R.

A razão de similaridade foi calculada na divisão da quantidade de vezes em que as distribuições eram julgadas significativamente semelhantes ( $p \leq 0,05$ ) no teste de Kruskal-Wallis, pela quantidade total de avaliações realizadas entre as duas distribuições.

Com o intuito de promover melhor visualização da informação, esse valor foi multiplicado por 255 e arredondado para ser usado como valor em uma escala de cinza, na coloração das células dessa tabela.



Tabela 3 - Diferença observada entre as distribuições com foco no fator Animal.

	Ge4	Ge5	Ge9	Ge12
Ge4	0			
Ge5	64,82 ± 45,48	0		
Ge9	25,6 ± 26,46	49,77 ± 33,85	0	
Ge12	37,04 ± 38,59	38,23 ± 30,23	18,3 ± 23,94	0

Tabela 4 - Razão de similaridade entre as distribuições com foco no fator Animal.

	Ge4	Ge5	Ge9	Ge12
Ge4	255			
Ge5	184	255		
Ge9	92	184	255	
Ge12	92	135	60	255

Tabela 5 - Diferença observada entre as distribuições com foco no fator Região do Cérebro.

	HP	S1	V1	Assembleia
HP	0			
S1	29,17 ± 25,53	0		
V1	44,29 ± 51,69	30,83 ± 27,44	0	
Assembleia	48,73 ± 55,06	30,80 ± 27,43	10,34 ± 11,65	0

Tabela 6 - Razão de similaridade entre as distribuições com foco no fator Região do Cérebro.

	HP	S1	V1	Assembleia
HP	255			
S1	117	255		
V1	94	117	255	
Assembleia	100	102	17	255

Tabela 7 - Diferença observada entre as distribuições com foco no fator Par de Objetos.

	Bola, Escova	Bola, Comida	Bola, Haste	Escova, Comida	Escova, Haste	Comida, Haste
Bola, Escova	0					
Bola, Comida	25,43 ± 28,81	0				
Bola, Haste	18,40 ± 21,43	23,80 ± 27,71	0			
Escova, Comida	18,93 ± 23,35	22,87 ± 25,57	21,59 ± 25,04	0		
Escova, Haste	20,04 ±24,14	35,49 ± 41,15	24,56± 29,46	22,46 ± 26,95	0	
Comida, Haste	24,74 ± 30,10	25,99 ± 29,88	23,74 ± 27,70	16,87 ± 26,95	23,96 ± 28,90	0

Tabela 8 - Razão de similaridade entre as distribuições com foco no fator Par de Objetos.

	Bola, Escova	Bola, Comida	Bola, Haste	Escova, Comida	Escova, Haste	Comida, Haste
Bola, Escova	0					
Bola, Comida	51	0				
Bola, Haste	22	38	0			
Escova, Comida	26	35	27	0		
Escova, Haste	28	69	41	39	0	
Comida, Haste	45	44	27	20	45	0

Tabela 9 - Diferença observada entre as distribuições com foco no fator Algoritmo de Agrupamento.

	K-Means	EM	Hier	Dens
K-Means	0			
EM	52,3 ± 34,28	0		
Hier	52,51 ± 34,61	52,78 ± 34,93	0	
Dens	50,34 ± 40,19	52,75 ± 36,12	51,88 ± 32,43	0

Tabela 10 - Razão de similaridade entre as distribuições com foco no fator Algoritmo de Agrupamento.

	K-Means	EM	Hier	Dens
K-Means	255			
EM	173	255		
Hier	167	166	255	
Dens	153	168	173	255

Tabela 11 - Diferença observada entre as distribuições com foco no fator Métrica

	Jaccard	Dunn	Silhueta
Jaccard	0		
Dunn	12,34 $\pm$ 16,42	0	
Silhueta	58,37 $\pm$ 19,54	69,61 $\pm$ 6,35	0

Tabela 12 - Razão de similaridade entre as distribuições com foco no fator Métrica.

	Jaccard	Dunn	Silhueta
Jaccard	255		
Dunn	55	255	
Silhueta	236	255	255

Tabela 13 - Diferença observada entre as distribuições com foco no fator *Bins*.

	50	100	150	200	250	300	350	400
50	0							
100	34,57 ±34,29	0						
150	49,02±45, 91	18,57±15, 98	0					
200	70,68±59, 66	39,40±32, 43	23,20±19, 51	0				
250	73,09 ±67,82	45,45±42, 31	33,34±29, 02	22,48±18, 75	0			
300	78,60±66, 33	51,05±43, 39	34,90±33, 33	24,71±25, 47	20,33±17, 35	0		
350	96,80±74, 15	68,03±51, 02	52,09±38, 95	33,04±28, 83	27,67±23, 72	26,46±24, 50	0	
400	93,59 ±74,47	67,99±56, 91	52,45±47, 52	36,20±37, 96	35,14±28, 07	31,04±26, 71	17,57 ±19,16	0

Tabela 14 - Razão de similaridade entre as distribuições com foco no fator *Bins*.

	50	100	150	200	250	300	350	400
50	255							
100	57	255						
150	81	0	255					
200	92	57	0	255				
250	113	74	32	3	255			
300	120	78	46	14	0	255		

350	145	103	74	28	7	11	255	
400	152	110	78	46	28	18	3	255

Tabela 15 - Diferença observada entre as distribuições com foco no fator Janela.

	5	10	15	20	25	30	35	40
5	0							
10	27,85±25,20	0						
15	49,74±49,94	26,01±25,41	0					
20	69,07±72,45	44,97±46,77	24,36±22,36	0				
25	83,87±88,36	58,51±64,34	38,16±38,22	18,94±18,14	0			
30	97,47±109,48	73,47±84,21	54,02±57,30	33,08±37,09	22,14±19,63	0		
35	106,93±119,62	81,99±95,41	61,86±69,37	42,06±48,20	28,30±30,67	15,23±13,92	0	
40	115,78 ±131,09	91,24±106,60	70,81±80,46	51,78±58,72	38,54±40,75	23,34 ±23,32	15,36±12,51	0

Tabela 16 - Razão de similaridade entre as distribuições com foco no fator Janela.

	5	10	15	20	25	30	35	40
5	255							
10	9	255						
15	85	14	255					
20	90	85	3	255				
25	90	87	69	0	255			
30	96	85	87	55	0	255		
35	96	87	87	73	39	0	255	
40	96	89	90	85	74	7	0	255

Tabela 17 - Diferença observada entre as distribuições com foco no fator Intervalo de Contato.

	Ic	i1	i2	i3
ic	0			
i1	8,72±10,54	0		
i2	10,56±10,99	15,31±18,62	0	
i3	7,58±7,23	10,76±13,02	10,86±12,16	0

Tabela 18 - Razão de similaridade entre as distribuições com foco no fator Intervalo de Contato.

	Ic	i1	i2	i3
ic	255			
i1	12	255		
i2	16	43	255	
i3	2	23	25	255

Todas as hipóteses nulas foram rejeitadas, e os resultados dos testes de hipóteses H1-0 à H8-0, serão apresentados pela discussão das Figura 13 até a Figura 20 e das Tabela 3 até a Tabela 18.

De acordo com a Figura 13 e as Tabelas 3 e 4, pode ser verificado que as relações de similaridade entre pares de objetos são processadas de forma diferente para cada animal. Assim, rejeitando a hipótese H1-0.

Também foi observado que o nível de similaridade entre as classes de padrões de ativação neuronal é sensível às regiões em que os padrões foram adquiridos, como pode ser visto na Figura 14 e nas Tabelas 5 e 6. Com isso, a hipótese H2-0 foi rejeitada.

Essa variação do nível de similaridade pode ser justificada também pelas diferentes quantidades de neurônios monitorados em cada região, conforme apresentado na Seção 3.3 e na Figura 21, em particular, o animal Ge4 teve 4, 13 e 28 neurônios monitorados nas regiões HP, S1 e V1, respectivamente.

Do ponto de vista da Computação, quanto menor o número de neurônios monitorados, menor a quantidade de fatores analisados para a construção dos padrões, o que torna a classificação desses padrões mais sensível às variações dos valores desses fatores (POLIKAR, 2006).

Os níveis menos acentuados de similaridade na assembleia indicam que a diferenciação dos objetos é feita de forma menos detalhada do que em grupos de neurônios de regiões mais específicas e especializadas, como HP, S1 e V1.

Contudo, um comportamento que pode ser observado entre as classes de padrões de ativação neuronal é a igualdade significativa quando utilizados o algoritmo de agrupamento EM, a métrica de similaridade Dunn, o tamanho de bin 250, o tamanho da janela 10 e intervalo de contato completo, em geral, das médias dos níveis de similaridade dos animais em cada par de objetos, entre as regiões HP, S1 e V1, conforme pode ser observado nas Figuras 22 a 27.

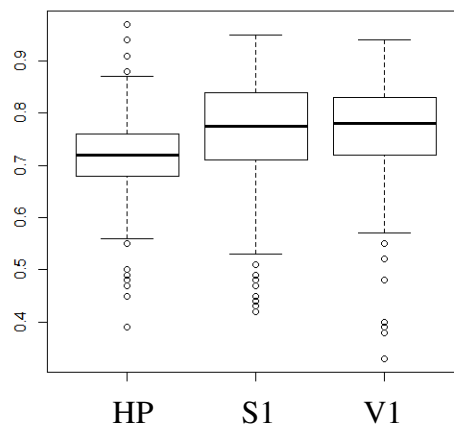


Figura 22: *Boxplots* das distribuições  $u\_MediaAnimal\_(HP,S1,V1)\_[Bola, Escova]\_EM\_Dunn\_250b\_10j\_ic$



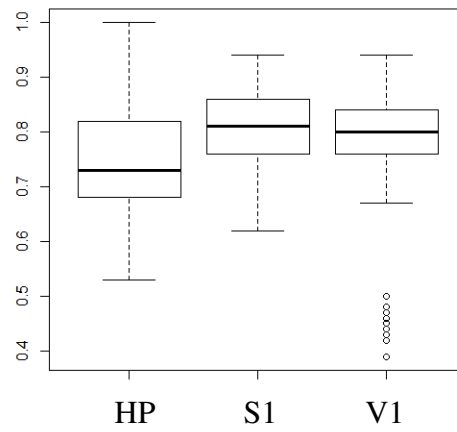


Figura 23: *Boxplots* das distribuições  $u\_MediaAnimal\_ (HP,S1,V1)\_ [Bola, Comida]\_EM\_Dunn\_250b\_10j\_ic$

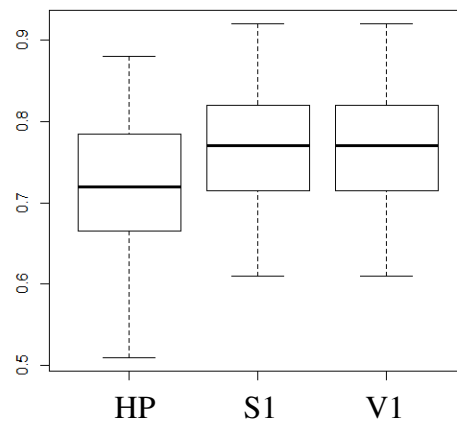


Figura 24: *Boxplots* das distribuições  $u\_MediaAnimal\_ (HP,S1,V1)\_ [Bola, Haste]\_EM\_Dunn\_250b\_10j\_ic$

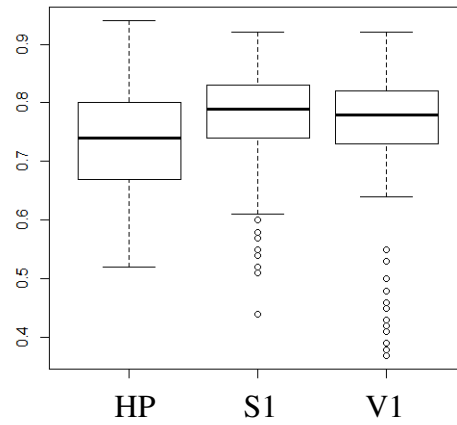


Figura 25: *Boxplots* das distribuições  $u\_MediaAnimal\_ (HP,S1,V1)\_ [Escova, Comida]\_ EM\_Dunn\_250b\_10j\_ic$

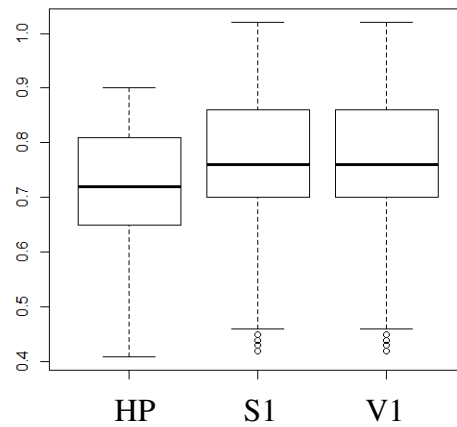


Figura 26: *Boxplots* das distribuições  $u\_MediaAnimal\_ (HP,S1,V1)\_ [Escova, Haste]\_ EM\_Dunn\_250b\_10j\_ic$

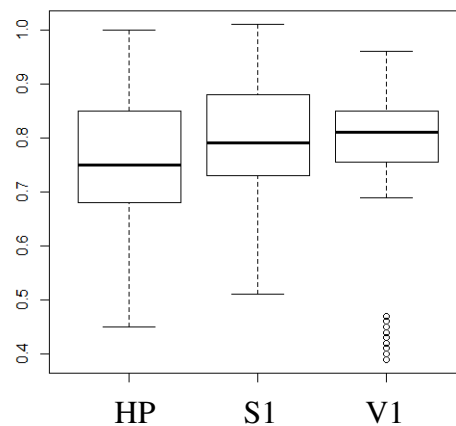


Figura 27: *Boxplots* das distribuições  $u\_MediaAnimal\_(HP,S1,V1)\_[Comida, Haste]\_EM\_Dunn\_250b\_10j\_ic$

Esses resultados estão em concordância com o trabalho de Vasconcelos *et al.* (2011), no qual, trabalhando sob os mesmos dados e utilizando uma abordagem supervisionada, foi observado que a qualidade média de classificação dos dados em cada objeto para todos os animais, mensurada pela AUROC, é similar entre as regiões S1 e V1.

Dessa maneira, as duas abordagens colaboram para inferir que dentro de um determinado contexto controlado, regiões do cérebro realizam a atividade de reconhecimento de objetos de maneira semelhante.

Ainda em relação ao trabalho de Vasconcelos *et al.* (2011), foi realizado o teste de correlação linear de Pearson (BOX, HUNTER, HUNTER, 2005) em relação à média do índice de Dunn e a AUROC para as mesmas classes utilizadas no referido trabalho: animais Ge4, Ge5, Ge9 e Ge12, regiões do cérebro HP, S1 e V1, e objetos Bola, Escova, Haste, utilizando o algoritmo de agrupamento EM, a métrica de similaridade Dunn, o tamanho de bin 250, o tamanho da janela 10 e intervalo de contato completo.

Como o valor da AUROC, no trabalho de Vasconcelos *et al.* (2011), foi relacionado ao nível de discriminação entre um objeto específico e os demais objetos, nesta dissertação foi utilizada a média entre os níveis de similaridade entre os pares de objetos das classes relacionadas que possuíam tal objeto específico (excluindo-se os pares envolvendo o objeto Comida).

Na Figura 28 é apresentada a disposição espacial de cada valor da média índice Dunn pela AUROC das distribuições analisadas.

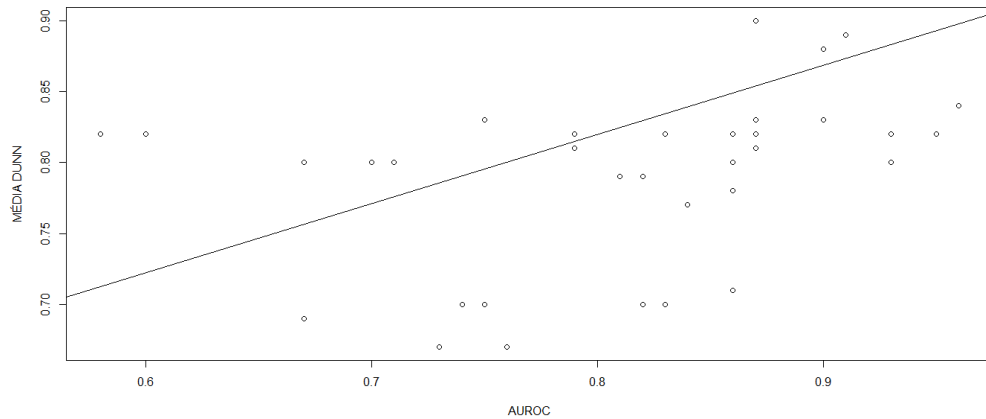


Figura 28: Disposição espacial das médias do índice Dunn pelo valor AUROC das distribuições  $u_{(Ge4,Ge5,Ge9,Ge12)}(HP,V1,S1)(Bola,Escova,Haste)_{EM\_Dunn\_250b\_10j\_ic}$

Para o nível de significância igual a 0,05, o coeficiente de correlação de Pearson entre essas variáveis é 0,32 e o p-valor igual a 0,057, o que indica baixa relação.

Outra observação que pode ser fitada é que a mediana dos valores de similaridades das distribuições referentes à região HP é sempre inferior às demais medianas dos valores de similaridades das distribuições referentes às regiões S1 e V1. Essa observação dá suporte a inferir que, do ponto de vista da Neurociência, na atividade de armazenamento de informações na região HP, responsável pela memória recente, são realizadas mais abstrações do que no processamento das informações táteis e visuais, nas regiões S1 e V1, respectivamente.

Com referência à Figura 15 e às Tabelas 7 e 8, a maioria dos valores de semelhança entre cada par de objetos é maior do que 1, no entanto, a variação entre estes valores não é a mesma. Com isso, a hipótese H3-0 foi rejeitada.

O que também podemos verificar, de acordo com a Figura 16 e as Tabelas 9 e 10, é que, entre os algoritmos de agrupamento, EM foi o único com mediana distinta (menor que 1). Com isso, a hipótese H4-0 foi rejeitada.

Quanto às métricas de similaridade, de acordo com a Figura 17 e as Tabelas 11 e 12, é possível ver que Jaccard apresentou a maior variância e existe diferença entre as distribuições. Assim, foi rejeitada a hipótese H5-0.

De acordo com a Figura 18 e as Tabelas 13 e 14, é possível observar que a variabilidade de níveis de semelhança é diretamente proporcional ao tamanho do *bin* e que a diferença entre as distribuições aumenta com o aumento do tamanho do *bin*. Com base nesses resultados, a hipótese H6-0 foi rejeitada.

De acordo com a Figura 19 e as Tabelas 15 e 16, também é possível verificar que a variabilidade de níveis de semelhança é diretamente proporcional ao tamanho da janela e que a diferença entre as distribuições aumenta com o aumento do tamanho da janela. Mas este comportamento não pode ser visto nos tratamentos que utilizam a métrica de Jaccard. Assim, é rejeitada a hipótese H7-0.

Finalmente, de acordo com a Figura 20 e as Tabelas 17 e 18, nota-se que o nível de similaridade é sensível ao tempo de contato, e as distribuições são diferentes entre si. Com isso, a hipótese H8-0 é rejeitada.

Na Figura 21, as disposições espaciais das medianas evidenciam a alta influência do número de neurônios e as regiões do cérebro dos animais para a definição dos níveis de similaridade.

#### *4.4 Considerações sobre o Capítulo*

Nesse capítulo, foram apresentadas as análises dos resultados relacionados ao teste da qualidade dos dados, o que demonstrou que a estratégia proposta para análise do nível de similaridade é sensível à informação contida nos dados, e das abordagens aplicadas nesse trabalho, os quais apresentaram indício quanto à aquisição dos dados, o processo de assimilação de objetos durante o tempo, entre outros.

No Capítulo 5 são apresentadas as considerações finais quanto a todos os resultados e atividades realizadas durante essa pesquisa, e sugestões de trabalhos futuros.

# Capítulo 5

## Considerações Finais e Sugestões para Trabalhos Futuros

Neste capítulo é apresentado um sumário da pesquisa, seguido de conclusões e de propostas para trabalhos futuros, nas subseções a seguir.

### *5.1 Sumário da pesquisa*

A presente pesquisa teve como objetivo colaborar para a atividade de avaliar a similaridade entre classes de padrões de ativação neuronal.

Os padrões neuronais avaliados foram extraídos, com o uso de microeletrodos, das regiões cerebrais, do hipocampo, do córtex somatossensorial primário e do córtex visual primário, de ratos da linhagem Long Evans, criados em cativeiro, durante a exploração de um ambiente no escuro povoado com objetos (bola, escova, comida e haste) com os quais os animais só tiveram contato no momento do experimento. Esses padrões foram classificados em classes de acordo com determinados fatores.

Esses fatores (descritos nas Seções 3.4.2 e 3.5.2) foram definidos de acordo com as características do experimento que poderiam influenciar à variável de resposta analisada na investigação de hipóteses (descritos nas Seções 3.4.4 e 3.5.4), que por sua vez foram definidas com o intuito de na sua realização serem respondidas questões (descritas nas Seções 3.4.1 e 3.5.1) de acordo com o objetivo da pesquisa.

Para o teste dessas hipóteses, foram aplicadas as aprendizagens Semi-Supervisionada, uma vez que se buscou reduzir o impacto de imprecisões presentes nos dados, e Não Supervisionada para analisar a características inerentes aos dados, como o número de agrupamentos.

Além dos resultados de cada abordagem, os quais serão discutidos na Seção 5.2, outro resultado a ser ressaltado é a contribuição do trabalho no que diz respeito à elaboração de uma estratégia para análise de classes de padrões de ativação neuronal, atividade esta que, conforme apresentado anteriormente, tem recebido pouca atenção e que pode ser empregada por desenvolvedores de software e pesquisadores para uma melhor compreensão do funcionamento do cérebro.

## 5.2 *Conclusões*

Com base nos resultados encontrados nos estudos utilizando abordagem não supervisionada, infere-se que o processo de aquisição e definição dos padrões de ativação por um observador foi sujeito a ruídos causados por motivos não controláveis, entre os quais podem ser imprecisões de registro ou o fato de que o animal, no momento do contato com o objeto, já estava mentalizando outras atividades ou experiências além da que ele estava demonstrando realizar.

Ainda referente a essa abordagem, pela natureza dos resultados, esta se apresentou como uma possível ferramenta para validação de experimentos que realizem processos semelhantes de aquisição de dados e remoção de imprecisões.

Essa última característica foi utilizada como base para a definição dos subgrupos mais representativos de cada classe para o cálculo de similaridade entre eles, na outra abordagem dessa pesquisa.

Pela abordagem semi-supervisionada foi observado que nem todos os graus de similaridade entre pares de classes de objetos são iguais a um dado tratamento. Em alguns tratamentos, certos pares possuíam o mesmo grau de semelhança, mas não permanecem com este comportamento em todos os tratamentos, por exemplo, alguns algoritmos de agrupamento tinham níveis estatisticamente semelhantes de similaridade (K-Means e Dens), enquanto alguns eram bem diferentes (EM e HIER).

Desta forma, observou-se que o nível de similaridade é sensível a todos os fatores estudados (Animal, Região do Cérebro, Par de Objetos, Algoritmo de Agrupamento, Métrica de Similaridade, Tamanho do Bin e da Janela, e Intervalo de Contato).

Quanto às regiões do cérebro, todos os pares de classes de objetos possuem diferentes

níveis de similaridade, contudo, em um determinado conjunto de classes, as regiões do cérebro S1 e V1 realizam a atividade de reconhecimento de objetos de maneira semelhante, em concordância com o trabalho de Vasconcelos *et al.* (2011), que encontrou essa mesma semelhança entre as qualidades de classificação para um conjunto de classes análogas.

Entre os algoritmos de agrupamento, observa-se que os agrupamentos formados por EM foram melhor discriminados do que em relação aos demais algoritmos.

Pode-se inferir que dentre as métricas de similaridade, Jaccard é mais sensível às diferenças entre os padrões de ativação neuronal e que, com o aumento do tamanho da janela e do *bin*, as diferenças entre os padrões de ativação neuronal tendem a estabilizar em altos valores de similaridade. Isso reforça que as diferenças entre as semelhanças são mais claramente observadas no nível micro do que no macro nível de ativação neuronal.

A análise do comportamento do grau de similaridade entre os padrões de ativação neuronal ao longo do tempo sugere que o processo de assimilação dos padrões pode ser sensível ao nível de atenção de um determinado animal em um intervalo de tempo particular ou pelo nível de novidade de um objeto para um animal em particular.

A nova estratégia elaborada nesse trabalho de dissertação para análise de classes de padrões de ativação neuronal é passível de ser aplicada em outros contextos em que os dados possam ser diferenciados em classes, o que aumenta a contribuição do trabalho.

Esse trabalho também colabora com o estado da arte no que diz respeito a ressaltar a complexidade da codificação neuronal, uma vez que foi observado que o processamento de similaridade entre classes de padrões de ativação neuronal, na livre exploração tátil de objetos, é sensível a uma lista de fatores investigados nesse trabalho.

### 5.3 *Proposta para trabalhos futuros*

Propõe-se um estudo que considere o fluxo de atividades do animal, isto é, considere que o padrão analisado em um determinado momento é um produto de toda a experiência que o animal adquiriu até aquele instante, para analisar o impacto da experiência do animal para o nível de similaridade entre as classes de padrões de ativação neuronal.

Além desse estudo, o emprego de outros algoritmos de agrupamento (JAIN, DUBES, 1988), como método de Ward e OPTICS, regiões cerebrais (por exemplo, o córtex auditivo



primário e córtex motor primário) e outras classes de objetos podem ser analisadas para testar o nível de generalização das inferências feitas anteriormente.

Na abordagem não supervisionada, possíveis análises também podem ser feitas envolvendo outras características dos agrupamentos, tais como nível de coesão e variação entre os elementos de um grupo (JAIN, DUBES, 1988).

Pode também ser relevante estudar o impacto da escolha do nível de significância estatística (por exemplo,  $\alpha = 0,01$ ) para os resultados encontrados neste trabalho e realizar uma análise mais específica para a identificação de subgrupos de comportamento dentro dos tratamentos elaborados na pesquisa.

## Referências Bibliográficas

AKEMANN, W. *et al.* Imaging brain electric signals with genetically targeted voltage-sensitive fluorescent proteins. **Nature Methods**, v. 7, p. 643–649, 2010.

ARONOFF, R.; PETERSEN, C. Layer, column and cell-type specific genetic manipulation in mouse barrel cortex. **Frontiers in Neuroscience**, v. 2, p. 1, 2008.

AZAMI, H.; SANEI, S. Three novel spike detection approaches for noisy neuronal data. **In Computer and Knowledge Engineering**. 2nd International eConference on, p. 44-49, 2012.

BARHAM, P. *et al.* Xen and the art of virtualization. **ACM SIGOPS Operating Systems Review**, v. 37, n. 5, p. 164-177, 2003.

BARTLETT, M. S. **Properties of sufficiency and statistical tests**. Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences, v. 160, n. 901, p. 268-282, 1937.

BEAR, M. F.; CONNORS, B. W.; PARADISO, M. A. **Neurociências: desvendando o sistema nervoso**. ArtMed, 3 ed, 2008.

BERGLES, D. E. *et al.* Excitatory actions of norepinephrine on multiple classes of hippocampal CA1 interneurons. **Journal of Neuroscience**, v. 16, n. 2, p. 572-585, 1996.

BEST, P. J.; WHITE, A. M.; MINAI, A. Spatial processing in the brain: the activity of hippocampal place cells. *Annual Review Neurosciencie*. v. 24, p. 459–486, 2001.

BESTEL, R.; DAUS, A. W.; THIELEMANN, C. A novel automated spike sorting algorithm with adaptable feature extraction. **Journal of neuroscience methods**, v. 211, n. 1, p. 168-178, 2012.

BOW, S. T.; BOW, B. T. Pattern Recognition and Image Preprocessing. **Marcel Dekker**, 2002.

BOX, G. E. P.; HUNTER, J. S.; HUNTER, W. G. **Statistics for Experimenters**. Wiley, 2005.

BOYER, D. W.; JAMES, M. R. A distributed cortical representation underlies crossmodal object recognition in rats. **Journal of Neuroscience**, v. 30, n. 18, p. 6253–6261, 2010.

BREIDBACH, O. The origin and development of the neurosciences. **Theory and method in the neurosciences**. University of Pittsburgh Press, Pittsburgh, p. 7-29, 2001.

BREVEGLIERI, R. *et al.* Visual, somatosensory, and bimodal activities in the macaque parietal area PEc. **Cerebral Cortex**, v. 18, p. 806–816, 2008.

BROWN, E. N. *et al.* A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. **The Journal of Neuroscience**, v. 18, n. 18, p. 7411-7425, 1998.

BROWN, E. N.; KASS, R. E.; MITRA, P. P. Multiple neural spike train data analysis: state-of-the-art and future challenges. **Nature neuroscience**, v. 7, n. 5, p. 456-461, 2004.

BUTTS, C. T. Revisiting the Foundations of Network Analysis. **Science**, v. 325, n. 5939, p. 414–416, 2009.

BUZSAKI, G. Large-scale recording of neuronal ensembles. **Nature Neuroscience**, v. 7, n. 5, p. 446–51, 2004,

CARL, C. Petersen. The functional organization of the barrel cortex. **Neuron**, v. 56, n. 2, p. 339–355, 2007.

CATTANEO, A.; MAFFEI, L.; MORRONE, C. Patterns in the discharge of simple and complex visual cortical cells. **Proceedings of the Royal Society of London. Series B. Biological sciences**, v. 212, n. 1188, p. 279-297, 1981

COPE, M. *et al.* Non-invasive measurement of tissue oxygenation using near infrared (NIR) spectroscopy. **Optical Techniques and Biomedical Applications**. IEE Colloquium on. IET, p. 14/1-14/4, 1991.

DARMANJIAN, S.; PRINCIPE, J. Hierarchical clustering of neural data using Linked-Mixtures of Hidden Markov Models for Brain Machine Interfaces. **Acoustics, Speech and Signal Processing**, p. 3505-3508, 2009.

DEADWYLER, Sam A.; HAMPSON, Robert E. Ensemble Activity and Behavior—What's the Code? **Science**, v. 270, n. 5240, p. 1316-1316, 1995.

DEMPSTER, A. P. *et al.* Maximum likelihood from incomplete data via the EM algorithm. **Journal of the Royal statistical Society**, v. 39, n. 1, p. 1-38, 1977.

DOUGLAS, R. J.; MARTIN, K. A.C. Opening the grey box. **Trends in neurosciences**, v. 14, n. 7, p. 286-293, 1991.

DUNN, J. C. Well-separated clusters and optimal fuzzy partitions. **Journal of cybernetics**, v. 4, n. 1, p. 95-104, 1974.

EICHENBAUM, H.; DUDCHENKO, P.; WOOD, E.; SHAPIRO, M.; TANILA, H. The hippocampus, memory, and place cells: is it spatial memory or a memory space? **Neuron**, v. 23, n. 2, p. 209–226, 1999.

ESTER, Martin *et al.* A density-based algorithm for discovering clusters in large spatial databases with noise. **Evangelos Simoudis, Jiawei Han, Usama M. Fayyad. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining**, p. 226-231, 1996.

FINGER, S. *Origins of Neuroscience: A History of Explorations into Brain Function.* **Oxford University Press**, 2001.

FLANAGAN, J. R.; LEDERMAN, S. J. Neurobiology: Feeling bumps and holes. **Nature**, v. 412, n. 6845, p. 389–391, 2001.

FLETCHER, M. *et al.* Neural network based pattern matching and spike detection tools and services—in the CARMEN neuroinformatics project. **Neural Networks**, v. 21, n. 8, p. 1076-1084, 2008.

FOROUZAN, B.; MOSHARRAF, F. **Fundamentos da Ciência da Computação-Tradução da 2ª Edição Internacional.** Cengage Learning, 2012.

FOX, J.; ANDERSEN, R. Using the R statistical computing environment to teach social statistics courses. **Department of Sociology, McMaster University**, 2005.

GELMAN, A. *et al.* **Bayesian data analysis**. CRC press, 2013

GERSTNER, W.; KISTLER, W. M. Spiking Neuron Models. **Cambridge University Press**, 2002.

GRILLNER, Sten *et al.* Mechanisms for selection of basic motor programs—roles for the striatum and pallidum. **Trends in neurosciences**, v. 28, n. 7, p. 364-370, 2005.

GRILL-SPECTOR, K.; MALACH, R. The human visual cortex. **Annual Review of Neuroscience.**, v. 27, p. 649-677, 2004.

HAIR, J. F.; TATHAM, R. L.; ANDERSON, R. E. **Multivariate Data Analysis**. Bookman, 7 ed., 2009.

HAYKIN, S. **Redes Neurais, Princípios e prática**. Bookman, 2 ed., 1999.

HEBB, D. O. **The organization of behaviour: a neuropsychological theory**. Psychology Press, 1949.

HODGKIN, A. L.; HUXLEY, A. F. Action potentials recorded from inside a nerve fiber. **Nature**, v.144, p. 710-711, 1939.

HOFFMAN, Kari L. *et al.* Frontiers: Category-specific responses to faces and objects in primate auditory cortex. **Frontiers in systems neuroscience**, v. 2, 2007.

HUNG, C. P.; *et al.* Fast readout of object identity from macaque inferior temporal cortex. **Science**, v. 310, n. 5749, p. 863–866, 2005.

IOANNIDES, A.A.; TAYLOR, J.G. Testing models of attention with MEG. **In Neural Networks**. Proceedings of the International Joint Conference on, v. 1, p. 287– 291, 2003.

JAIN, A. K.; DUBES, R. C. Algorithms for Clustering Data, **Prentice Hall**, 1988.

JAIN, A. K.; MURTY, M. N.; FLYNN, P.J. Data Clustering: A Review. **ACM Computing Surveys**, v. 31, n. 3, p. 264–322, 1999.

JAMMALAMADAKA, A. *et al.* Statistical analysis of dendritic spine distributions in rat hippocampal cultures. **BMC bioinformatics**, v. 14, n. 1, p. 287, 2013

JOACHIMSTHALER, B. *et al.* Quantitative analysis of neuronal response properties in primary and higher-order auditory cortical fields of awake house mice (*Mus musculus*). **European Journal of Neuroscience**, v. 39, n. 6, p. 904-918, 2014.

JONES, E. G.; MENDELL, L. M. Assessing the decade of the brain. **Science**, v. 284, n. 5415, p. 739-739, 1999.

JOSHUA, M. *et al.* Quantifying the isolation quality of extracellularly recorded action potentials. **Journal of neuroscience methods**, v. 163, n. 2, p. 267-282, 2007.

KANDEL, E. R. *et al.* **Principles of neural science**. McGraw-Hill, v. 4, 2000.

KETCHEN, D. J.; SHOOK, C. L. The application of cluster analysis in strategic management research: an analysis and critique. **Strategic management journal**, v. 17, n. 6, p. 441-458, 1996.

KREPKE, R. Design and implementation of an online BCI system for the control in gaming applications and virtual limbs. **Tese (PhD in Computer Science)**, Berlin University, 2004.

KRIEGER, Hans-Peter *et al.* Density-based clustering. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, v. 1, n. 3, p. 231-240, 2011.

KRUPA, D. J., *et al.* Behavioral properties of the trigeminal somatosensory system in rats performing whisker-dependent tactile discriminations. **The Journal of Neuroscience**, v. 21, n. 15, p. 5752-5763. 2001.

KRUSKAL, W. H.; WALLIS, W. A. Use of ranks in one-criterion variance analysis. **Journal of the American statistical Association**, v. 47, n. 260, p. 583-621, 1952.

LAKATOS, E. M.; MARCONI, M. A. **Metodologia científica**. Atlas, 7 ed, 2010.

LEE, Chia-Wei. *et al.* Comparing The Spatial and Temporal Reproducibility of Brain Activation Using Three fMRI Techniques: BOLD, FAIR, and VASO. **In: Noninvasive Functional Source Imaging of the Brain and Heart and the International Conference**

on **Functional Biomedical Imaging. Joint Meeting of the 6th International Symposium on**, p. 258-261, 2007.

LENT, R. Cem Bilhões de Neurônios. **Atheneu**, 3 ed., 2010.

LI, N. *et al.* What response properties do individual neurons need to underlie position and clutter “invariant” object recognition? **Journal of neurophysiology**, v. 102, n. 1, p. 360, 2009.

LOMBROSO, P. Learning and memory. *Revista Brasileira de Psiquiatria*. v. 26, n. 3, p. 207–210, 2004.

LOPES-DOS-SANTOS, V., RIBEIRO, S., TORT, A. B. L. Detecting cell assemblies in large neuronal populations. **Journal of neuroscience methods**, v. 220, n. 2, p. 149-166, 2013.

LOPES-DOS-SANTOS, V. *et al.* Neuronal assembly detection and cell membership specification by principal component analysis. **PloS one**. v.6, n.6, p. 1-16, 2011.

MACQUEEN, J. *et al.* Some methods for classification and analysis of multivariate observations. **Proceedings of the fifth Berkeley symposium on mathematical statistics and probability**, p. 14, 1967.

MATSUMOTO, N. *et al.* Visualization of multi-neuron activity by simultaneous optimization of clustering and dimension reduction. **Neural Networks**, v. 23, n. 6, p. 743-751, 2010.

MENDELSON, Everett; SMITH, Merritt Roe; WEINGART, Peter. **Science, technology and the military**. Kluwer Academic Pub, 1989.

MISHKIN, Mortimer; UNGERLEIDER, Leslie G. Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. **Behavioural brain research**, v. 6, n. 1, p. 57-77, 1982.

MOURA, E. S. Agrupamento de faces em coleções de fotografias digitais. **Tese** (Doutorado em Ciências da Computação), Universidade Federal de Campina Grande, 2012.

NICOLELIS, M. A. L. *et al.* Reconstructing the engram: simultaneous, multisite, many single neuron recordings. **Neuron**, v. 18, n. 4, p. 529–537, 1997.

NICOLELIS, M. A. L. **Methods for neural ensemble recordings**. CRC press, 2 ed. 2007.

NICOLELIS, Miguel AL; LEBEDEV, Mikhail A. Principles of neural ensemble physiology underlying the operation of brain–machine interfaces. **Nature Reviews Neuroscience**, v. 10, n. 7, p. 530-540, 2009.

O'KEEFE, J.; NADEL, L. **The hippocampus as a cognitive map**. Oxford: Clarendon Press, 1978.

PERKEL, D. H.; BULLOCK, T. H. Neural Coding. **Neuroscience Research Progress Bulletin**, v. 6, n. 3, p. 221-348. 1968.

POLIKAR, R. **Pattern Recognition**. Wiley, New York, 2006.

RIBEIRO, S. *et al.* Novel experience induces persistent sleep-dependent plasticity in the cortex but not in the hippocampus. **Frontiers in Neuroscience**. v. 1, n. 1, p. 43-55, 2007.

RIEKE, F. **Spikes: exploring the neural code**. MIT press, 1999.

ROBLES-DE-LA-TORRE, G.; HAYWARD, V. Force can overcome object geometry in the perception of shape through active touch. **Nature**, v. 412, n. 6845, p. 445-448, 2001.

ROLLS, E. T.; TREVES, A. The neuronal encoding of information in the brain. **Progress in Neurobiology**. v. 95, n. 3, p. 448-490, 2011.

ROUSSEEUW, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. **Computational and Applied Mathematics**, v. 20, p. 53–65. 1987.

RUSSELL, S.; NORVIG, P. Artificial Intelligence – A Modern Approach, **Prentice Hall**, 3 ed, 2010.

SCHLOGL, A.; BRUNNER, C. Biosig: A free and open source software library for BCI research. **Computer**, v. 41, n. 10, p 44-50, 2008.



SHALCHYAN, V.; FARINA, D. A non-parametric Bayesian approach for clustering and tracking non-stationarities of neural spikes. **Journal of Neuroscience Methods**. v. 223, p. 85-91, 2014.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, 52 ed, p. 591-611, 1965.

SQUIRE, L. R. *et al.* Fundamental Neuroscience. **Academic Press**, 4 ed, 2012.

STERRATT, D. *et al.* Principles of computational modelling in neuroscience. **Cambridge University Press**, 2011.

STORK, D. G.; DUDA, R. O.; HART, P. E. **Pattern classification**. John Wiley & Sons, 2 ed, 2001.

TAMURA, S. Random bin for analyzing neuron spike trains. **Computational intelligence and neuroscience**, v. 2012, p. 1, 2012.

THEODORIDIS, S. *et al.* **Introduction to Pattern Recognition: A Matlab Approach: A Matlab Approach**. Academic Press, 2010.

VAN ESSEN, D. C.; GALLANT, J. L. Neural mechanisms of form and motion processing in the primate visual system. **Neuron**, v. 13, n. 1, p. 1-10, 1994.

VANCE, A. **Data Analysts Captivated by R's Power**, New York Times, 2009.

VASCONCELOS, N. *et al.* Cross-modal responses in the primary visual cortex encode complex objects and correlate with tactile discrimination. **Proceedings of the National Academy of Sciences**, v. 108, n. 37, p. 15408-15413, 2011.

WAZLAWICK, R. **Metodologia de pesquisa para ciência da computação**. Elsevier Brasil, 2009.

WILK, M. B.; GNANADESIKAN, R. Probability plotting methods for the analysis of data, **Biometrika (Biometrika Trust)**, v. 55, n.1, p. 1-17, 1968.

WU, M. C.-K.; DAVID, S. V.; GALLANT, J. L. Complete functional characterization of sensory neurons by system identification. **Annual Review of Neuroscience**., v. 29, p. 477-505, 2006.

YANG, Y. *et al.* Millisecond-scale differences in neural activity in auditory cortex can drive decisions. **Nature neuroscience**, v. 11, n. 11, p. 1262-1263, 2008.

## Apêndice A – Amostra de registros analisados na abordagem semi-supervisionada

Neste apêndice está uma amostra dos registros utilizados na análise semi-supervisionada realizada nesse trabalho de dissertação.

Conforme a Seção 3.5.3, 294912 distribuições foram analisadas em agrupamentos de acordo com as hipóteses descritas na Seção 3.5.4.

Na Seção 4.3, foram apresentados um grupo de *boxplots* para cada hipótese testada. A seguir são apresentados outros três grupos de *boxplots* (Figura A1 a Figura A24), selecionados ao acaso entre demais grupos, que também foram analisadas em cada teste de hipótese, e o que é inferido sobre as hipóteses as quais estão relacionados.

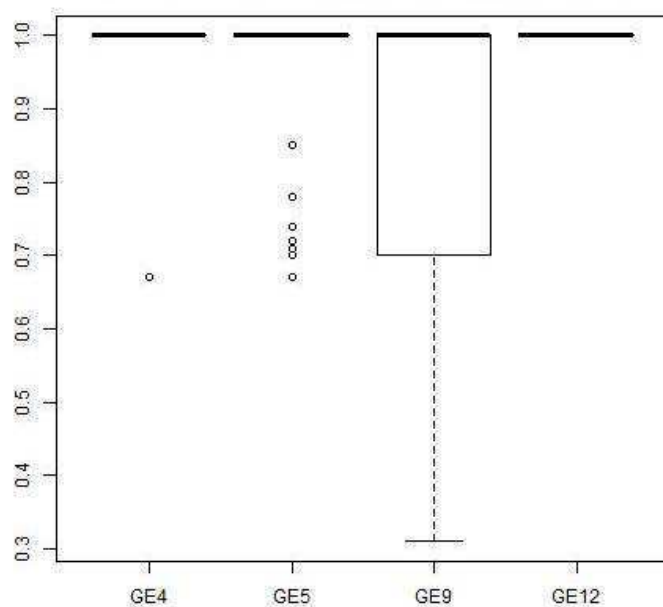


Figura A1: *Boxplot* das distribuições  $u_{(GE4,GE5,GE9,GE12)}_{HP_{[Bola,Escova]}_K-Means\_Jaccard\_50b\_5j\_c1}$

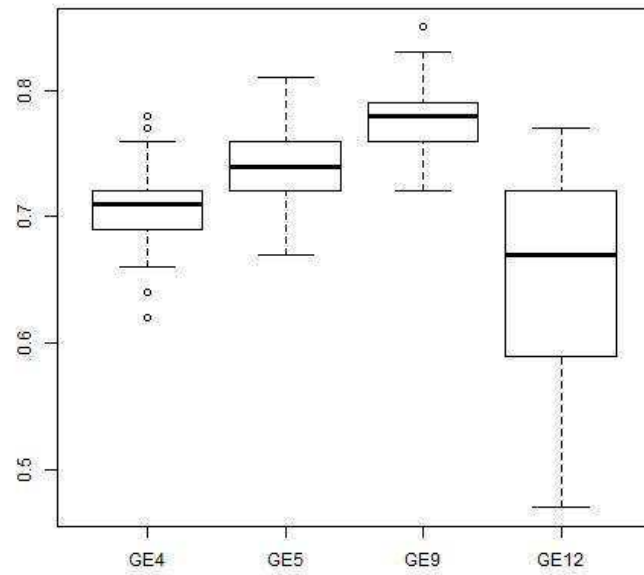


Figura A2: *Boxplot* das distribuições  $u_{(GE4,GE5,GE9,GE12)}_{S1}_{[Escova,Comida]}_{EM\_Dunn\_50b\_10j\_c3}$

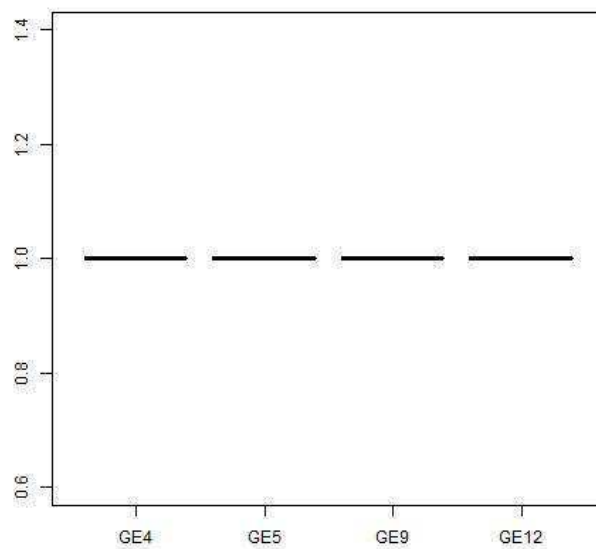


Figura A3: *Boxplot* das distribuições  $u_{(GE4,GE5,GE9,GE12)}_{V1}_{[Comida, Haste]}_{Hier\_Silhueta\_100b\_15j\_ic}$

De acordo com a análise das Figuras A1 a Figura A3, foi rejeitada a hipótese  $H1-0$ , em

que é defendido que independente dos animais, os níveis de similaridade são os mesmos.

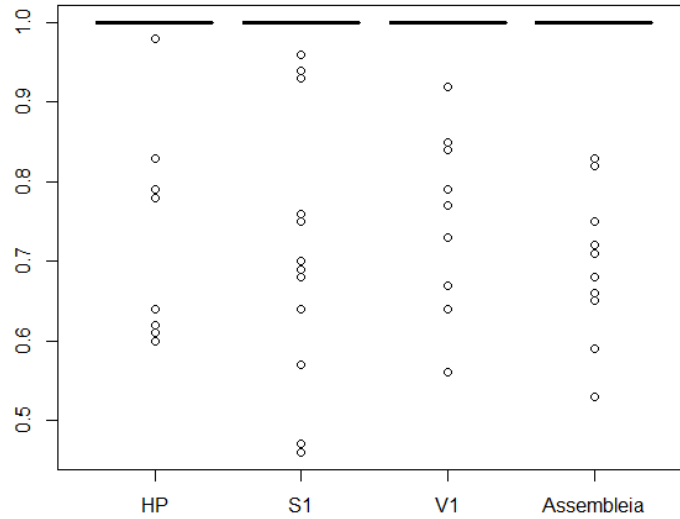


Figura A4: *Boxplot* das distribuições  $u_{GE4}(HP, S1, V1, Assembleia)_{[Bola, Escova]}_{K-Means\_Jaccard\_50b\_5j\_ic}$

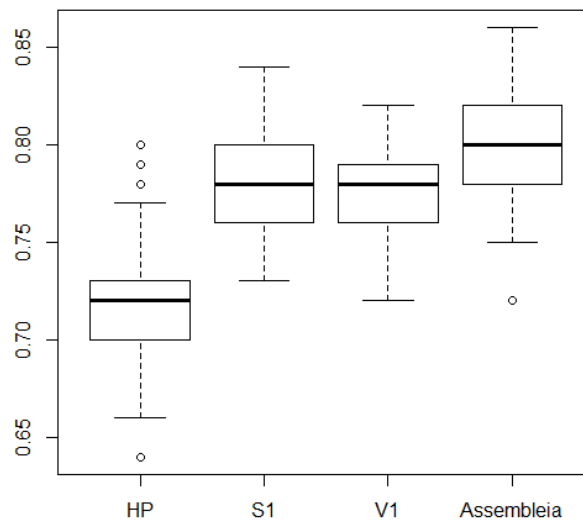


Figura A5: *Boxplot* das distribuições  $u_{GE5}(HP, S1, V1, Assembleia)_{[Escova, Comida]}_{EM\_Dunn\_50b\_10j\_i3}$

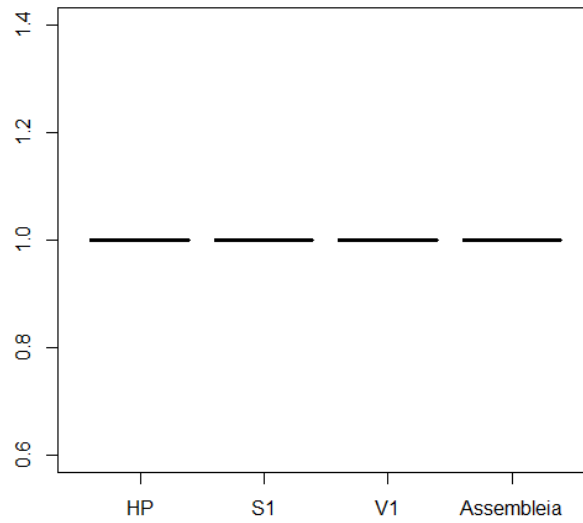


Figura A6: *Boxplot* das distribuições  $u_{GE9}$  (HP, S1, V1, Assembleia)\_[Comida, Haste]\_Hier\_Silhueta\_100b\_15j\_ic

Conforme o que é apresentado nas Figuras A4 a Figura A6, foi rejeitada a hipótese H2-0, em que é defendido que independente das regiões do cérebro, os níveis de similaridade são os mesmos.

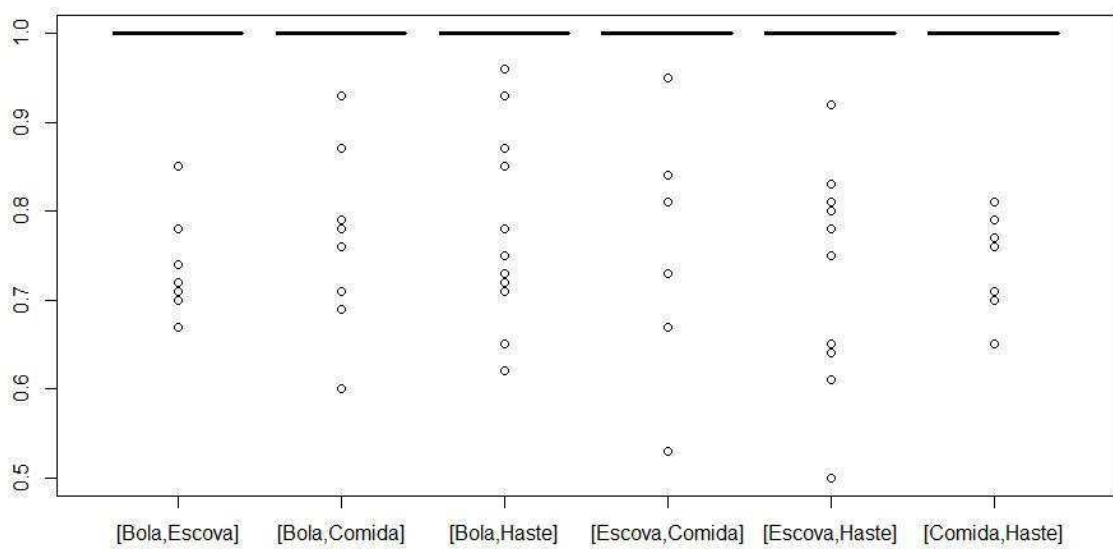


Figura A7: *Boxplot* das distribuições  $u_{GE4\_HP}$  ([Bola, Escova], [Bola, Comida], [Bola, Haste], [Escova, Comida], [Escova, Haste], [Comida, Haste])\_K-Means\_Jaccard\_50b\_5j\_i1

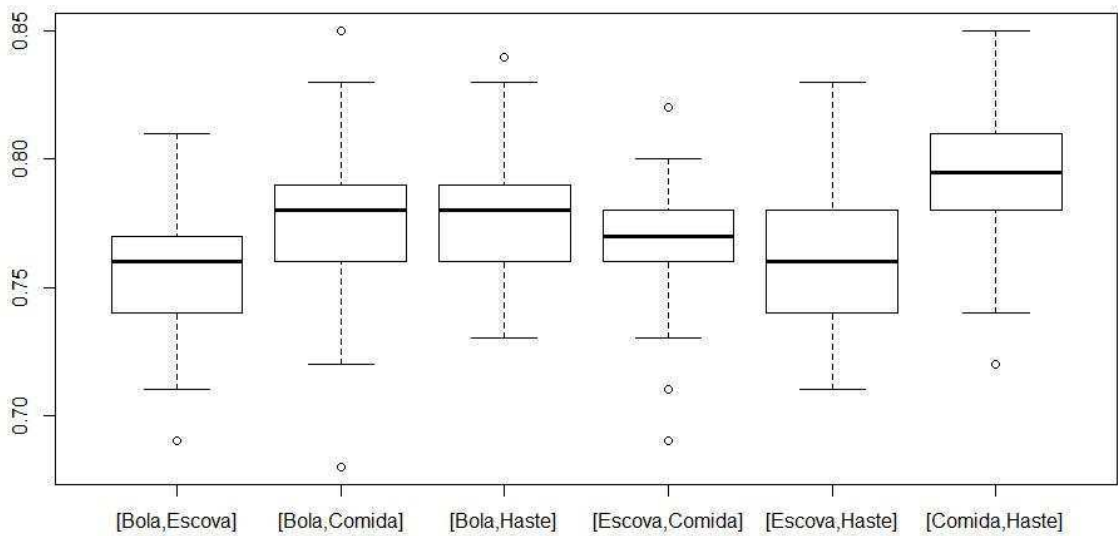


Figura A8: *Boxplot* das distribuições  $u_{GE5\_V1\_}$ ([Bola,Escova], [Bola,Comida], [Bola,Haste], [Escova, Comida], [Escova, Haste], [Comida,Haste])\_Escova, Comida)\_EM\_Dunn\_50b\_10j\_i3

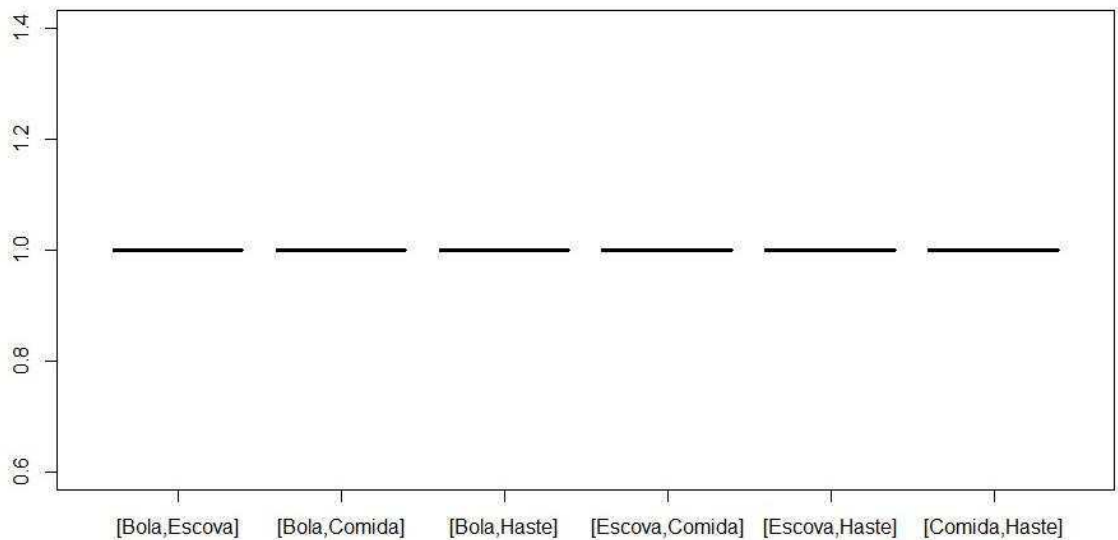


Figura A9: *Boxplot* das distribuições  $u_{GE9\_S1\_}$ ([Bola,Escova], [Bola,Comida], [Bola,Haste], [Escova, Comida], [Escova, Haste], [Comida,Haste])\_Hier\_Silhueta\_100b\_15j\_i0

Nas Figuras A7 a Figura A9, é evidenciado que o nível de similaridade entre as classes de padrões de ativação neuronal é dependente do par de objetos analisado. Desta forma, foi

rejeitada a hipótese H3-0.

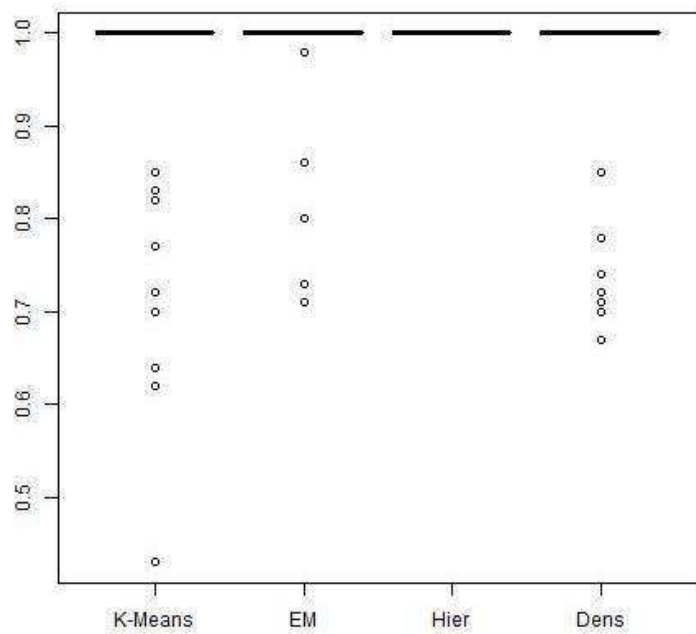


Figura A10: *Boxplot* das distribuições  $u_{GE4\_HP\_}[Bola,Escova]$  (K-Means, EM, Hier, Dens)\_Jaccard\_50b\_5j\_i1

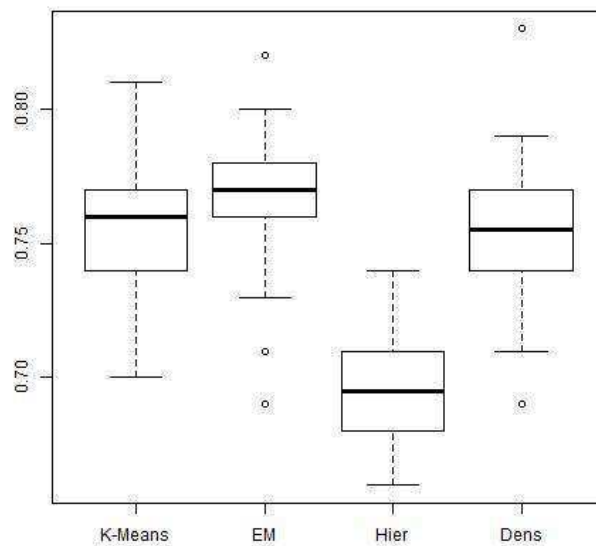


Figura A11: *Boxplot* das distribuições  $u_{GE5\_V1\_}[Escova, Comida]$  (K-Means, EM, Hier, Dens)\_Dunn\_50b\_10j\_i3



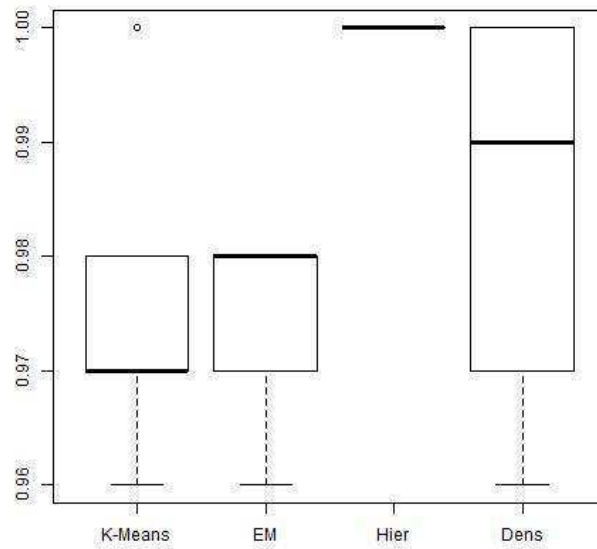


Figura A12: *Boxplot* das distribuições  $u_{GE9\_S1\_}[Comida,Haste]_{(K-Means, EM, Hier, Dens)}_{Silhueta\_100b\_15j\_i0}$

De acordo com a análise das Figuras A10 a Figura A12, foi rejeitada a hipótese H4-0, em que é defendido que independente dos algoritmos de agrupamento, os níveis de similaridade são os mesmos.

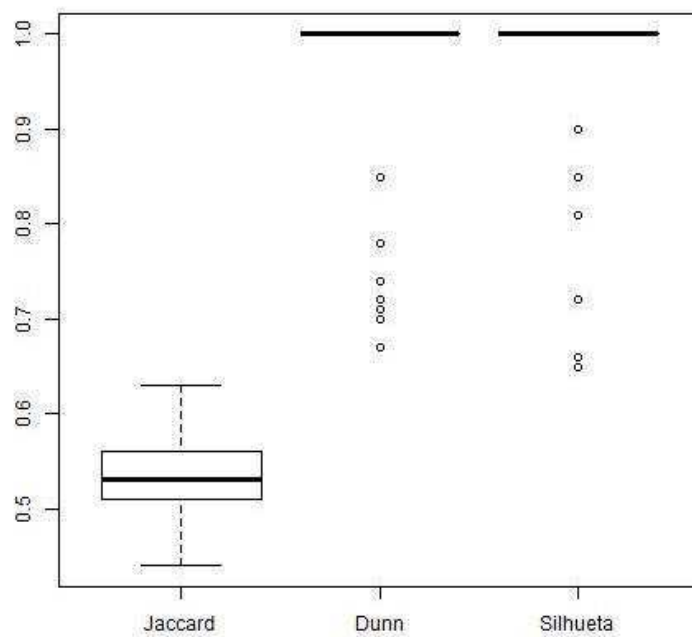


Figura A13: *Boxplot* das distribuições  $u_{GE4\_HP\_}[Bola,Escova]_{K-Means_{(Jaccard, Dunn, Silhueta)}_{50b\_5j\_i1}}$

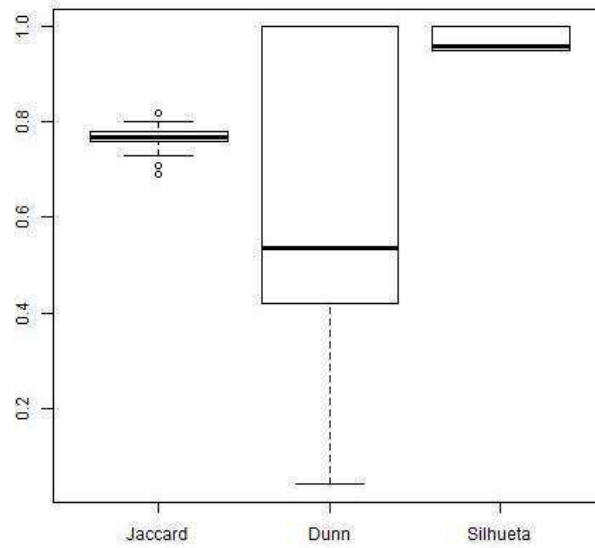


Figura A14: *Boxplot* das distribuições  $u_{GE5\_V1\_}[\text{Escova, Comida}]_{EM\_}(\text{Jaccard, Dunn, Silhueta})_{50b\_10j\_i3}$

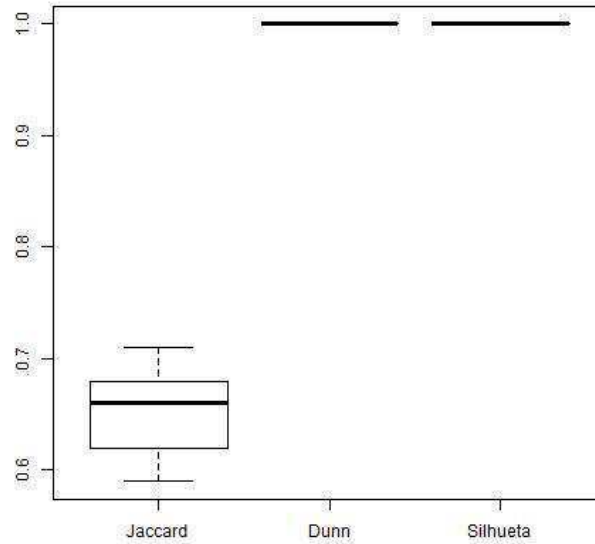


Figura A15: *Boxplot* das distribuições  $u_{GE9\_S1\_}[\text{Comida,Haste}]_{Hier\_}(\text{Jaccard, Dunn, Silhueta})_{100b\_15j\_i0}$

Nas Figuras A13 a Figura A15, é evidenciado que o nível de similaridade entre as classes de padrões de ativação neuronal é dependente da métrica. Desta forma, foi rejeitada a hipótese H5-0.

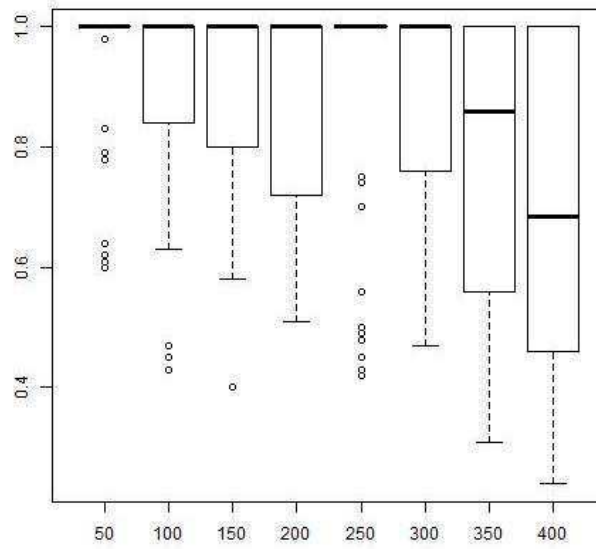


Figura A16: *Boxplot* das distribuições  $u_{GE4\_HP\_}[Bola,Escova]_{K-Means\_Jaccard\_}(50-400)b_{5j\_ic}$

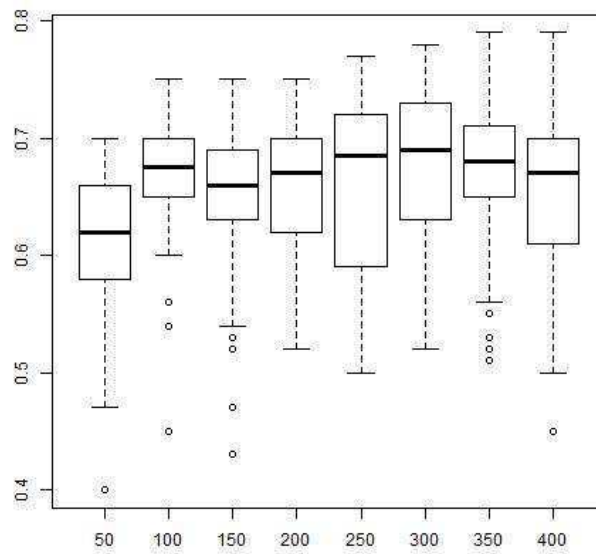


Figura A17: *Boxplot* das distribuições  $u_{GE9\_V1\_}[Escova, Comida]_{Dunn\_}(50-400)b_{5j\_ic}$

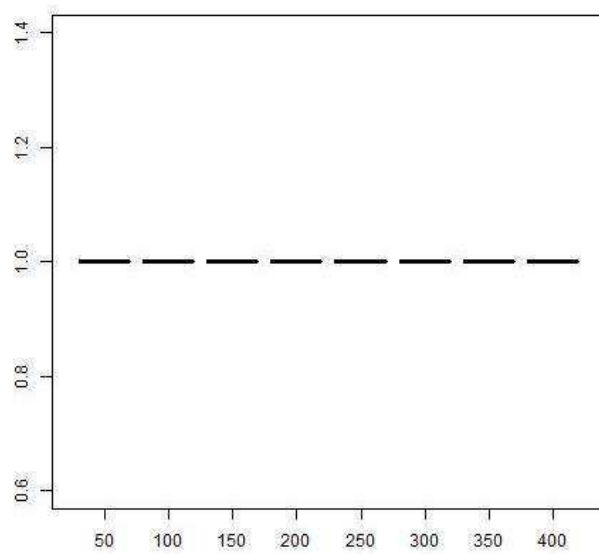


Figura A18: *Boxplot* das distribuições  $u_{GE12\_S1\_}[Comida,Haste]_{Hier\_Silhueta\_}(50-400)b_{10j\_ic}$

Em relação a  $H6-0$ , em que é defendido que independentemente do tamanho do *bin*, os níveis de similaridade são os mesmos, a análise das Figuras A16 a Figura A18 suporta rejeição desta hipótese.

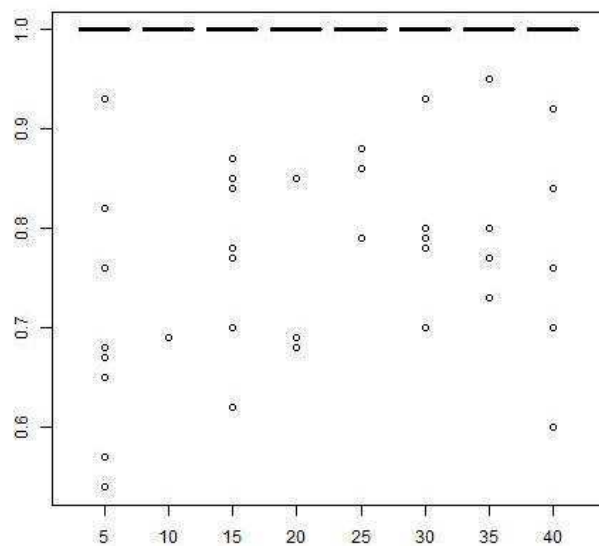


Figura A19: *Boxplot* das distribuições  $u_{GE4\_HP\_}[Bola,Escova]_{K-Means\_Jaccard\_50b\_}(5-40)j_{i1}$

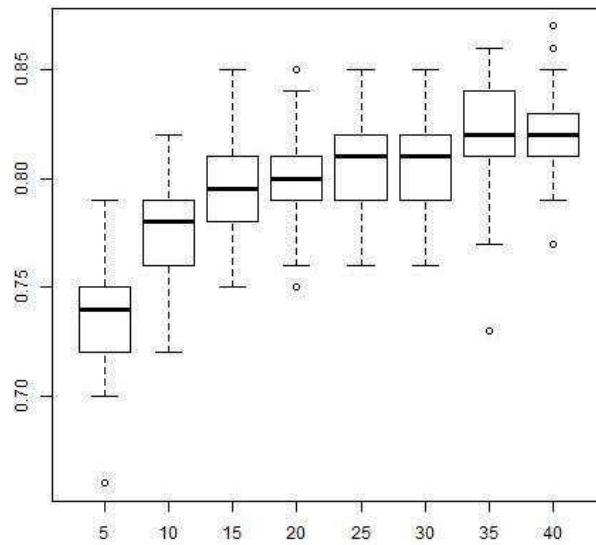


Figura A20: *Boxplot* das distribuições  $u_{GE5\_V1\_} [Escova, Comida]_{EM\_Dunn\_50b\_} (5-40)j_{ic}$

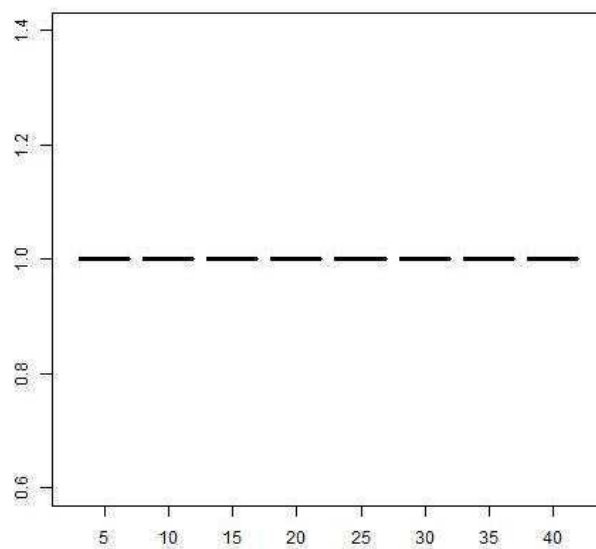


Figura A21: *Boxplot* das distribuições  $u_{GE9\_S1\_} [Comida,Haste]_{Hier\_Silhueta\_100b\_} (5-40)j_{ic}$

De acordo com a análise das Figuras A19 a Figura A21, foi rejeitada a hipótese  $H7_0$ , em que é defendido que independentemente do tamanho da janela, os níveis de similaridade são os mesmos.

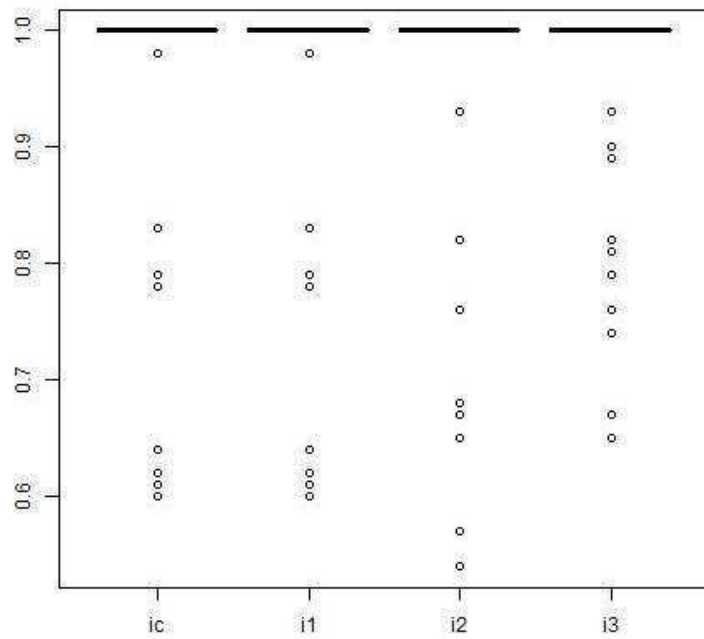


Figura A22: *Boxplot* das distribuições  $u_{GE4\_HP\_}[Bola,Escova]_{K-Means\_Jaccard\_50b\_5j\_}(ic, i1, i2, i3)$

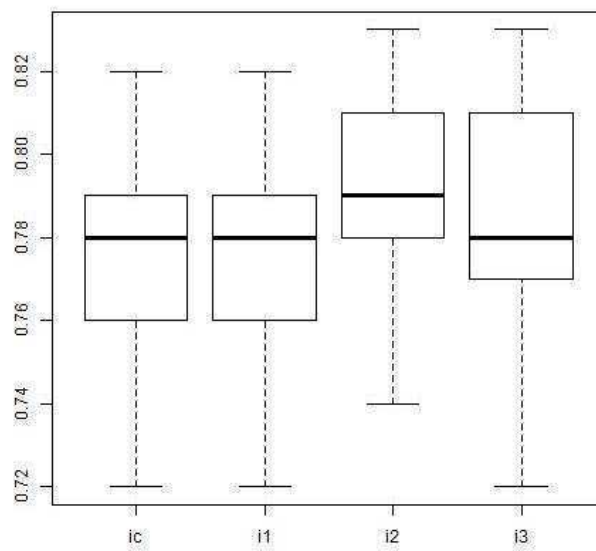


Figura A23: *Boxplot* das distribuições  $u_{GE5\_V1\_}[Escova, Comida]_{EM\_Dunn\_50b\_10j\_}(ic, i1, i2, i3)$

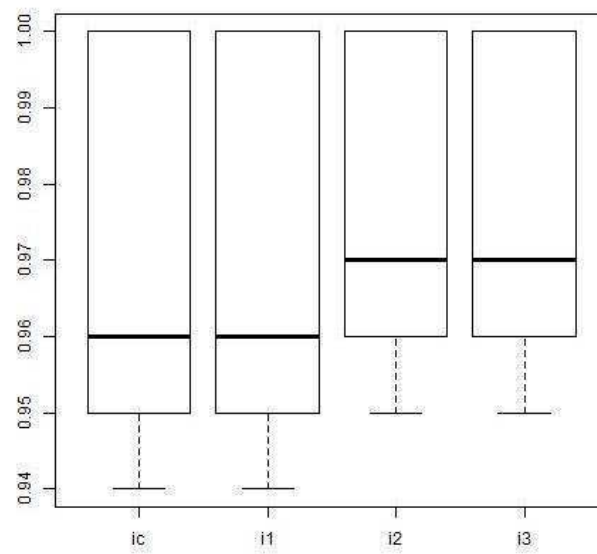


Figura A24: *Boxplot* das distribuições  
u\_GE12\_S1\_[Escova,Haste]\_Hier\_Silhueta\_100b\_15j\_(ic, i1, i2, i3)

Conforme o que é apresentado nas Figuras A22 a Figura A24, foi rejeitada a hipótese H8-0, em que é defendido que os níveis de similaridade são os mesmos ao longo do tempo.

## Apêndice B – Fundamentação Teórica

A seguir, são apresentados os conceitos julgados importantes para uma maior compreensão do trabalho. Os conceitos estão divididos nos subtópicos intitulados Conceitos da Ciência da Computação, que aborda os conceitos de Aprendizagem, Análise de Agrupamento e Métricas, Conceitos da Neurociência, que trata das Regiões de Interesse no Cérebro, Neurônios, Bin, Janela, Métodos de Aquisição de Dados Neurais, Matrizes de Microelétrodos e Padrões de Ativação Neuronal, e Conceitos da Metodologia Científica, que contém Tipo de Estudo, Fatores, Projeto do Experimento e Análise Estatística.

### *B.1 Conceitos da Ciência da Computação*

A Ciência da Computação é área do conhecimento em que é realizado o estudo dos métodos e das técnicas destinadas a modelar, analisar e resolver problemas utilizando recursos computacionais (FOROUZAN, MOSHARRAF, 2012). Essa possui um grande conjunto de conceitos relacionados às atividades e aos objetos de seu domínio, tais como Aprendizagem e Agrupamento de Dados, os quais são explorados neste trabalho de Dissertação, sendo descritos a seguir.

#### B.1.1 Aprendizagem

A aprendizagem de máquina é um dos ramos da Inteligência Artificial, dedicado à aquisição de conhecimento a partir de dados ou observações empíricas, e ao desenvolvimento de algoritmos e técnicas que permitam a otimização de tarefas, utilizando de raciocínio indutivo (extração de regras e padrões) para realização de generalização, classificação e sumarização (RUSSEL, NORVIG, 2010).

A aprendizagem de máquina pode ser tratada como um problema de aproximação de funções, na medida em que, com base em dados já conhecidos, é possível estimar regras de produção para expressão de uma decisão sobre dados não conhecidos.

Em termos didáticos, de acordo com Russell e Norvig (2010), a aprendizagem pode



classificada em três categorias, conforme descritas a seguir.

- “Aprendizagem Não Supervisionada: O agente aprende padrões na entrada, embora nenhuma realimentação explícita seja fornecida. A tarefa mais comum da aprendizagem não supervisionada é o agrupamento - a detecção de grupos potencialmente úteis de exemplos de entrada. Por exemplo, um agente de táxi pode gradualmente desenvolver um conceito de ‘bons dias de trânsito’ e ‘dias de tráfego ruim’ sem nunca terem sido dados exemplos rotulados de cada por outro agente.
- Aprendizagem por Reforço: O agente aprende a partir de uma série de reforços de recompensas ou de punições. Por exemplo, a falta de uma ponta na extremidade do trajeto dá o agente táxi uma indicação de que ele fez algo de errado. Os dois pontos por vitória, no final de um jogo de xadrez dizem que o agente fez algo certo. Cabe ao agente decidir as ações antes do reforço foram mais responsáveis por ela.
- Aprendizagem Supervisionada: O agente observa alguns exemplos de pares de entrada-saída e aprende uma função que mapeia da entrada à saída. As entradas são percepções e as saídas são fornecidos pelo agente.”

Ainda de acordo com Russell e Norvig (2010),

“Na prática, essas distinções nem sempre são tão nítidas, tornando possível a classificação de aprendizado semi-supervisionado, na qual, com base em alguns exemplos rotulados, busca-se rotular uma grande quantidade de exemplos. Mesmo os rótulos em si podem não ser as verdades oraculares esperadas.

Uma ilustração desse conceito é feita por meio do seguinte exemplo: imagine o desenvolvimento de um sistema que adivinha a idade de uma pessoa a partir de uma foto. São disponibilizados algumas fotos rotuladas pelas próprias pessoas fotografadas. Até então, a aprendizagem do sistema é classificada como supervisionada. Mas, na realidade algumas das pessoas mentiram sobre sua idade. Não é apenas um fator de ruído aleatório nos dados, mas

sim as imprecisões são sistemáticas, e descobri-las é um problema de aprendizado não supervisionado envolvendo imagens, idades auto-relatadas e verdadeiras idades (desconhecidas). Assim, tanto as imprecisões e a falta de rótulos exigem uma classificação de aprendizagem entre supervisionada e não-supervisionada (semi-supervisionada).”

## B.1.2 Análise de Agrupamentos de Dados

A atividade de análise de agrupamentos de dados (*cluster analysis*) pode ser definida como a categorização de um conjunto total de elementos em diferentes subgrupos, cada um contendo elementos mais semelhantes entre si e mais distintos dos demais elementos dos outros grupos, segundo alguma função de distância (JAIN, DUBES, 1988; JAIN, MURTY, FLYNN, 1999).

Para a realização dessa atividade, existem algoritmos, tais como K-Means (MACQUEEN, 1967), DBScan (ESTER *et al.*, 1996) e EM (DEMPSTER, LAIRD, RUBIN, 1977), os quais possuem parâmetros que necessitam ser configurados. Para a configuração desses parâmetros, em geral não é necessário conhecimento prévio quanto aos possíveis grupos existentes em um conjunto de dados (THEODORIDIS *et al.*, 2010).

Uma vez agrupados os dados, pode ser iniciada a análise de características internas, como por exemplo, o número de elementos em cada subgrupo ou o nível de coesão do grupo e a distância entre dois subgrupos resultantes, e externas, como a distância de um determinado subgrupo resultante e um grupo de referência (JAIN, DUBES, 1988; THEODORIDIS *et al.*, 2010; JAIN, MURTY, FLYNN, 1999).

Diante da grande variedade de algoritmos de agrupamento existentes (JAIN, DUBES, 1988; JAIN, MURTY, FLYNN, 1999), e buscando analisar o nível generalidade das inferências realizadas neste trabalho de pesquisa, foi escolhido um subconjunto representativo dos principais tipos de algoritmos de agrupamento (JAIN, DUBES, 1988; JAIN, MURTY, FLYNN, 1999), o qual é apresentado a seguir.

### **Agrupamento com base em centróides (K-Means)**

O método K-Means (MACQUEEN, 1967) realiza a partição dos dados em  $k$  agrupamentos em que cada observação pertence ao grupo mais próximo da média.

Matematicamente, o algoritmo pode ser descrito conforme a Equação (1) e descrições apresentadas a seguir.

**Parâmetros:** Um conjunto de elementos  $(x_1, x_2, \dots, x_n)$ . Cada elemento é um vetor  $d$ -dimensional e um número inteiro  $k$ .

**Objetivo:** Agrupar os  $n$  elementos em  $k$  conjuntos ( $k < n$ ),  $S = \{S_1, S_2, \dots, S_k\}$ , de modo a minimizar a soma dos quadrados dentro de cada grupo.

$$\arg_S \min \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \mu_i\|^2 \quad (1)$$

$\mu_i$  é a média dos elementos em  $S_i$ .

### Agrupamento com base em distribuições (Maximização da Expectativa)

A Maximização da Expectativa - *Expectation-maximization* - (EM) é um método de refinamento de agrupamento que objetiva encontrar estimativas de máxima verossimilhança ou máxima probabilidade *a posteriori* (MAP) de parâmetros em modelos estatísticos, em que o modelo depende de variáveis latentes não observadas (DEMPSTER, LAIRD, RUBIN, 1977).

Conforme Schön (2009), matematicamente, o algoritmo pode ser descrito como:

**Parâmetros:**

**X:** Dados observados em um conjunto de  $m$  instâncias ocorridas independentemente;

**Z:** Dados não observados destas mesmas instâncias;

**Y:**  $X \cup Z$ ;

$\theta$ : Sendo  $Z$  uma variável aleatória,  $\theta$  é o conjunto de parâmetros desconhecidos que justificam a distribuição de probabilidade de  $Z$  e dos dados observados  $X$ ;

$h$ : Hipótese atual dos parâmetros  $\theta$ ; e

$h'$ : Hipótese revisada que é estimada em cada iteração do algoritmo EM.

**Objetivo:** Encontrar  $h'$  que maximize alguma função desta quantidade  $P(Y|h')$  e o logaritmo desta quantidade  $\log P(Y|h')$ .

**Algoritmo:**

1. Considerando a Equação (2)

$$Q(h'|h) = E[\ln P(Y|h')|h, X] \quad (2)$$

2. Enquanto não converge fazer:
  - a. **Passo da Expectativa (E):** Calcular  $Q(h'|h)$  usando a hipótese atual  $h$  e os dados observados  $X$  para estimar a distribuição de probabilidade sobre  $Y$ .
  - b. **Passo de Maximização (M):** Trocar a hipótese  $h$  pela hipótese  $h'$  que maximiza a função  $Q$ .

**Agrupamento com base em conectividade (Agrupamento Hierárquico)**

Algoritmos de agrupamento que têm como base conectividade buscam formar grupos a partir de estimativas de distância entre os elementos. Deste modo, os grupos podem ser descritos pela distância máxima entre dois membros do grupo.

Quando alteramos o quão deve ser a máxima a distância entre os itens de um grupo, novos grupos poderão ser formados, criando assim uma árvore, também definida como

dendrograma, com níveis de agrupamento, conforme pode ser visualizado na Figura B1.

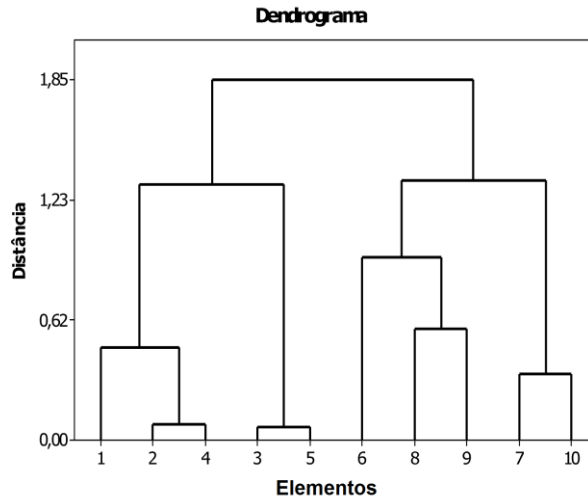


Figura B1: Dendrograma resultante de um processo de agrupamento hierárquico.

Por fornecer grupos hierarquicamente ordenados a partir de grupos que se fundem uns com os outros em determinadas distâncias, esses algoritmos também são conhecidos como algoritmos de agrupamento hierárquico.

Os vários algoritmos dessa categoria se diferenciam pela maneira com que estabelecem as ligações entre os elementos de um grupo e calculam a distância entre eles.

Dentre as estratégias utilizadas para se estabelecer as ligações entre os elementos de um grupo, as mais utilizadas são as ligações simples, completa e por média.

As ligações simples determinam que dois elementos estão no mesmo grupo sempre que a distância entre eles for a menor dentre as distâncias de cada um deles para com os demais elementos.

As ligações completas determinam que dois elementos estão no mesmo grupo sempre que a distância entre eles for a maior dentre as distâncias de cada um deles para com os demais elementos.

As ligações por média determinam que dois elementos estão no mesmo grupo sempre que a distância entre eles for menor ou igual a média entre as distâncias encontradas entre os

elementos de cada um dos grupos desses elementos. Um exemplo de algoritmo que utiliza essa estratégia de ligação é o UPGMA (Método não ponderado de Agrupamento de Pares, utilizando Média Aritmética”, do Inglês: *Unweighted Pair Group Method with Arithmetic Mean*), o qual foi empregado neste trabalho de dissertação e cuja a formulação é apresentada na Equação (3).

Sejam  $A$  e  $B$  conjuntos de elementos, o cálculo do UPGMA é definido como:

$$\frac{1}{|A| \cdot |B|} \sum_{x \in A} \sum_{y \in B} d(x, y) \quad (3)$$

Além disso, agrupamento hierárquico pode ser classificado como divisivo (*top-down*), quando os agrupamentos são formados pela divisão de grupos maiores em menores, e como aglomerativo (*bottom-up*), quando os agrupamentos são formados a partir de elementos únicos até a formação dos grupos.

Em casos gerais, algoritmos hierárquicos possuem complexidade computacional igual a  $O(n^3)$ , o que os tornam onerosos para grandes conjuntos de dados.

### **Agrupamento com base em densidade**

Algoritmos de agrupamento com base em densidade têm como ideia central a definição de áreas de densidade para a classificação de elementos, os quais são ditos pertencente ao grupo com maior densidade (KRIEGEL *et al.*, 2011).

Dentre os algoritmos desse tipo, o mais utilizado pela comunidade científica é o DBScan (ESTER *et al.*, 1996), e o qual também foi empregado neste trabalho.

Esse algoritmo apresenta um modelo de agrupamento definido como "densidade - acessibilidade" o qual estabelece ligação entre dois elementos com base na densidade dos objetos formados pelos grupos de cada um deles, e de um certo limite de distância.

Nesse algoritmo, a forma de um conjunto pode ser arbitrária.

Em relação à complexidade computacional, algoritmos desse tipo, tem custo, em casos gerais, igual a  $O(n^2)$ .

As desvantagens desse tipo de algoritmo de agrupamento são a necessidade de queda de densidade para detectar fronteiras entre os grupos e a impossibilidade de detectar estruturas aglomeradas intrinsecamente, problema esse bastante comum em dados reais (ESTER *et al*, 1996).

### B.1.3 Métricas

No contexto de análise de agrupamento de dados, métricas podem ser definidas como instrumentos utilizados para mensurar a qualidade de um agrupamento de acordo com determinada característica e indicar uma possível quantidade de subgrupos presentes em um conjunto de dados. A seguir, são apresentadas as métricas utilizadas nesta pesquisa para mensurar o nível de similaridade entre os agrupamentos de padrões de ativação neuronal, as quais foram selecionadas com intuito de testar a generalidade das inferências realizadas neste trabalho de pesquisa

#### Índice de Dunn

O índice de Dunn (1974) é definido como a razão entre a distância mínima entre agrupamentos e a maior distância dentre as internas de cada agrupamento.

Essa métrica é formulada conforme a Equação (4).

$$D = \min_{1 \leq i \leq n} \left\{ \min_{1 \leq j \leq n, i \neq j} \left\{ \frac{d(i, j)}{\max_{1 \leq k \leq n} d'(k)} \right\} \right\} \quad (4)$$

Em que  $n$  é o número de agrupamentos;  $d(i, j)$  é a distância entre os agrupamentos  $i$  e  $j$ ; e  $d'(k)$  é a distância interna do agrupamento  $k$ .

Por analisar o nível de coesão interna dos grupos pelo nível de coesão externo dos grupos, quanto maior o valor do índice de Dunn, melhor o resultado do agrupamento analisado.

### Índice de Jaccard

O índice de Jaccard (JAIN, DUBES 1988) quantifica o quão semelhantes são dois conjuntos, assumindo valores entre 0 (quando os conjuntos são totalmente diferentes) e 1 (totalmente iguais). Esse valor é calculado pela razão entre o número de elementos da interseção entre os dois conjuntos e o número de elementos da união entre os dois conjuntos.

É definido pela seguinte Equação (5).

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \quad (5)$$

em que  $A$  e  $B$  são dois conjuntos.

### Índice Silhueta (Silhouette index)

O índice Silhueta (ROUSSEEUW, 1987) mensura o quão adequado está um elemento dentro de um agrupamento, com base em uma medida de proximidade intra e intergrupo. Essa métrica assume valores entre -1 (mais adequado ao grupo vizinho) e 1 (mais adequado ao próprio grupo). Valores iguais a zero, indicam que o elemento está na fronteira entre o seu grupo e um grupo vizinho.

Matematicamente, pode ser descrito conforme a Equação (6).

Em que  $k$ : Número de grupos;  $i$ : Índice que referencia um elemento do conjunto total de elementos;  $a(i)$ : Nível de dissimilaridade médio do elemento  $i$  com todos os outros elementos de seu grupo; e  $b(i)$ : Menor nível de dissimilaridade médio do elemento  $i$  a qualquer outro grupo no qual ele não pertença.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (6)$$



Quando analisada a média de todos os  $s(i)$  do conjunto total de elementos, tem-se mensurado o qual adequado está o agrupamento dos elementos.

Essa métrica também é empregada para determinar o número de subgrupos presentes em um conjunto de dados.

### Índice Cotovelo (*Elbow index*)

O índice Cotovelo (KETCHEN, SHOOK, 1996) é uma métrica utilizada para estipular o número ideal de subgrupos presentes em um conjunto de dados.

O valor dessa métrica é função do percentual de variância do agrupamento. O percentual de variância é a razão entre a variância entre os grupos e a variância total.

O aumento do número de agrupamentos é proporcional ao percentual de variância. O Índice Cotovelo é o ponto na curva onde o ganho no percentual de variância é reduzido em relação ao número anterior de agrupamentos. Esse conceito é ilustrado em um exemplo na Figura B2.

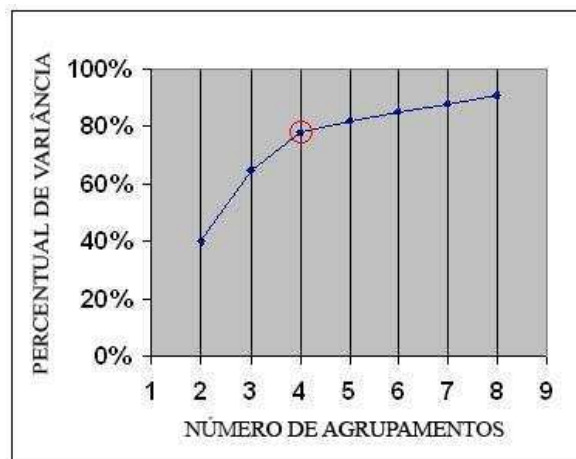


Figura B2: Exemplo de aquisição do Índice Cotovelo. Adaptado da Wikimedia.  
[http://upload.wikimedia.org/wikipedia/commons/c/cd/DataClustering\\_ElbowCriterion.JPG](http://upload.wikimedia.org/wikipedia/commons/c/cd/DataClustering_ElbowCriterion.JPG)

Conforme ilustrado na Figura B2, nesse exemplo, o número de agrupamentos ideal de acordo com o Índice Cotovelo é 4.

## B.2 *Conceitos da Neurociência*

A Neurociência é uma ciência multidisciplinar que analisa o sistema nervoso para entender a biologia básica do comportamento (LENT, 2010; SQUIRE *et al.*, 2012). Esta Ciência contempla conceitos específicos das atividades e dos objetos de estudo utilizados em suas pesquisas. A seguir, são apresentados alguns desses conceitos julgados importantes para melhor compreensão deste trabalho de dissertação.

### B.2.1 Regiões de Interesse no Cérebro

O cérebro é o órgão central do sistema nervoso dos animais vertebrados, sendo também presente em muitos animais invertebrados, como insetos e moluscos (BEAR, CONNORS, PARADISO, 2008).

Dentre as funções exercidas pelo cérebro, estão o controle e o monitoramento do comportamento, a partir de estímulos eletrofisiológicos (LENT,2010).

Contudo, existem comportamentos que não são necessariamente processados no cérebro. A maioria dos vertebrados possui circuitos neurais na própria coluna vertebral capazes de processar respostas reflexas e menos sofisticadas, assim como padrões motores simples (GRILLNER *et al.*, 2005).

Compreender o cérebro sempre foi um dos grandes objetivos da ciência, embora ainda hoje, muitas questões quanto ao seus mecanismos de processamento, aprendizagem continuem não solucionadas (BEAR, CONNORS, PARADISO, 2008).

Uma grande característica do cérebro é a plasticidade, com a qual, pode alocar diferentes seguimentos de sua composição para a realização de tarefas variadas. No entanto, também é sabido que o cérebro possui regiões com processamentos preferenciais para determinadas tarefas.

Essas correntes de explicação sobre o funcionamento do cérebro são conhecidas como princípio da localidade e processamento cross-modal, respectivamente (BEAR, CONNORS, PARADISO, 2008; LENT,2010).

Na Figura B3, são ilustradas as principais áreas do cérebro, as quais, por sua vez, podem ser segmentadas em diversas sub-regiões conforme sua função.

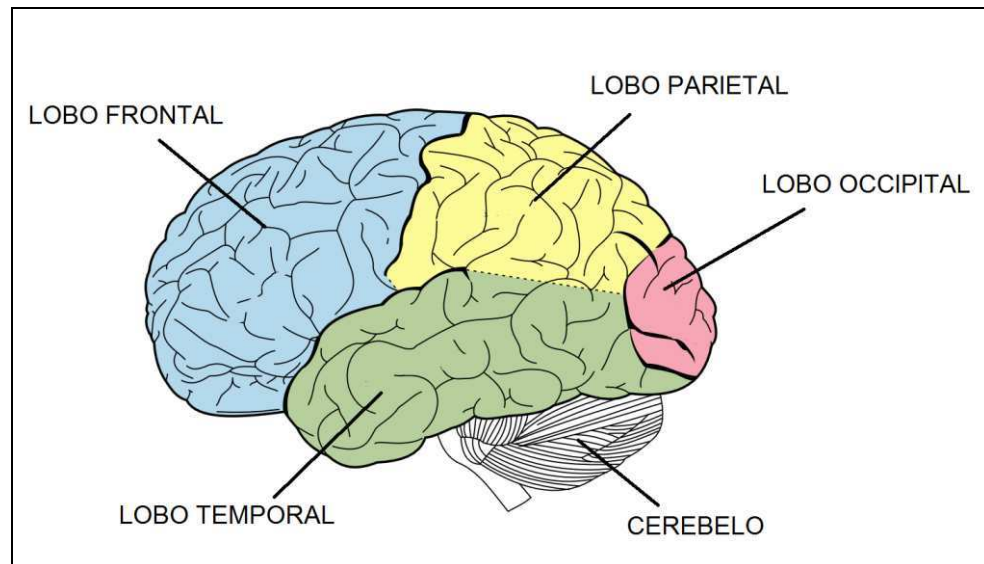


Figura B3: Ilustração das principais áreas do cérebro. Adaptado da Wikimedia (<http://upload.wikimedia.org/wikipedia/commons/1/1a/Gray728.svg>).

A seguir, serão apresentadas as regiões que foram utilizadas neste trabalho de Dissertação.

### **Hipocampo**

O hipocampo é a sub-região do cérebro, localizada no lobo temporal, responsável pelas memórias episódicas (LOMBROSO, 2004), relacionadas a eventos e experiências pessoais. Além dessa função, os neurônios dessa região também apresentam relacionamento com o deslocamento do animal (BEST, WHITE, MINAI, 2001).

Os locais no espaço, nos quais ocorrem potenciais de ação, são denominados campos de lugar (*place fields*), e as células que os ativaram são denominadas células de lugar (*place cells*).

Conforme a Figura B4 é possível avaliar os campos de lugar de acordo com a intensidade da atividade eletrofisiológica das células de lugar. Na Figura B4.(A), é ilustrado o

processo a partir do qual são adquiridas e processadas as informações relevantes para a definição dos campos de lugar e das células de lugar. Na Figura B4.(B), é apresentada uma matriz de possíveis espaços, nos quais o animal poderia percorrer. As cores de cada célula da matriz estão relacionadas à quantidade de potenciais de ação realizados por determinada célula de lugar.

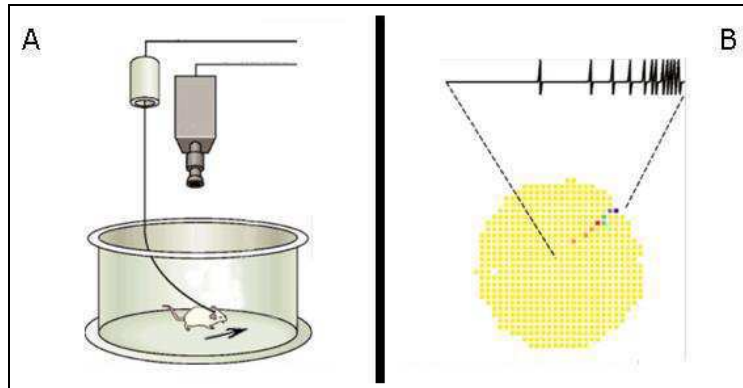


Figura B4: Ilustração de um experimento para análise dos campos de lugar com base na atividade a intensidade atividade eletrofisiológica das células de lugar. (A) Imagem de um rato explorando um ambiente cilíndrico, em que sua movimentação e atividade neuronal são registradas. (B) Resultado do processamento das informações adquiridas em (A). Adaptada de (KANDEL, SCHWARTZ, JESSELL, 2000).

A escolha de realizar a análise de padrões de ativação neuronal do hipocampo nesta Dissertação é justificada pelo fato dessa região ser responsável pelo tipo de memória que acredita-se estar presente durante o experimento comportamental que utilizado.

Na Figura B5, é ilustrada a posição do hipocampo no cérebro de um rato.



Figura B5: Ilustração da região hipocampo em ratos.  
Adaptada de (KANDEL, SCHWARTZ, JESSELL, 2000).

Uma das teorias também existentes quanto à função do hipocampo, é quanto ao suporte ao que se denominam mapas cognitivos (O'KEEFE, NADEL, 1978). Essa teoria sugere que o ambiente seria representado por uma matriz de células de lugar, de maneira que cada célula representa um local no qual o indivíduo possa estar.

Com isso, as configurações da ativação das células de lugar seriam utilizadas pelo indivíduo para orientação espacial. Na Figura B6 é possível visualizar uma ilustração dessa teoria.

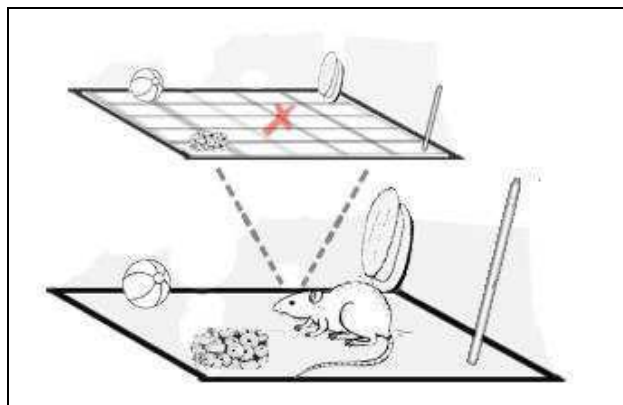


Figura B6: Ilustração da teoria dos Mapas Cognitivos. Adaptado de (EICHENBAUM *et al.*, 1999)

A consolidação da teoria dos Mapas Cognitivos recebeu a contribuição do trabalho de Brown *et al.* (1998), no qual foi utilizada a taxa de disparo de um conjunto de células de lugar para a predição das coordenadas do animal no plano.

### **Córtex Somatossensorial Primário**

A sub-região do cérebro, localizada no lobo parietal, especializada no processamento de informações táteis é o córtex somatossensorial primário.

O tato é um dos principais sentidos de seres vivos que exploram o ambiente em que estão inseridos, pois por ele são processadas informações da interação com objetos, tais como forma, textura, temperatura etc, e informações biológicas internas do indivíduo, tais como dor, movimento etc (BUTTS, 2009; BARHAM *et al.*, 2003). Sendo assim, um grande fornecedor de informações úteis para os indivíduos (FLANAGAN, LEDERMAN, 2001; ROBLES-DE-LA-TORRE, HAYWARD, 2001).

O córtex somatossensorial primário possui dinâmica de síntese de informação complexa, recebendo dados neuronais do sistema nervoso periférico, em especial, das células sensitivas, quanto por estimulação de diversas fontes externas (SQUIRE *et al.*, 2012).

Essa informação tem chamado a atenção de diversas pesquisas e desenvolvedores de aplicações de realidade virtual, e demonstram a importância do entendimento do sistema somatossensorial para a computação (FLANAGAN, LEDERMAN, 2001; ROBLES-DE-LA-TORRE, HAYWARD, 2001).

Em alguns mamíferos, como em roedores, além dos membros, são bastante utilizadas as vibrissas (pelos localizados em algum lugar da face) para a aquisição de informação tátil do ambiente.

Como o conjunto de padrões de ativação neuronal utilizado nesse trabalho de dissertação são originados de registros em ratos, a seguir, serão apresentados alguns detalhes do sistema somatossensorial desse animal.

Os ratos são animais noturnos acostumados a viver em locais escuros onde buscam comida e se protegem de predadores (PETERSEN, 2007), e utilizam, como mencionado anteriormente, vibrissas para amplificar sua aquisição de informações sobre o ambiente.

As vibrissas geram uma espécie de matriz altamente sensíveis ao toque para aquisição de informação tátil, com alta precisão da localização e discriminação de objetos (ARONOFF, LAYER, 2008; PETERSEN, 2007; KRUPA *et al.*, 2001).

Durante a exploração tátil desses animais, têm-se duas grandes funções corticais atuando: a função do processamento sensorial tátil, e a função motora responsável pelo movimento das vibrissas, o que permite um bom contexto para estudos dos aspectos neuronais de integração sensório-motor.

Na Figura B7, é apresentado um diagrama do fluxo de informação do sistema somatossensorial em ratos. As informações captadas pelas vibrissas são enviadas ao córtex somatossensorial primário, e as informações quanto ao movimento, são enviadas do córtex motor às vibrissas.

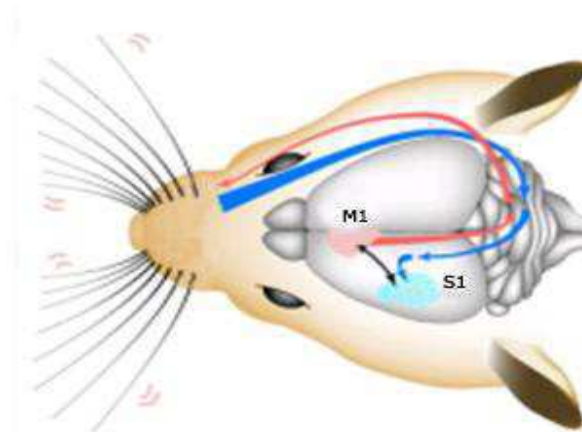


Figura B7: Diagrama do fluxo de informação do sistema somatossensorial de ratos.

Adaptado de (ARONOFF, LAYER, 2008).

### **Córtex Visual Primário**

A sub-região do cérebro, localizada no lobo occipital, especializada no processamento de informação visual é o córtex visual primário, também conhecida como área de Brodmann, área 17, ou ainda, como é denominada nesta Dissertação, V1.

A visão possui grande colaboração para, dentre outras modalidades, a formação das

memórias e compreensão do ambiente (KANDEL, SCHWARTZ, JESSELL, 2000), em todos os seres vivos, e realiza as funções de captação, registro e criação de representações, com base em experiências anteriores do indivíduo, da informação visual no mundo exterior.

Deste modo, é válida a concepção de que o processamento visual no cérebro é um processo dinâmico e ativo, e não estático e passivo (BREVEGLIERI *et al.*, 2008; GRILL-SPECTOR, MALACH, 2004; KANDEL, SCHWARTZ, JESSELL, 2000).

Um exemplo disso, são as imagens com ilusões de óptica. Na Figura B8, é ilustrado um casal à beira de um lago. Por se tratar de um casal, inconscientemente, nosso cérebro associa a imagem com o nosso conceito de família, o que favorece a sintetizarão e visualização de um bebê, formado pelos galhos de uma árvore.



Figura B8: Imagem com ilusão de óptica (Autor desconhecido).

O fluxo de informação do sistema visual humano é ilustrado na Figura B9. Na Figura B9.A, os estímulos do mundo externo são recebidos pelas retinas e enviados até o córtex visual primário, via as estruturas da Fóveas e da Periferia, as quais, primeiramente passam pelo Núcleo Geniculado Lateral, onde ocorre o cruzamento das vias. Quando a informação chega em V1, é distribuída para a via ventral, para o córtex inferior temporal, e para via dorsal que segue para o córtex parietal posterior, como é ilustrado na Figura B9.B.



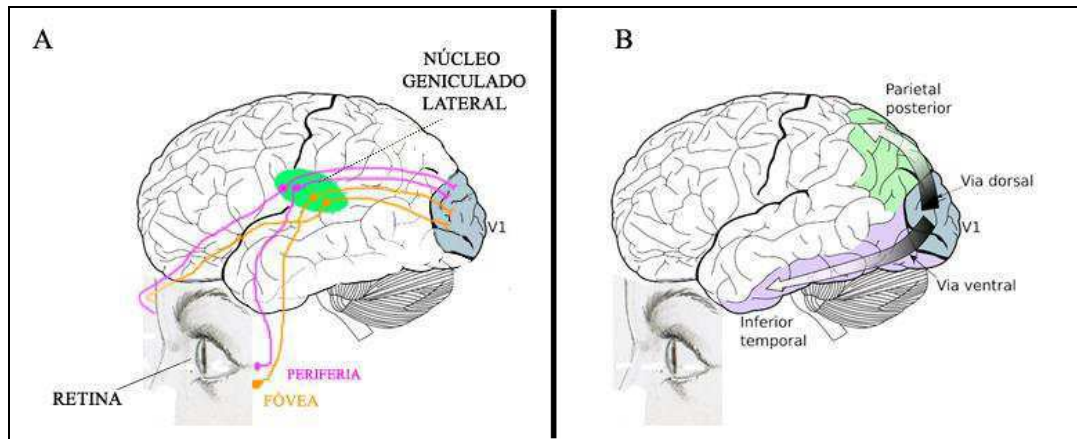


Figura B9: Dinâmica do sistema visual humano. (A) Recepção de estímulos visuais do mundo externo pela retina e envio ao V1. (B) Distribuição da informação do V1 pelas vias dorsal e ventral. Adaptada de (SQUIRE *et al.*, 2012).

O processamento realizado na via ventral tem como função a atividade de discriminação de objetos, enquanto que o processamento realizado na via dorsal tem como objetivo a identificação espacial. Em uma concepção simplificada, esses dois processamentos distribuídos são integrados para gerar a percepção completa do objeto visual, em seus diversos atributos (KANDEL, SCHWARTZ, JESSELL, 2000; MISHKIN, 1982; SQUIRE *et al.*, 2012).

Dentre as possíveis consequências de lesões na via dorsal, há a perda de resposta a objetos visuais no campo visual contralateral. Nessa via, encontra-se o córtex parietal posterior e o córtex temporal medial, cujas lesões podem ocasionar dificuldades no movimento dos membros para alcançar objetos (GRILL-SPECTOR, MALACH, 2004; KANDEL, SCHWARTZ, JESSELL, 2000; VAN ESSEN, GALLANT, 1994) e a perda da percepção de movimentos e profundidade, respectivamente.

Dentre as eventuais consequências de lesões na via ventral, mais especialmente no córtex inferior temporal, estão distúrbios relacionados à identificação de objetos e faces (HOFFMAN *et al.*, 2007; HUNG, 2005; KANDEL, SCHWARTZ, JESSELL, 2000).

## B.2.2 Neurônios

Os neurônios representam cerca de apenas 10% da massa encefálica de seres vivos que possuem sistema nervoso. No entanto, são responsáveis pela maior carga de processamento de informações em nosso organismo (BEAR, CONNORS, PARADISO, 2008). Nos seres humanos, há cerca de 86 bilhões de neurônios, possuindo cada um mais de 10 mil ligações estabelecidas entre eles (LENT, 2010). A estrutura do neurônio é ilustrada na Figura B10.

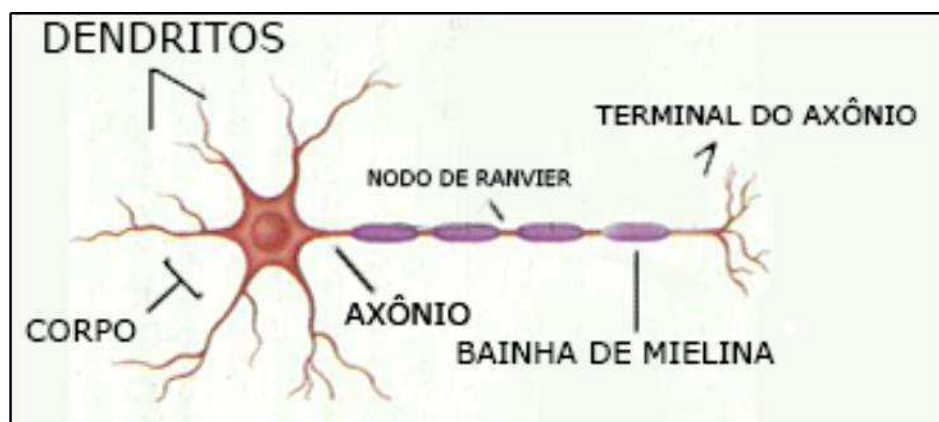


Figura B10: Ilustração estrutural de um neurônio. Adaptado de (LENT, 2010)

Neurônios são envolvidos por células da glia (também intituladas neuroglia), as quais possuem a função de fixação e suprimento neuronal.

Estruturalmente, os neurônios são constituídos pelas seguintes partes: Corpo celular (também conhecido como Soma), Dendritos e Axônio. O Soma engloba o conjunto de organelas existentes no neurônio, assim como o núcleo da célula (BEAR, CONNORS, PARADISO, 2008).

O núcleo é a parte central da célula onde o processamento proteico é realizado no neurônio.

Os Dendritos são os canais pelos quais são recebidos os sinais eletrofisiológicos, enviados por outros neurônios (GERSTNER, KISTLER, 2002).

O Axônio é parte da célula neuronal em que o sinal eletrofisiológico é processado e encaminhado aos terminais axoniais para ser enviado para outros neurônios, via sinapses. Esse componente se conecta ao Soma por uma estrutura denominada Cone de implantação, e

possui uma membrana quase completamente coberta por certas células gliais que estabelecem uma espessa camada isolante chamada bainha de mielina (LENT, 2010), que possibilita a condução ultrarrápida dos sinais elétricos produzidos pelos neurônios. Os trechos que não são cobertos pela bainha de mielina, são denominados Nodos de Ranvier (BEAR, CONNORS, PARADISO, 2008).

### **Assembleias**

As assembleias são subconjuntos de neurônios com comportamento de co-ativação significativa (HEBB, 1949), isto é, que possuem atividade eletrofisiológica correlacionada. Elas foram objetos de várias pesquisas realizadas entre as décadas de 80 e 90, nos conceitos de circuitos neuronais orientados (DEADWYLER, HAMPSON, 1995; DOUGLAS, KAC, 1991).

A questão relevante neste contexto é encontrar critérios que atribuam neurônios a uma determinada assembleia, que realiza uma codificação conjunta.

No modelo original de Hebb, as assembleias são diferenciadas entre si quando os neurônios de um subconjunto mostram ativação maior do que os demais neurônios de uma população. Porém, mais tarde, também foi introduzido o critério de disparo síncrono entre os neurônios para a formação das assembleias.

De modo mais geral, quaisquer conjuntos de neurônios em que se observa o comportamento coordenado de potenciais de ação, durante a realização da atividade de processamento de informação, podem formar uma população (de CHARMS, 1998).

Com base na Figura B11 e nas definições apresentadas anteriormente, é exemplificado, por análise visual que, dentre todos os seguintes 20 neurônios, uma possível assembleia foi formada entre os tempos 400 e 500ms, constituída pelos neurônios 1, 3, 7, 9, 14, 15 e 19.

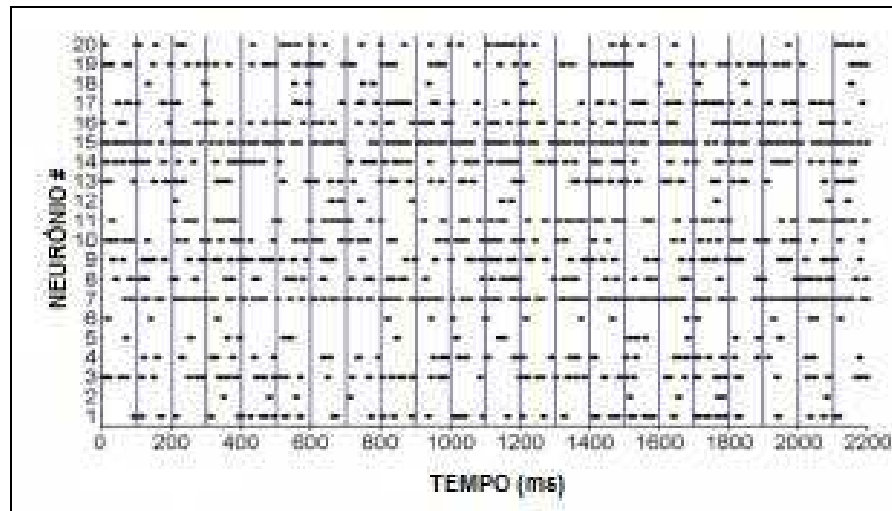


Figura B11: Disposição dos potenciais de ativação de neurônios pelo tempo. Adaptado de (LOPES-DOS-SANTOS *et al.*, 2011)

### Eletrofisiologia

É fundamental para os organismos a comunicação entre os seus neurônios, e entre neurônios e músculos (SQUIRE *et al.*, 2012).

Apesar de ser possível encontrar registros realizados pelo médico grego Claudius Galeno, nos anos 200 A.C, que sugeria que existia um fluxo de comunicação entre o cérebro e os músculos (SQUIRE *et al.*, 2012), foi apenas após a descoberta da eletricidade que se tornou possível o estudo do funcionamento dos sistemas vivos do ponto de vista elétrico (também denominada como Eletrofisiologia).

Na área da Neurociência, uma das pesquisas que trabalhou com esse foco foi a realizada por Luigi Galvani (1737-1798), a partir da qual foi demonstrado que músculos e células nervosas poderiam produzir eletricidade (FINGER, 2001; SQUIRE *et al.*, 2012).

Outro importante estudo sobre a eletrofisiologia do sistema nervoso foi realizado por Julius Bernstein (1839-1917), que resultou na definição do impulso nervoso com uma onda de negatividade, baseada em um conceito, proposto pelos pesquisadores Carlo Matteucci (1811-1868) e Emil du Bois-Reymond (1818-1896), intitulado "Variação negativa", observada por meio do galvanômetro, o qual apresentou um decréscimo de corrente elétrica, durante a contração no músculo (FINGER, 2001; NICOLELIS, 2007), fenômeno que os experimentos de monitoramento da atividade eletrofisiológica de ratos na exploração de objetos, realizados nessa pesquisa, também estavam sujeitos.

Com os avanços das técnicas da eletrofisiologia foi possível o aumento da precisão na aferição da atividade elétrica do sistema nervoso (FINGER, 2001; SQUIRE *et al.*, 2012), possibilitando o estudo da atividade de um único neurônio ou de populações de neurônios (BUZSAKI, 2004).

### **Potencial de Ação/ Spike**

Atualmente, ainda não está completo o entendimento de como é processada a informação no cérebro. No entanto, há um considerável consenso sobre um elemento fundamental de tratamento de informação neste contexto: o neurônio. Pesquisadores acreditam que o processamento de informação realizado pelo cérebro pode ser estudado do ponto de vista da dinâmica das propriedades elétricas (RIEKE *et al.*, 1999).

Conforme Gerstner e Kistler (2002):

“Os sinais neuronais consistem de pulsos elétricos curtos que podem ser observados colocando um fino eletrodo perto do soma ou do axônio do neurônio (...) Os pulsos são então chamados potenciais de ação ou *spikes*, têm uma amplitude em torno de 100mV e uma duração típica de 1-2 ms. A forma do pulso não muda quando este se propaga ao longo do axônio”

Desta forma, a informação de saída é apresentada no axônio por efeito de um potencial, denominado potencial de ação.

Na Figura B12, pode ser vista uma ilustração da coleta da informação do potencial de ação em um neurônio.

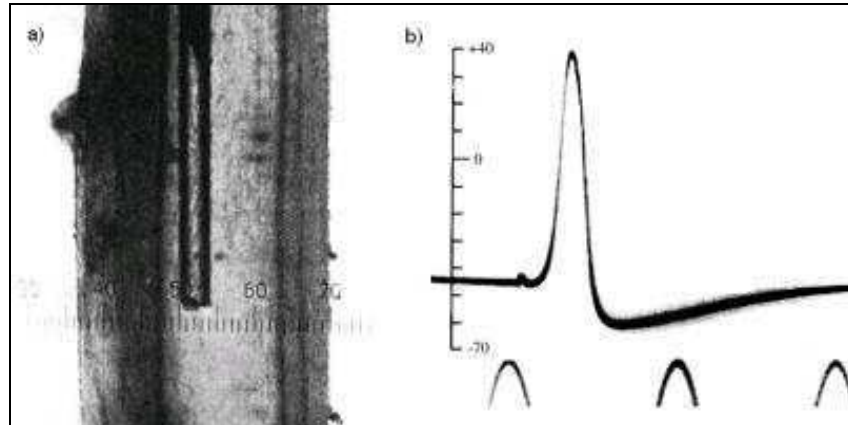


Figura B12: Ilustração da coleta do sinal do potencial de ação intracelular no neurônio. (A) Micropipeta de vidro, de aproximadamente 100 $\mu$ m de diâmetro inserida no axônio de uma lula. (B) Imagem referente ao potencial de ação gravado no axônio. Fonte: (HODGKIN, HUXLEY, 1939).

Na Figura B13, é apresentada a esquematização da curva do potencial de ação. Para o registro do valor do potencial de ação no neurônio tornam-se necessários o uso de técnicas de registro eletrofisiológicas, invasivas ou não à célula. Técnicas invasivas são mais precisas do que técnicas não invasivas, porém podem lesionar a célula.

Um exemplo de técnica invasiva, utilizada nesse contexto é a implantação de matrizes de microeletrodos. Nesse caso, é necessário posicionar os microeletrodos de forma a medir a diferença de potencial entre a parte interna e externa da célula.

Nos momentos anteriores às produções dos potenciais de ação, ocorre a polarização positiva no interior da célula, ocorrendo pela entrada de íons com carga positiva no citoplasma da célula, tornando seu interior mais positivo que a parte externa da célula, até atingir um pico de aproximadamente 40 mV, quando esta diferença de potencial cai e torna-se mais negativa que na situação de equilíbrio, voltando depois ao valor de equilíbrio inicial, em torno de -70 mV, este momento é denominado período refratário (BEAR, CONNORS, PARADISO, 2008; BERGLES *et al.*, 1996). Quando o pico do potencial de ação ultrapassa determinado potencial limiar, diz-se a que célula disparou ou ativou.

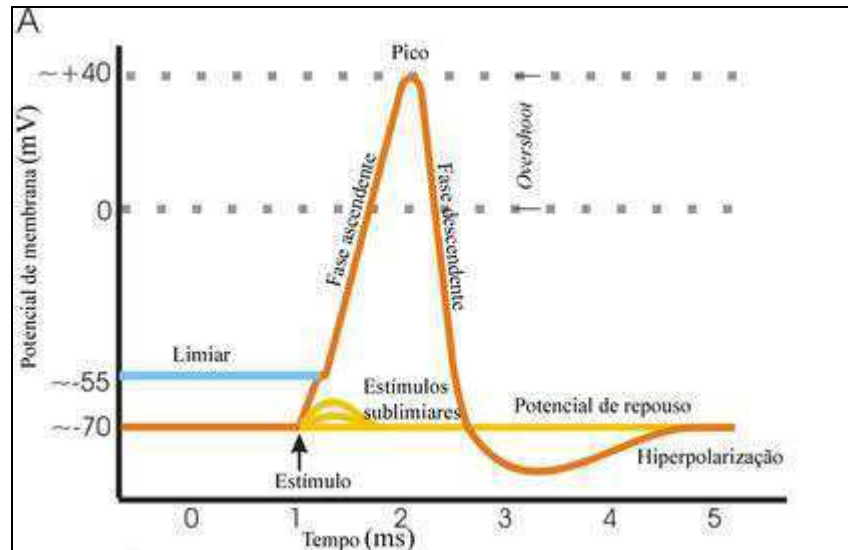


Figura B13: Ilustração da curva do potencial de ação. Adaptada de Wikimedia ([http://upload.wikimedia.org/wikipedia/commons/4/41/Potencial\\_de\\_acao.png](http://upload.wikimedia.org/wikipedia/commons/4/41/Potencial_de_acao.png)).

De acordo com a Figura B13, é possível perceber que todo o processo, entre o momento de início da polarização e o instante que o potencial volta ao valor de equilíbrio, e o período refratário, dura aproximadamente 4ms.

Diversos trabalhos têm sido realizados a fim de estudar o comportamento do potencial de ação, com intuito de analisar mais informações e remover ruídos dos registros (MARK *et al.*, 2006).

### B.2.3 Bin

No contexto desta pesquisa, o termo *bin* significa um intervalo de tempo escolhido para a contagem de ocorrências de potenciais de ação de um determinado neurônio. O tamanho do *bin* é uma característica importante para determinar a configuração dos padrões de ativação (TAMURA *et al.*, 2012).

Para auxiliar na consolidação do conceito, é ilustrado um exemplo, com base na Figura B14, a qual apresenta um gráfico da atividade potencial de membrana de um neurônio ao longo do tempo.

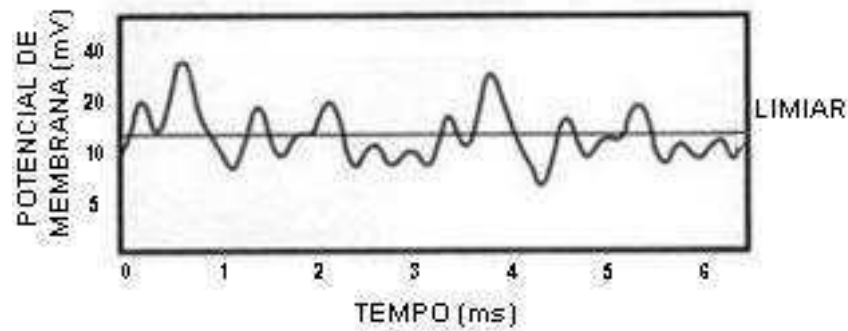


Figura B14: Gráfico da atividade potencial de membrana de um neurônio.

Se for determinado um bin de duração de 2 ms, é possível encontrar com base na figura anterior, 3 bins (0-2ms, 2-4ms e 4-6ms). No primeiro bin, é possível contar 3 potenciais de ação, 3 potenciais de ação no segundo bin e no último bin apenas 2.

#### B.2.4 Janela

O conceito de Janela é bastante utilizado na área da Neurociência, na realização da atividade de segmentação de matrizes de representação da atividade neuronal em um intervalo de tempo, pois determina o número de bins que será utilizado na construção dos padrões de ativação neuronal.

Esse conceito é ilustrado na Figura B15, na qual podemos ver a matriz A, em que nas linhas são listados N neurônios os quais estão sendo analisados as ativações, nas colunas são dispostos M instantes de tempo, e nas células (n,m) são sinalizadas a ocorrência de potenciais de ação, e a janela W, uma sub-matriz de A de tamanho L.



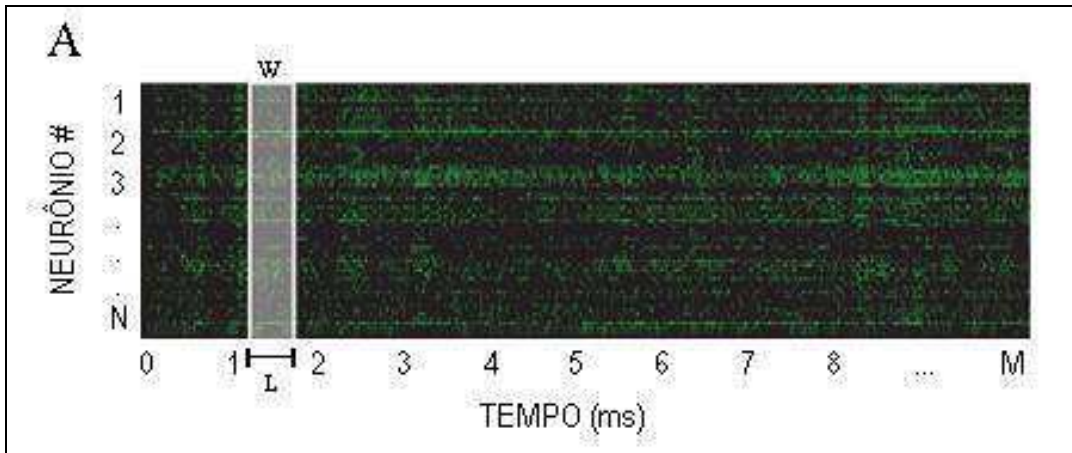


Figura B15: Ilustração de uma janela  $W$  de largura  $L$ , na matriz  $A$  (disposição de potenciais de ação de  $N$  neurônios pelo tempo). Adaptado de (RIBEIRO *et al.*, 2007)

### B.2.5 Métodos de Aquisição de Dados Neurais

Atualmente, existe uma grande variedade de métodos de aquisição de dados neuronais, os quais podem ser classificados como invasivos ou não invasivos (LENT, 2010). Os métodos invasivos consistem na aquisição de determinadas características cerebrais de um indivíduo diretamente no alvo do estudo, ou seja, é necessária intervenção cirúrgica para o uso desses métodos.

Os métodos não invasivos permitem a aquisição sem intervenção cirúrgica. Esses métodos reduzem o impacto da atividade de coleta sobre o comportamento estudado, contudo, possuem menor precisão do que os métodos invasivos (SCHLOGL, BRUNNER, 2008).

Entre os métodos não invasivos, se destacam uso das tecnologias Magnetoencefalografia (MEG) (IOANNIDES, TAYLOR, 2003), Ressonância Magnética (LEE *et al.*, 2007), Espectroscopia de Infravermelho Próximo (COPE *et al.*, 1991) e Eletroencefalografia (EEG) (KREPKE, 2004), entre outras (AKEMANN *et al.*, 2010).

Entre os métodos invasivos, a matriz de microeletrodos é uma das tecnologias mais utilizadas, a qual, por ter sido utilizada para a aquisição dos dados analisados neste trabalho de Dissertação, foi descrita em maior detalhe na Seção B.2.6.

## B.2.6 Matrizes de Microeletrodos

Uma das ferramentas utilizadas para realização do registro da atividade eletrofisiológica intra e extracelular de um conjunto de neurônios é a matriz de microeletrodos.

Atualmente, matrizes de microeletrodos são capazes de registrar a atividade eletrofisiológica de mais de 100 neurônios ao mesmo tempo (MATSUMOTO, 2010).

Os registros são captados a partir de eletrodos de espessura microscópica (em geral, aproximadamente  $35\mu\text{m}$ ), dispostos em arranjos geométricos de forma variada, e implantados em regiões específicas do sistema nervoso (NICOLELIS *et al.*, 1997; NICOLELIS, 2007).

Na Figura B16, pode ser visualizada uma foto onde é apresentada uma matriz de microeletrodos, na qual é possível observar, tomando como referência uma moeda, o tamanho do dispositivo, e o implante da mesma no cérebro de um humano.

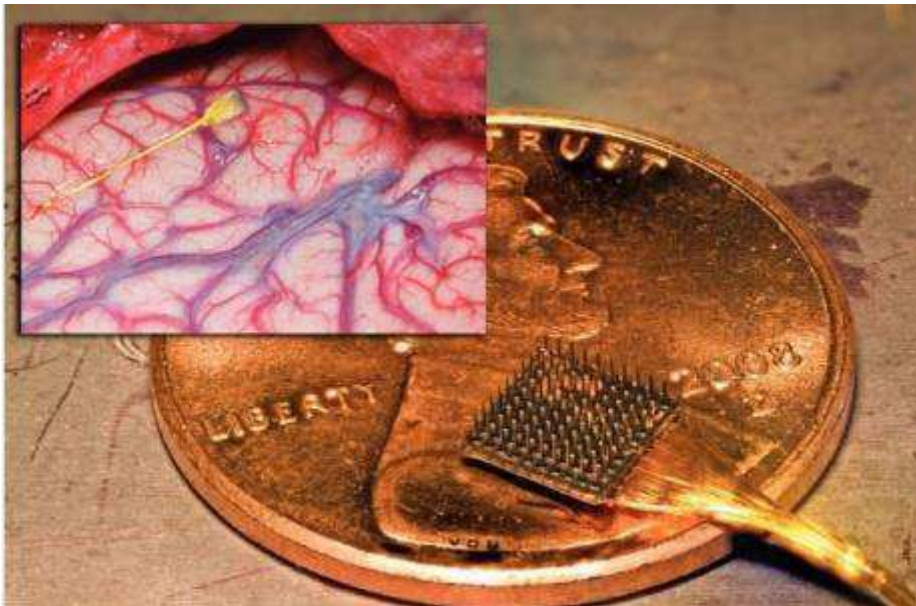


Figura B16: Fotografia do implante das matrizes microeletrodos. Fonte: CITYWEEKLY (<http://www.cityweekly.net/utah/article-9318-brain-chip-may-help-the-blind-see.html> - Acessado em 26/02/2014).

### B.2.7 Padrão de Ativação Neuronal

Padrões de ativação neuronal são regularidades na atividade neuronal de um indivíduo, cuja aparência dá suporte a uma informação específica sobre a tarefa que um neurônio ou grupos de neurônios desempenham (CATTANEO, MAFFEI, MORRONE, 1981).

Pesquisadores supõe que os padrões devem ajudar a explicar a codificação ou o cálculo executado pelo sistema. Por exemplo, os padrões de ativação podem expressar a mensagem a ser codificada, a assinatura do mecanismo de geração eletrofisiológica ou até mesmo alterações que afetam o sistema (BROWN, KASS, MITRA, 2004).

Um padrão é uma descrição quantitativa ou estrutural de um objeto, ou alguma outra entidade de interesse. Donde segue que uma classe de padrões é definida como um conjunto de padrões que compartilham alguma propriedade em comum (BOW, BOW, 2002).

Os eventos que formam um padrão estão dispostos ao longo de um eixo de tempo. Eles também podem ter um caráter espacial, no sentido de que os eventos são construídos a partir de ativações provenientes de diferentes neurônios.

A construção de um padrão requer a criação de experiências nas quais aparecem regularidades que possam ser definidos em parâmetros relevantes para a discriminação de cada padrão.

## *B.3 Metodologia Científica*

Uma Metodologia Científica pode ser definida como um conjunto de abordagens, técnicas e processos utilizados pela ciência para formular e resolver problemas de aquisição objetiva do conhecimento, de uma maneira sistemática (WAZLAWICK, 2009; LAKATOS, MARCONI, 2010).

Em geral, uma metodologia científica é composta e descrita pelo tipo de estudo, variáveis, fatores, projeto do experimento e análise estatística (WAZLAWICK, 2009; LAKATOS, MARCONI, 2010), utilizados em toda a realização dessa pesquisa de Dissertação, os quais serão descritos nas seções seguintes.

### B.3.1 Tipos de Estudo

Dentre os possíveis tipos de estudos que podem ser realizados seguindo uma metodologia científica, existem os estudos Transversais, Longitudinais (também definidos como de Coorte) e de Caso Controle (WAZLAWICK, 2009; LAKATOS, MARCONI, 2010).

Os estudos Transversais são caracterizados pela análise de um determinado estado de uma população em um específico momento.

Os estudos Longitudinais têm como objetivo analisar o progresso de um indivíduo durante um determinado intervalo de tempo.

Por fim, os estudos de Caso Controle são definidos como trabalhos que analisam uma determinada amostra (caso), na qual tenha sido realizado algum tipo de intervenção, com relação à outra de referência (controle).

Neste trabalho de dissertação, foi realizado um estudo do tipo Longitudinal, pois foram analisados os progressos (potenciais de ação) de ratos na exploração de objetos ao longo de um determinado intervalo de tempo.

### B.3.2 Fatores

Fatores são características ou atributos de um experimento que podem ser identificados como influenciadores do resultado (variável de resposta), a qual é analisada em um determinado experimento, na investigação dos testes de hipóteses, como por exemplo, o tempo, a temperatura, o número de indivíduos utilizados na pesquisa, especificações de um computador etc. (WAZLAWICK, 2009; LAKATOS, MARCONI, 2010).

Quando um fator só assume um determinado valor, dar-se o nome de Constante, como por exemplo, a velocidade da luz ou a gravidade, em determinados contextos (LAKATOS, MARCONI, 2010).

Quando os valores de um determinado fator podem ser configurados ou mudam de forma espontânea ao longo do tempo, estes fatores são definidos como variáveis. As variáveis também podem ser subdivididas em dependentes e independentes, caso os valores por elas assumidos levem em consideração outros fatores ou não, respectivamente

(LAKATOS, MARCONI, 2010).

As variáveis configuradas de acordo com um objetivo de determinado experimento são chamadas de variáveis de controle, e a variável dependente, cujo valor é analisado ao término de cada experimento é chamada variável de resposta (LAKATOS, MARCONI, 2010).

Outra classificação para variáveis está na elegibilidade dos valores por ela assumidos, no qual pode ser qualitativa, onde os valores não possuem uma ordem ou hierarquia, ou quantitativa, caso contrário. (LAKATOS, MARCONI, 2010).

As variáveis também podem ser especificadas quanto ao conjunto de valores que elas podem assumir, sendo ditas discretas, quando o conjunto de valores por elas assumido é finito ou enumerável, e contínuas, quando o conjunto de valores formam um intervalo ou união de números reais.

Cada valor que uma variável pode assumir é contabilizado como níveis da variável, como por exemplo, digamos que uma variável pode assumir os valores inteiros de 1 a 10, logo, podemos dizer que essa variável possui 10 níveis. (LAKATOS, MARCONI, 2010).

Sendo  $n$  o número de fatores, ao vetor,  $x = [a_1; a_2; \dots; a_n]$ , no qual cada um de seus componentes, guarda um valor assumido do  $i$ -ésimo fator, dá-se o nome de vetor de atributos, ou vetor de características, ou ainda padrão (POLIKAR, 2006).

Cada um destes vetores identifica unicamente um objeto ou padrão. Quando referente à configuração de um experimento, usualmente os vetores são definidos como tratamentos.

### 2.3.5 Projeto do Experimento

No projeto do experimento, são registrados todos os passos e atividades que deverão ser realizadas para que o resultado de uma determinada pesquisa seja reencontrado.

Nele também são descritos os tratamentos que serão analisados durante uma pesquisa.

A quantidade de tratamentos que uma pesquisa irá analisar pode defini-la como Fatorial Completa ou Fatorial Fracionária.

A Fatorial Completa tem como objetivo analisar todos os níveis de um conjunto

particular de fatores.

A Fatorial Fracionária visa o estudo apenas de alguns níveis de um conjunto particular de fatores.

### B.3.6 Análise Estatística

A análise estatística permite obter, de conjuntos complexos, representações simples e constatar se essas verificações simplificadas têm relações entre si. Assim, o método estatístico objetiva a redução de fenômenos sociológicos, políticos, econômicos etc., a termos quantitativos e a manipulação estatística que permitem comprovar as relações dos fenômenos entre si, e obter generalizações sobre sua natureza, ocorrência ou significado (LAKATOS, MARCONI, 2010).

Entretanto, a estatística pode ser considerada mais do que apenas um meio de descrição racional; é, também, um método de experimentação e prova, pois é método de análise (LAKATOS, MARCONI, 2010).

A seguir, serão apresentados os testes e fatores estatísticos que foram analisados para dar suporte às inferências realizadas nesse trabalho de Dissertação.

#### **Média Aritmética**

Também definida como “valor esperado”, é valor que indica o ponto de equilíbrio de uma frequência ou distribuição.

Matematicamente, seja  $n$  o número de elementos analisados, a média aritmética ( $\mu$ ) pode ser definida pela Equação (7).

$$\mu = \frac{1}{n} \sum_{i=1}^n a_i \quad (7)$$

### Variância

Mensura o nível de dispersão estatística de uma variável, indicando quão longe, em geral, os seus valores se encontram da média.

Se  $\mu$  é a média da variável  $X$ , então a variância de  $X$ ,  $var(X)$ , pode ser definida matematicamente pela Equação (8).

$$var(X) = E((X - \mu)^2) \quad (8)$$

### Desvio Padrão

É a medida de dispersão estatística correlacionada à variância de uma variável.

Matematicamente, o desvio padrão, representada pelo símbolo sigma ( $\sigma$ ), define-se como a raiz quadrada da variância, conforme a Equação (9).

$$\sigma = \sqrt{var(X)} \quad (9)$$

### QQ-plot

QQ-plot é uma representação gráfica para comparação entre duas distribuições de probabilidades utilizando valores que dividem um conjunto ordenado de dados em quatro partes iguais, definidos como quartis.

O gráfico é construído a partir do intervalo escolhido para os quartis. Um ponto  $(x, y)$  na curva corresponde ao quartil da segunda distribuição (coordenada  $y$ ) em função do mesmo quartil da primeira distribuição (coordenadas  $x$ ). Assim, a linha representa uma curva paramétrica para o intervalo de percentil. (WILK, GNANADESIKAN, 1968).

Se as duas distribuições comparadas são semelhantes, os pontos do QQ-plot serão aproximadamente a linha  $y = x$ .

Na Figura B17, é ilustrado um exemplo de QQ-plot.

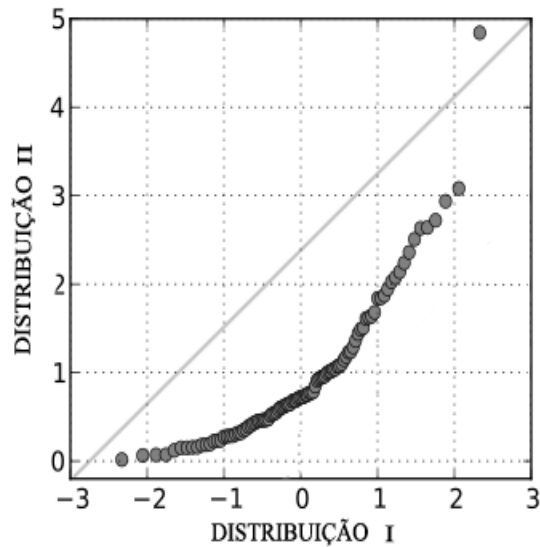


Figura B17: Exemplo de QQ-plot. Adaptada de Wikimedia ([http://upload.wikimedia.org/wikipedia/commons/1/11/Normal\\_exponential\\_qq.svg](http://upload.wikimedia.org/wikipedia/commons/1/11/Normal_exponential_qq.svg) - Acessado em 17/04/2014 ).

### ***Boxplot***

*Boxplot* é uma ferramenta gráfica, também denominada Diagrama de Caixas ou Diagrama de Whisker, que possibilita a análise de distribuições de dados usando o conceito de quartis.

Dado um conjunto de dados ordenados, o 1º quartil representa 25% dos dados e Q1 o valor maior dessa sequência, o 2º quartil representa 50% dos dados e Q2 a mediana dos valores dos dados, o 3º quartil representa 75% dos dados e Q3 o valor maior dessa sequência, e o 4º quartil é o conjunto inteiro dos dados e Q4 o maior valor dentro desse conjunto.

Na Figura B18 é ilustrado 4 *boxplots* ladeados.



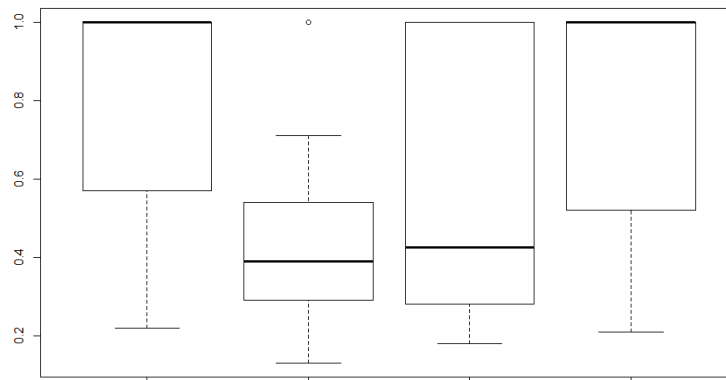


Figura B18: Exemplo de Boxplots.

Os lados inferiores e superiores dos retângulos dos *Boxplots* representam o 1º e o 3º quartil, respectivamente. A linha do interior é o 2º quartil. As linhas nas extremidades simbolizam os limites superiores e inferiores.

Valores acima do limite superior e abaixo do limite inferior, são sinalizados por pequenos círculos. Esses valores são definidos como ruídos, ou valores que se distinguem bastante de suas populações (*outliers*).

### Distribuição Normal (Normalidade)

A distribuição normal (também conhecida como Distribuição Gaussiana, ou ainda como curva em forma de sino) é uma função de probabilidade contínua, comumente encontrada na ocorrência de um evento.

Esse tipo de distribuição é bastante utilizado na modelagem de variáveis aleatórias reais, cujas distribuições não são conhecidas, devido ao teorema do limite central, que estabelece que a distribuição dos valores de uma variável aleatória independente é aproximadamente normal.

Matematicamente, a distribuição normal pode ser definida de acordo com a Equação (10).

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (10)$$

De maneira que a curva gerada por essa função é semelhante à ilustrada na Figura B19.

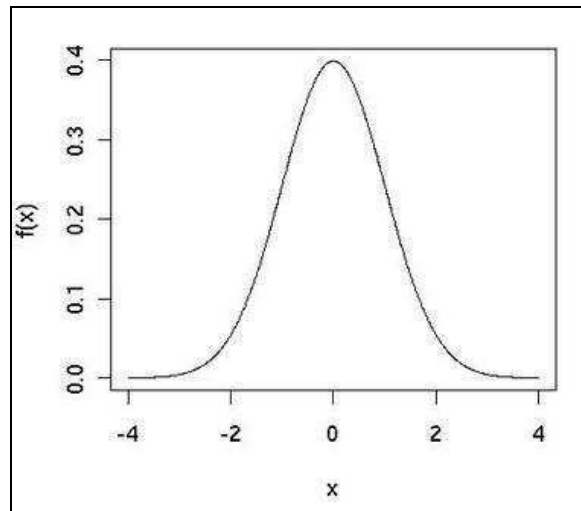


Figura B19: Exemplo de gráfico de uma Distribuição Normal

### Teste de Shapiro-Wilk

O teste de Shapiro-Wilk (1965) analisa se determinada amostra veio de uma população normalmente distribuída – hipótese nula.

A expressão matemática do teste é definida pela Equação (11):

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (11)$$

Em que:  $X$  é uma amostra de  $n$  elementos e  $X(i)$  é o elemento de posição  $i$  dentro da amostra;  $\bar{x}$  é a média dos elementos da amostra  $X$ ; e  $a_1$  é uma constante, definida pela Equação (12):

$$(a_1, \dots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{\frac{1}{2}}} \quad (12)$$

Que por sua vez:  $m$  é o vetor transposto de  $n$  valores esperados de uma ordem estatística de uma variável aleatória independente e identicamente distribuída, amostrada de uma distribuição normal padrão; e  $V$  é a matriz de covariância da ordem estatística.

Se  $W$  possuir um valor muito pequeno, podemos rejeitar a hipótese nula, nesse caso, a amostra não viria de uma distribuição normal.

### Teste de Bartlett

O teste de Bartlett (1937) usado para testar se  $k$  amostras são de populações com variâncias iguais (ou homogêneas).

Matematicamente, o teste pode ser definido pela Equação (13).

Em que:  $k$ : Número de distribuições;  $X_i$ : Distribuição de index  $i$ ; e  $QMRE$ : Quadrado Médio Residual.

$$H = \frac{(X-k) \log QMRE - \sum_{i=1}^k (X_i - 1) \log var(X)}{1 + \frac{1}{3(k-1)} \left[ \sum_{j=1}^k \frac{1}{X_j} - \frac{1}{X-k} \right]} \quad (13)$$

Com isso, as distribuições são ditas de variância iguais se  $H$  não for maior que o valor crítico da cauda superior da distribuição  $X^2_{k-1}$ , isto é, o nível de diferenciação entre as variâncias é estatisticamente não significativo.

### Teste de Kruskal-Wallis

O teste de Kruskal-Wallis (1952) tem como objetivo comparar várias amostras independentes por meio de uma análise de variância por postos.

Este teste deve ser usado considerando-se duas ou mais amostras aleatórias independentes e ordinais, para que seja possível atribuir posições.

Esse teste não pode ser usado para testar diferenças em amostras pareadas.

É uma análise da variância que emprega posições (soma de filas) em lugar de mensurações como critério de avaliação.

A variável básica deve ter distribuição contínua e o tamanho mínimo de cada amostra deve ser 6. Nesse casos, a significância do teste pode ser determinada pela Tabela do Qui-quadrado (BOX, HUNTER, HUNTER, 2005).

Matematicamente, o teste de Kruskal-Wallis é definido de acordo com a Equações (14).

Em que:  $N$ : número total de observações;  $K$  : número de amostras;  $n_j$  : número de observações na  $j$ -ésima amostra;  $R_j$  : soma dos postos da  $j$ -ésima amostra; e  $X_{critico}^2(\alpha, gl = k - 1)$  : Valores críticos da Tabela do Qui-quadrado.

$$H = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{(R_j)^2}{n_j} - 3(N + 1) \quad (14)$$

Se  $H > X_{critico}^2$ , então as amostras pertencem à mesma distribuição.