

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA E ESTATÍSTICA

Anderson David de Souza Campelo

**Análise de dados de Imagem por Ressonância Magnética
Funcional utilizando Mapas auto-organizáveis**

Belém

2010

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA E ESTATÍSTICA

Anderson David de Souza Campelo

**Análise de dados de Imagem por Ressonância Magnética
Funcional utilizando Mapas auto-organizáveis**

Dissertação apresentada ao Programa de Pós-graduação em Matemática e Estatística - PPGME - da Universidade Federal do Pará, como pré-requisito para obtenção do título de Mestre em Matemática.

ORIENTADOR: Prof. Dr. Valcir João da Cunha Farias

Belém

2010

Campelo, Anderson David de Souza

Análise de dados de Imagem por Ressonância Magnética Funcional utilizando Mapas auto-organizáveis/(Anderson David de Souza Campelo); orientador, Valcir João da Cunha Farias. - 2010.

61 f. il. 28 cm

Dissertação (Mestrado) - Universidade Federal do Pará. Instituto de Ciências Exatas e Naturais. Programa de Pós-graduação em Matemática e Estatística. Belém, 2010.

1. Redes Neurais. 2. Entropia - Teoria da Informação. 3. Imagem por ressonância Magnética. I. Farias, Valcir João da Cunha, orient. II. Universidade Federal do Pará, Instituto de Ciências Exatas e Naturais, Programa de Pós-Graduação em Matemática e Estatística. III. Título.

CDD 22. ed. 006.32

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA E ESTATÍSTICA

Anderson David de Souza Campelo

Análise de dados de Imagem por Ressonância Magnética Funcional utilizando Mapas auto-organizáveis

Dissertação apresentada ao Programa de Pós-graduação em Matemática e Estatística da Universidade Federal do Pará, como pré-requisito para obtenção do título de Mestre em Matemática.

Data da defesa: 18 de fevereiro de 2010.

BANCA EXAMINADORA

Prof. Dr. Valcir João da Cunha Farias (Orientador)
PPGME - UFPA

Prof. Dr. Marcus Pinto da Costa da Rocha (Co-orientador)
PPGME - UFPA

Prof. Dr. Heliton Ribeiro Tavares
PPGME - UFPA

Prof. Dr. Antônio Pereira Júnior
UFRN

À minha família.

Agradecimentos

- Agradeço primeiramente a Deus, por ter me dado a vida e a oportunidade de estudar.
- Aos meus pais que transformaram essa oportunidade em realização, através de seus incentivos e carinho.
- A Gilmara, pessoa que esteve presente todos esses anos desde os tempos de graduação tornando a vida mais agradável.
- Agradeço aos professores e amigos Marcus Rocha e, particularmente, Valcir Farias que como co-orientador e orientador, dedicaram tempos para discussões, aprendizados e sem os quais esta dissertação não seria possível.
- A todos do *LAM*, Alessandra, André, Daniel, Edinaldo, Josiane, Liliane, Luís, Marcos, Nayane, Phillip e Ulisses pelo companheirismo, momentos de discussão e apoio ao trabalho.
- Agradeço ainda, aos outros membros da banca, professores Heliton e Antônio Pereira, que se dispuseram a ler criticamente o texto e auxiliar em possíveis dúvidas correntes.
- Ao Programa de Pós-Graduação em Matemática e Estatística da Universidade Federal do Pará de modo geral, pelo apoio de todos os professores que ainda que indiretamente, tenham contribuído com valiosos ensinamentos.
- A CAPES, pela concessão da bolsa de mestrado e pelo apoio financeiro para a realização desta pesquisa.

Lista de Abreviaturas

BOLD	Contraste dependente do nível de oxigenação sanguínea (<i>Blood Oxygenation Level Dependent Contrast</i>)
EPI	Echo-planar imaging
fMRI	Imagem por ressonância magnética funcional (<i>functional magnetic resonance imaging</i>)
FWHM	Largura no ponto de meia-altura (<i>full-width at half maximum</i>)
GLM	Modelo linear geral (<i>general linear model</i>)
HC	Agrupamento Hierárquico (<i>hierarchical clustering</i>)
HRF	Função de resposta hemodinâmica (<i>hemodynamic response function</i>)
ICA	Análise de componentes independentes (<i>independent component analysis</i>)
MR	Ressonância Magnética (<i>magnetic resonance</i>)
PCA	Análise de componentes principais (<i>principal component analysis</i>)
RF	Rádio-freqüência
RNA	Redes Neurais Artificiais
SNR	Relação sinal-ruído (<i>signal-to-noise ratio</i>)
SOM	Mapas auto-organizáveis (<i>self-organizing maps</i>)

Lista de Figuras

1.1	(a) Neurônio biológico; (b) Neurônio artificial; (c) Esquema ilustrativo de uma RNA.	4
1.2	Diferentes arquiteturas de mapas auto-organizáveis: (a) mapa unidimensional; (b) mapa bidimensional retangular.	5
1.3	As três distâncias da <i>U-matrix</i> . Imagem retirada e adaptada a partir de Costa & Netto, 2007.	11
1.4	(a) Dendrograma; (b) Diagrama de agrupamentos.	14
1.5	Entropia no caso de duas possibilidades com probabilidades p e $1 - p$	16
2.1	Exemplo de orientação de três diferente cortes: (a) Fatia Coronal; (b) Fatia Sagital; (c) Fatia Axial.	18
2.2	Função Resposta Hemodinâmica para um estímulo de curta duração. Imagem retirada e adaptada a partir de Mulert & Lemieux, 2010	20
2.3	Experimentos em fMRI: (a) Paradigma em Bloco, quatro períodos de repouso intercalados com três de ativação; (b) Um paradigma misto, paradigma em bloco interpretado como evento-relacionado; (c) um paradigma evento-relacionado aleatório. Imagem retirada de Sarty, 2007.	21
2.4	Realinhamento espacial de um experimento auditivo. Na legenda, os 6 parâmetros de corpo rígido, para translação e rotação. Figura obtida após a etapa de realinhamento realizada no SPM8 (Statistical Parametric Mapping).	22
2.5	Série temporal de voxels ativos e não ativos de uma única fatia dos volumes de imagens. Imagem retirada e adaptada a partir de Sarty, 2007.	24
2.6	(a) Representação de um paradigma em bloco, o valor 0 indica o repouso e 1 a ativação; (b) HRF modelada segundo a proposta de Friston (1998), usando parâmetros de Glover (1999); (c) Convolução da HRF com o paradigma em bloco.	26

2.7	Diagrama descritivo do cálculo da entropia de Shannon. As probabilidades podem ser obtidas sobre cada janela separadamente (W_1 e W_2) e para cada nível I_k . Imagem retirada de Araújo et al. 2003.	27
3.1	Esquema ilustrativo da distribuição espacial de voxels ativos e a intensidade destes voxels ao longo do tempo em dados sintéticos com SNR = 5dB	29
3.2	Séries temporais para diferentes valores de SNR. Para valores altos de SNR, a HRF é mais evidente.	30
3.3	(a) Sinal apresentando baixa entropia, os pontos concentram-se no intervalo superior, com as seguintes distribuição de probabilidade de cada intervalo, $p(I_1) = 0.1083$ e $p(I_2) = 0.8917$ e entropia $H = 0.4949$; (b) Sinal apresentando alta entropia, com distribuição de probabilidade para os intervalos $p(I_1) = 0.4667$ e $p(I_2) = 0.5333$ e entropia $H = 0.9968$	33
4.1	(a) Mapa entrópico dos dados artificiais; (b)Análise entrópica dos dados sintéticos. Os pontos pretos na imagem foram eliminados, o equivalente a 6% dos dados de entrada.	35
4.2	(a) Resultado do método de agrupamento hierárquico aplicado aos pesos do SOM, 72 agrupamentos foram formados; (b) Visualização tridimensional da <i>U-matriz</i> dos dados sintéticos.	36
4.3	Configuração final do SOM, grade 10×10 para os dados sintéticos, o neurônios em destaque representam os padrões ativos dos dados de entrada.	36
4.4	O sinal em azul corresponde a HRF com SNR de 5dB atribuída aos dados sintéticos. O sinal em vermelhos representa a média da série temporal dos neurônios candidatos a ativação.	37
4.5	As regiões pretas no interior do cérebro correspondem às regiões ativas computadas pelo SOM com 100 iterações para um $CC > 0.7056$	37
4.6	(a) Mapa entrópico dos dados reais da oitava fatia; (b)Análise entrópica dos dados reais. Os pontos pretos na imagem foram eliminados, o equivalente a 8% dos dados de entrada.	38
4.7	Gráfico do erro de quantização para um total de 250 iterações. (a) erro de quantização com aplicação de entropia; (b) erro de quantização sem entropia.	38
4.8	(a)Resultado do método de agrupamento hierárquico aplicado aos pesos do SOM, 49 agrupamentos foram formados; (b) Visualização tridimensional da <i>U-matriz</i> do experimento auditivo.	39

4.9	Configuração final do SOM (10×10) para os dados reais, o neurônios em destaque representam os padrões ativos dos dados de entrada.	40
4.10	À esquerda, resultado de uma estimulação auditiva no córtex cerebral segundo análise do SOM ($CC > 0.6$). À direita, média da série temporal dos pesos dos neurônios ativos.	40
4.11	Resultado de uma estimulação auditiva no córtex cerebral segundo análise do SOM, fatias 9 e 10 ($CC > 0.6$).	41
4.12	Resultado de uma estimulação auditiva no córtex cerebral segundo análise do SPM.	41

Resumo

A imagem por ressonância magnética funcional (fMRI) tem se tornado uma importante técnica no estudo do cérebro humano em ação, principalmente por permitir uma investigação de forma não-invasiva e por possuir uma boa resolução espaço-temporal. Nos últimos anos, os mais diferentes métodos tem surgido com o intuito de analisar os sinais coletados em fMRI a fim de detectar ativações cerebrais funcionais. Dentre esses métodos, incluem-se os mapas auto-organizáveis (SOM) que juntamente com técnicas de agrupamento hierárquico (HC) mostraram-se bastante consistentes, especialmente por permitirem identificar atividades que surgem de diferentes fontes de sinal. No presente estudo, foi proposto uma análise dos dados fMRI, na qual antes da aplicação do SOM e HC, foi utilizada a entropia de Shannon visando eliminar sinais que podem estar relacionados a fontes de ruídos. Desta forma, aumentando a eficiência do SOM e HC na busca de padrões de atividade. Além disso, a entropia de Shannon, assim como, outras técnicas baseadas na Teoria da Informação, já mostrou-se satisfatória em experimentos fMRI. Os três métodos foram aplicados conjuntamente em dados simulados e em um experimento auditivo real, o que revelou ser possível, efetivamente, encontrar ativações cerebrais funcionais.

Palavras-chave: Mapas auto-organizáveis, fMRI, entropia de Shannon, agrupamento hierárquico, coeficiente de correlação.

Abstract

A magnetic resonance imaging (fMRI) has become an important technique in the study of human brain in action, especially for allowing an investigation of a non-invasive form and to have a good spatial and temporal resolution. In recent years, several methods have emerged in order to analyze the signals collected in fMRI to detect brain functional activation. Among these methods, include the self-organizing maps (SOM) that together with the techniques of hierarchical clustering (HC) were revealed very consistent, especially for allowing to identify activities that appear of different sources of signal. In this study, we proposed an analysis of fMRI data, which predated the implementation of the SOM and HC, we used the Shannon's entropy to eliminate signals that may be related to noise sources. Thus, increasing the efficiency of the SOM and HC in the search for patterns of activity. In addition, the Shannon's entropy, as well as other techniques based on Information Theory, has proved to be satisfactory in fMRI experiments. The three methods were applied together in simulated data and a real auditory experiment, which proved to be possible, effectively, find functional brain activation.

Keywords: Self-organizing maps, fMRI, Shannon's entropy, Hierarchical Clustering, correlation coefficient.

Sumário

Introdução	1
1 Mapas auto-organizáveis	3
1.1 Introdução	3
1.2 O mapa auto-organizável	4
1.3 O Algoritmo	5
1.3.1 O Processo Competitivo	5
1.3.2 O Processo Cooperativo	6
1.3.3 O Processo Adaptativo	7
1.4 Avaliação da qualidade da aprendizagem	9
1.5 Formando grupos no Mapa	10
1.5.1 <i>U-matrix</i>	10
1.5.2 Técnicas de Agrupamento Hierárquico	11
1.6 Entropia de Shannon	15
2 Imagem Funcional por Ressonância Magnética	17
2.1 Introdução	17
2.2 Aquisição das imagens	18
2.3 Efeito BOLD	19
2.4 Experimentos em fMRI	20
2.5 Pré-processamento das Imagens	21
2.5.1 Realinhamento	22
2.5.2 Filtros Espaciais e Temporais	23
2.6 Análise dos dados fMRI	23
2.6.1 O modelo linear geral	25
2.6.2 Método Entrópico	27

3	Materiais e métodos	29
3.1	Dados Simulados	29
3.2	Dados Reais	30
3.3	Análise de Entropia	31
3.4	Procedimentos	32
4	Resultados	35
4.1	Dados Simulados	35
4.2	Dados Reais	37
	Conclusão	42
	Referências Bibliográficas	43

Introdução

A imagem funcional por ressonância magnética (fMRI - *functional magnetic resonance imaging*) é uma ferramenta amplamente usada em estudos do cérebro humano em ação. Ela permite mapear as respostas funcionais quantitativamente usando uma tecnologia que é não-invasiva, oferecendo uma boa resolução espaço-temporal (Logothetis, 2002). A técnica tem sido aplicada de forma bastante consistente em processos cognitivos elevados, como no monitoramento do crescimento de tumores cerebrais, mapeamento pré-cirúrgico, estudos de cronometria mental, e também como método de diagnóstico em doenças de Alzheimer, epilepsia (Fountura et al., 2008).

A fMRI surgiu no início da década de 90 como extensão dos convencionais exames de ressonância magnética. Em um simples procedimento fMRI o sujeito é submetido a estímulos variados enquanto sucessivas imagens vão sendo adquiridas. O scanner MR detecta as atividades funcionais cerebrais a partir de mudanças nos níveis oxigenação do sangue, as quais causam perturbações no campo magnético aplicado. Estas alterações na taxa de oxigenação no sangue são conhecidas como contraste ou efeito BOLD (*Blood Oxigenation Level Dependent*), e podem ser modeladas através da função resposta hemodinâmica (HRF - *Hemodynamic Response Function*).

As imagens obtidas em um experimento fMRI, geralmente, possuem uma baixa relação sinal ruído (SNR - *signal to noise ratio*), o que, somado às outras fontes de ruído dificultam a inspeção visual direta das áreas de atividade cerebral, tornando necessária a utilização de métodos computacionais de análises. Contudo, não existe um consenso entre os pesquisadores em fMRI na escolha de um melhor método, sendo crescente o número de trabalhos que desenvolvem metodologias alternativas de análise dos sinais provenientes da fMRI (de Araújo, 2002).

Uma classe particular do conjunto de métodos de análise envolve técnicas de agrupamento, que consiste em atribuir à série temporal de cada ponto (voxel) dos dados fMRI a diferentes grupos homogêneos, de maneira a maximizar a similaridade entre voxels do mesmo grupo, e minimizar a similaridade entre voxels de grupos distintos. Esperando-se encontrar um grupo que represente um efeito dinâmico de interesse, tal como padrões de ativação. Os mais populares

algoritmos de agrupamento em fMRI são K -médias (Balslev et al. 2002), agrupamento fuzzy (Baumgartner et al., 1998; Fadili et al., 2000) e métodos de agrupamento hierárquico (HC - *Hierarchical Clustering*) (Stanberry et al., 2003).

Dentro da mesma classe de métodos, ainda se pode destacar os mapas auto-organizáveis (SOM - *self-organizing maps*), um tipo de rede neural artificial (RNA) para agrupamento e visualização de dados de altas dimensões, e que preserva uma organização espacial, utilizando uma aprendizagem competitiva (Kohonen, 2001). O SOM já foi utilizado como método de análises de dados de fMRI por Fischer e Hennig (1999), que aplicaram o SOM em fMRI, bem como também em exames de mamografias. Ngan et al. (2002), utilizaram o método juntamente com análises de correlação cruzada para identificar as regiões associadas as estímulos visuais e motor. Peltier, et al. (2003), aplicaram o SOM para identificar conectividade funcionais de baixa-freqüência em fMRI. Chen et al. (2006) e Liao et al. (2008) utilizaram SOM associados a um agrupamento hierárquico para detectar diferentes desenhos experimentais no mesmo mapa.

A partir, principalmente, desses dois últimos trabalhos, desenvolveu-se um estudo com os mapas auto-organizáveis e técnicas de agrupamento hierárquico como ferramentas de análises de dados fMRI. Ao método, foi associada a entropia de Shannon com intuito de torná-lo mais eficaz na detecção de padrões relativos a atividades cerebrais. Para medir esta eficácia é calculado o erro de quantização do SOM, uma medidade que avalia a qualidade de aprendizagem da rede.

O estudo estrutura-se da seguinte maneira: no primeiro capítulo, estudou-se os mapas auto-organizáveis, além entropia de Shannon e técnicas de agrupamento hierarquico. No segundo capítulo, foi feita uma introdução sobre alguns fundamentos e conceitos do fMRI, passando pela aquisição de imagens até uma breve visão do modelo linear geral (GLM - *General Linear Model*) e entropia de Shannon aplicados ao fMRI. O terceiro capítulo é caracterizado pela descrição dos fMRI utilizados no estudo e procedimento metodológico, além da configuração dos parâmetros do SOM e HC. No quarto capítulo são exibidos os resultados das análises do SOM tanto pra dados sintéticos como em dados reais. Por fim, são apresentados conclusão e referencial bibliográfico.

Capítulo 1

Mapas auto-organizáveis

1.1 Introdução

As redes neurais artificiais (RNA) são modelos matemáticos e computacionais baseados nos métodos de processamento cerebral. O cérebro pode ser considerado como um computador, o qual recebe informações, processa-as e toma decisões apropriadas. A unidade de processamento cerebral são os neurônios (Figura 1.1a), que por sua vez são constituídos por três principais componentes: os *dendritos* que são os canais de entrada dos sinais transmitidos por outros neurônios; o *corpo celular* que tem por função coletar e processar estes sinais; e o *axônio*, o qual é responsável por transmitir os estímulos para outras células. Estima-se que haja aproximadamente 10 bilhões de neurônios no córtex humano e 60 trilhões de sinapses¹ (Shepherd e Kock, 1990).

Em um neurônio matemático (Figura 1.1b), as sinapses são caracterizadas por um peso $\mathbf{m}_i = [m_{i1}, m_{i2}, \dots, m_{in}]^T$ associado a um neurônio i de entrada $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$. Estes vetores são, então, combinados linearmente $v = \sum_{j=1}^n m_{ij}x_j$ e uma função de ativação $f(v)$ determina a forma e a intensidade dos valores transmitidos para um outro neurônio. Existem variações deste modelo básico, assim como várias funções de ativação. Sendo uma RNA (Figura 1.1c) formada pela combinação destes neurônios artificiais em uma ou mais camadas, as quais podem conter um ou mais neurônios interligados pelas sinapses.

Uma outra característica importante de uma RNA é a sua habilidade de aprender a partir de seu ambiente e de melhorar o seu desempenho através da aprendizagem. O processo de aprendizagem em uma rede neural está associado a alterações aplicadas aos seus pesos sinápticos, processo este conhecido como *treinamento*. Atualmente, existem uma diversidade de algoritmos

¹As sinapses são pequenas estruturas biológicas responsáveis pela conexão entre neurônios.

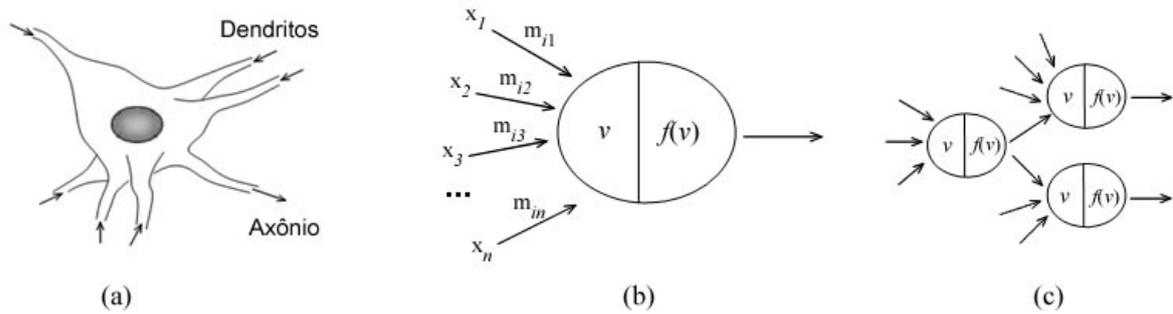


Figura 1.1: (a) Neurônio biológico; (b) Neurônio artificial; (c) Esquema ilustrativo de uma RNA.

de aprendizagem, dentre eles o processo de *aprendizagem competitiva*, o qual forma a base para o estudo do presente capítulo.

A aprendizagem competitiva é um tipo de aprendizado não-supervisionado ou auto-organizado, ou seja, a rede é treinada conforme informações de entrada e sem envolver um agente externo (saídas desejadas). Essencialmente, o princípio da aprendizagem competitiva está na competição entre os neurônios a partir da apresentação de um sinal de entrada, o *neurônio vencedor* tem, então, seus pesos ajustados para melhor responder ao sinal empregado, e por *interação lateral* os neurônios vizinhos ao vencedor também recebem reajustes ao seus pesos. Como resultado, os neurônios tornam-se gradualmente sensíveis aos diversos padrões de entrada e a conjuntos de amostras num domínio específico do espaço de entrada.

1.2 O mapa auto-organizável

Sabe-se que o cérebro, em suas várias regiões, especialmente o córtex cerebral, está dividido de maneira ordenada em diferentes modalidades sensoriais, realizando tarefas específicas, tais como, tarefas motoras, somestésicas, visuais, auditivas, etc. É esta estrutura cerebral que serve como motivação para construção de mapas topográficos artificiais e que possam aprender através de auto-organização de maneira inspirada na neurobiologia.

A formação destes mapas topográficos obedece a um princípio formulado por Kohonen, que diz que a localização espacial dos neurônios de saída são indicativas das características particulares contidas nos dados de entrada (Kohonen, 1990). Por isto o nome mapa auto-organizável (SOM - *self-organizing map*), o qual transforma um padrão de sinal incidente de dimensão arbitrária em um mapa discreto e realiza esta transformação adaptativamente de uma maneira topologicamente ordenada (Haykin, 2001).

Entre os modelos de mapas auto-organizáveis, o mais conhecido na literatura é o de Kohonen

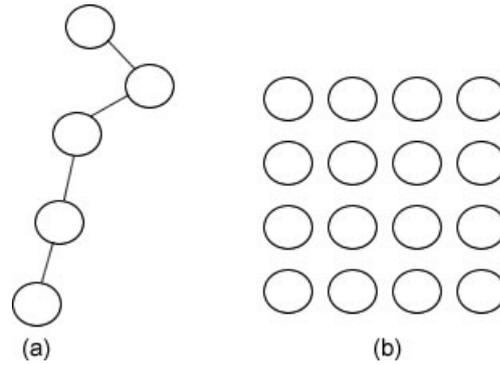


Figura 1.2: Diferentes arquiteturas de mapas auto-organizáveis: (a) mapa unidimensional; (b) mapa bidimensional retangular.

proposto em 1982. O mapa auto-organizável de Kohonen é uma RNA na qual os neurônios ficam dispostos em uma grade (Figura 1.2), normalmente, uni ou bidimensional. Cada neurônio da grade é ligado a todos os componentes da fonte de entrada, isto é, a dimensão dos pesos sinápticos é mesma dos dados de entrada. E, para que o processo de auto-organização seja bem sucedido, todos os neurônios da grade devem ser apresentados a um número suficiente de diferentes padrões de entrada.

1.3 O Algoritmo

O algoritmo da construção do SOM começa inicializando os pesos sinápticos da grade com valores aleatórios, para que nenhuma organização prévia seja imposta ao mapa. O processo, então, é dividido em três fases (Haykin, 2001). Na primeira etapa, a fase competitiva, para cada padrão de entradas, os neurônios da grade competem entre si para escolha de um único vencedor. Na segunda fase, cooperativa, é definida a vizinhança do neurônio vencedor. Por fim, a fase adaptativa, o neurônio vencedor e sua vizinhança tem seus pesos ajustados. As três fases são mais detalhadamente descritas nas subseções seguintes.

1.3.1 O Processo Competitivo

Considere um mapa auto-organizável como descrito anteriormente. Seja $\mathbf{x} \in \mathfrak{R}^n$ um vetor de entradas selecionado aleatoriamente dado por:

$$\mathbf{x} = [x_1, x_2, \dots, x_n]^T \quad (1.1)$$

Para cada neurônio i da grade é associado um vetor peso sináptico $\mathbf{m}_i \in \mathfrak{R}^n$ representado por:

$$\mathbf{m}_i = [m_{i1}, m_{i2}, \dots, m_{in}]^T, \quad i = 1, 2, \dots, k \quad (1.2)$$

onde k indica o número total de neurônios na grade. A fim de se encontrar o neurônio vencedor, o vetor de entrada \mathbf{x} é então combinado com todos os pesos \mathbf{m}_i 's. Um simples critério para essa busca pode ser o produto interno $\mathbf{x}^T \mathbf{m}_i$ para $i = 1, 2, \dots, k$ e selecionar o maior produto. Contudo, outras métricas podem ser usadas, dentre elas é freqüente o uso da distância euclideana entre \mathbf{x} e \mathbf{m}_i , onde a mínima distância indica o neurônio vencedor c . Esta seleção é definida por:

$$c = \arg \min_i \{\|\mathbf{x} - \mathbf{m}_i\|\}, \quad i = 1, 2, \dots, k \quad (1.3)$$

1.3.2 O Processo Cooperativo

O neurônio vencedor indica o centro de uma vizinhança topológica de neurônios cooperativos. E esta relação de vizinhança é definida conforme uma função que busque seguir o modelo neurobiológico em que para uma determinada ativação cerebral, o grau de ativação dos neurônios diminui à medida que se aumenta a distância da região de ativação inicial, isto é, a interação lateral entre o neurônio vencedor e seus vizinhos decresce ligeiramente com a distância lateral.

O parâmetro *vizinhança topológica* h_{ci} indica o grau de interação entre o neurônio c e seu vizinho i , assim como d_{ci} representa a *distância lateral* entre os mesmos neurônios. Para que a função h_{ci} obedeça aos critérios neurobiológicos, é necessário que a mesma seja simétrica em relação ao ponto máximo onde define-se $d_{ci} = 0$, e que decresça monotonicamente com o aumento da distância lateral d_{ci} , em que se tem $h_{ci} \rightarrow 0$ quando $d_{ci} \rightarrow \infty$. Uma função freqüentemente usada e que satisfaz estas exigências é a função *gaussiana*, dada por:

$$h_{ci}(t) = \exp\left(\frac{-d_{ci}^2}{2\sigma^2(t)}\right), \quad t = 0, 1, 2, \dots, \quad (1.4)$$

A distância lateral d_{ci} , no caso de uma grade bidimensional, é a distância euclideana e o parâmetro $\sigma(t)$ corresponde a largura da função gaussiana, ou ainda, definida como raio da vizinhança topológica, o qual deve diminuir monotonicamente com o tempo discreto t . O valor de σ é freqüentemente adotado como um decaimento exponencial:

$$\sigma(t) = \sigma(0) \cdot \exp\left(\frac{-t}{\tau_1}\right) \quad (1.5)$$

onde τ_1 é uma constante de tempo e pode ser igual ao número de iterações do algoritmo, ou ainda, $\tau_1 = \frac{n.\text{º de iterações}}{\log \sigma_0}$.

1.3.3 O Processo Adaptativo

Como mencionado anteriormente, o aprendizado de uma RNA, particularmente de um mapa auto-organizável, se dá pelo ajuste dos pesos sinápticos. Sendo assim, depois de se encontrar o neurônio vencedor c , seu peso \mathbf{m}_c é então ajustado, tornando-se ainda mais similar ao vetor \mathbf{x} do espaço de entradas. A forma mais comum de se ajustar este peso deriva do postulado de aprendizado de Hebb (Hebb, 1949):

“Quando o axônio de uma célula A está próximo o bastante para excitar uma célula B e participa do seu disparo repetida ou persistentemente, então algum processo de crescimento ou mudança metabólica acontece em uma das células ou ambas, de maneira que a eficiência de A, como uma das células que dispara B, é aumentada.”

Analicamente, isto significa que os pesos são ajustados de acordo com (Kohonen, 2001):

$$\Delta \mathbf{m}_i(t) = \alpha y_i(t) \mathbf{x}_i(t), \quad (1.6)$$

em que t denota o tempo, $\alpha(t)$ é uma constante positiva que determina a *taxa de aprendizagem*, \mathbf{x} é um sinal de entrada (atividade pré-sináptica) e y_i é um sinal de saída (atividade pós-sináptica). Da Eq. (1.6), então, adiciona-se o termo $-g(y_i)\mathbf{m}_i$, onde \mathbf{m}_i é o vetor peso sináptico do neurônio i e $g(y_i)$ é uma função escalar positiva. Desta forma, podemos expressar o ajuste do vetor peso do neurônio i por:

$$\Delta \mathbf{m}_i(t) = \alpha y_i(t) \mathbf{x}_i(t) - g(y_i) \mathbf{m}_i \quad (1.7)$$

A única exigência da função $g(y_i)$ é que deve passar pela origem, sendo assim:

$$g(y_i) = 0 \quad y_i = 0. \quad (1.8)$$

Para satisfazer a Eq. (1.8), tomamos $g(y_i)$ como uma função linear dada por:

$$g(y_i) = \alpha y_i, \quad (1.9)$$

adotando

$$y_i = h_{ci}. \quad (1.10)$$

A Eq. (1.7), conforme as Eqs. (1.9) e (1.10), pode agora ser escrita como:

$$\begin{aligned} \Delta \mathbf{m}_i(t) &= \alpha y_i(t) \mathbf{x}_i(t) - \alpha h_{ci} \mathbf{m}_i \\ &= \alpha h_{ci} (\mathbf{x} \mathbf{m}_i) \end{aligned} \quad (1.11)$$

Deste modo, o vetor de peso atualizado $\mathbf{m}_i(t+1)$ para o neurônio i , no passo $t+1$, é descrito pela seguinte equação (Haykin, 2001; Kohonen, 2001):

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + \alpha(t) h_{ci}(t) [\mathbf{x}(t) - \mathbf{m}_i(t)] \quad (1.12)$$

que modifica toda a vizinhança topológica do neurônio vencedor c , de maneira que o grau de ajuste dos pesos \mathbf{m}_i é regido pela função vizinhança topológica h_{ci} (Eq.1.4). Através da apresentação repetida dos dados de treinamento, os vetores de peso sináptico tendem a seguir a distribuição dos vetores de entrada devido à atualização da vizinhança. Desta maneira, a grade, após o processo de treinamento concluído, transforma-se em um mapa de características ordenado topologicamente, no sentido de que neurônios que são adjacentes na grade, tenderão a ter vetores de pesos similares.

O parâmetro taxa de aprendizagem $\alpha(t)$, normalmente é definido por um decaimento exponencial:

$$\alpha(t) = \alpha(0) \cdot \exp\left(\frac{-t}{\tau_2}\right) \quad (1.13)$$

onde τ_2 é outra constante de tempo, normalmente adotada como sendo igual ao número de iterações. A taxa de aprendizagem decresce com o tempo, evitando assim, que dados novos apresentados após um longo treinamento, venham a comprometer o conhecimento que já foi sedimentado.

1.4 Avaliação da qualidade da aprendizagem

Durante o processo de treinamento, o objetivo, a cada iteração, é aproximar os vetores pesos sinápticos aos dados de entrada, caracterizando o aprendizado. Para avaliar a qualidade do aprendizado, introduziu-se um erro de quantização vetorial, que calcula a distância entre o vetor de entrada $\mathbf{x} \in \mathfrak{R}^n$ e o peso $\mathbf{m}_c \in \mathfrak{R}^n$ do neurônio vencedor. Desta maneira, o erro de quantização pode ser descrito como um erro de distorção, expresso por (Kohonen, 2001; Samarasinghe, 2006):

$$e = d(\mathbf{x}, \mathbf{m}_c), \quad (1.14)$$

que é a distância entre \mathbf{x} e \mathbf{m}_c . Contudo, pode ser mais apropriado definir o erro de distorção em função da vizinhança topológica h_{ci} (Eq. 1.4). Sendo assim, o erro de distorção para um vetor de entrada \mathbf{x} torna-se:

$$e = \sum_i h_{ci} d(\mathbf{x}, \mathbf{m}_i), \quad (1.15)$$

onde m_i é um vetor peso sináptico pertencente a vizinhança do neurônio vencedor c . Calculando esta medida de distorção para todos os vetores \mathbf{x} do espaço de entrada, a média do erro de distorção E , é calculado por:

$$E = \frac{1}{N} \sum_j \sum_i h_{ci} d(\mathbf{x}^j, \mathbf{m}_i), \quad (1.16)$$

ou ainda, caso a vizinhança topológica não seja usada, a equação (1.16) simplifica-se a:

$$E = \frac{1}{N} \sum_j d(\mathbf{x}^j, \mathbf{m}_i). \quad (1.17)$$

Nas equações (1.16) e (1.17), o termo \mathbf{x}^j denota o j -ésimo vetor de entradas e N o número total de vetores de entrada. Desta maneira, o objetivo do SOM pode ser, alternativamente expressado, encontrar o conjunto de vetores de pesos sinápticos \mathbf{m}_i que minimize globalmente o erro E .

1.5 Formando grupos no Mapa

Após o processo de aprendizagem do SOM, padrões dos dados de entradas similares aparecem em grupos de neurônios vizinhos no mapa. Contudo, muitas vezes, a quantidade de grupos é desconhecida ou de complexa distinção, tornando-se difícil segmentá-los. Para uma melhor visualização destas estruturas alguns métodos tem sido propostos, tais como a matriz de distâncias entre os neurônios vizinhos, denominada matriz de distância unificada ou *U-matrix* (Ultsch, 1993), a qual utiliza da mesma métrica usada durante o processo de treinamento para calcular distâncias entre pesos sinápticos de neurônios adjacentes. O resultado gerado a partir da aplicação da *U-matrix* é uma imagem em que a intensidade de cor de cada região corresponde a uma distância calculada. Uma proposta alternativa para detecção e segmentação automática de grupos no mapa SOM é por meio de técnicas de agrupamento hierárquico (HC - *Hierarchical Clustering*) para aglomerar os neurônios (Vesanto & Alhoniemi, 2000). As técnicas de agrupamento, de um ponto de vista geral, tem por objetivo reunir objetos (neurônios do SOM) em grupos, de forma que os objetos no mesmo grupo são mais parecidos uns com os outros do que com os objetos de outros grupos, sendo assim os agrupamentos resultantes devem então exibir elevada homogeneidade interna (dentro dos agrupamentos) e elevada heterogeneidade externa (entre agrupamentos) (Hair et al., 2009).

1.5.1 *U-matrix*

Para um mapa bidimensional retangular $X \times Y$, encontra-se a *U-matrix* de dimensão $(2X - 1) \times (2Y - 1)$, calculando-se as distâncias dx , dy e dxy (Figura 1.3) para cada neurônio $i(x, y)$ com peso m_i , conforme as equações abaixo:

$$dx(x, y) = \|\mathbf{m}_{i(x,y)} - \mathbf{m}_{i(x+1,y)}\| \quad (1.18)$$

$$(1.19)$$

$$dy(x, y) = \|\mathbf{m}_{i(x,y)} - \mathbf{m}_{i(x,y+1)}\| \quad (1.20)$$

$$(1.20)$$

$$dxy(x, y) = \frac{1}{2} \left(\frac{\|\mathbf{m}_{i(x,y)} - \mathbf{m}_{i(x+1,y+1)}\|}{\sqrt{2}} \frac{\|\mathbf{m}_{i(x,y+1)} - \mathbf{m}_{i(x+1,y)}\|}{\sqrt{2}} \right)$$

em que (x, y) corresponde as coordenadas do neurônio i na grade do mapa. O valor du da *U-matrix* é calculado em função dos valores dos elementos circunvizinhos ao neurônio i , podendo ser a média, valor máximo ou valor mínimo destes valores.

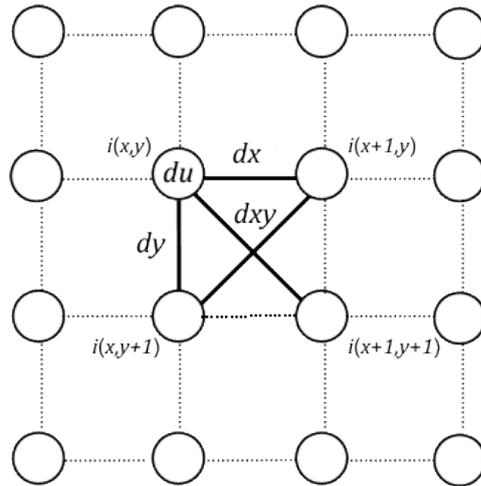


Figura 1.3: As três distâncias da *U-matrix*. Imagem retirada e adaptada a partir de Costa & Netto, 2007.

A interpretação da *U-matrix* se dá a partir da análise da intensidade de cor de cada termo da matriz, em que valores que se apresentam relativamente elevados indicam dissimilaridades entre neurônios vizinhos podendo estar associados a fronteiras de agrupamentos de neurônios, por outro lado, valores baixos correspondem a neurônios vizinhos similares. Também pode-se pensar em uma imagem como uma função tridimensional onde vales, neste relevo topográfico, correspondem a regiões de neurônios que são similares, enquanto que montanhas, i.e., valores relativamente elevados na *U-matrix*, refletem a dissimilaridade entre neurônios vizinhos e podem ser associadas a fronteiras de agrupamentos de neurônios. Pelo fato da *U-matrix* ser uma imagem relativamente complexa, principalmente em problemas de análise de dados reais, geralmente seu uso se restringe a visualização do mapa para a separação manual dos agrupamentos da rede SOM (Costa & Netto 2001).

1.5.2 Técnicas de Agrupamento Hierárquico

Como dito anteriormente, a proposta das técnicas de agrupamento é formar grupos parecidos, isto é, grupos que apresentem um alto grau de correspondência ou semelhança. Portanto, é necessário um coeficiente que possa ser usado como um quantificador desta parenceza. A rigor, o coeficiente de parenceza dividi-se em duas categorias: medidas de similaridade e dissimilaridade (ou distância). Na primeira quanto maior o valor observado menos parecido (mais dissimilares) serão os objetos. Desta maneira, diversos métodos são possíveis, incluindo a correlação, exemplo de medida de similaridade, e a distância euclideana, que é um exemplo de dissimilaridade. A seguir, veremos alguns exemplos de coeficientes de parenceza entre par de objetos (Bussab et

al., 1990).

Considere o conjunto $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_2, \dots, \mathbf{x}_N)$, em que cada $\mathbf{x}_j = (x_{j1}, x_{j2}, \dots, x_{jd})^T \in \mathfrak{R}^d$ representa um objeto descrito pelas suas d componentes reais x_{jk} , $k = 1, \dots, d$.

- Distância Euclideana

$$d(\mathbf{x}_i, \mathbf{x}_j) = \left(\sum_{k=1}^d |x_{ik} - x_{jk}|^{\frac{1}{2}} \right)^2. \quad (1.21)$$

- Distância City-block

$$d(\mathbf{x}_i, \mathbf{x}_j) = \left(\sum_{k=1}^d |x_{ik} - x_{jk}| \right). \quad (1.22)$$

- Distância de Minkowsky

$$d(\mathbf{x}_i, \mathbf{x}_j) = \left(\sum_{k=1}^d |x_{ik} - x_{jk}|^{\frac{1}{p}} \right)^p, \quad (1.23)$$

para $p = 1$ e $p = 2$ passa a ser o caso das métricas city-block e euclideana, respectivamente.

- Distância de Mahalanobis

$$d(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T S^{-1} (\mathbf{x}_i - \mathbf{x}_j) \quad (1.24)$$

onde S^{-1} é a matriz de covariância da amostra.

- Coeficiente de correlação de Pearson

$$d(\mathbf{x}_i, \mathbf{x}_j) = (1 - r(\mathbf{x}_i, \mathbf{x}_j))/2 \quad (1.25)$$

onde

$$r(\mathbf{x}_i, \mathbf{x}_j) = \frac{\sum_{k=1}^d (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum_{k=1}^d (x_{ik} - \bar{x}_i)^2} \sqrt{\sum_{k=1}^d (x_{jk} - \bar{x}_j)^2}}. \quad (1.26)$$

os valores do coeficiente de correlação $r(\mathbf{x}_i, \mathbf{x}_j)$ variam no intervalo real $[-1, 1]$, e se $r(\mathbf{x}_i, \mathbf{x}_j) = 0$, diz-se que os vetores objetos não estão correlacionados.

As técnicas de agrupamento hierárquico envolvem uma série de $N - 1$ (sendo N o número total de objetos) decisões de agrupamento que combinam objetos em uma estrutura de hierarquia. Os dois tipos básicos de procedimentos hierárquicos são aglomerativos e divisivos. Nos métodos aglomerativos, iniciam-se a análise dos dados considerando a existência de N agrupamentos, isto é, cada objeto começa como seu próprio agrupamento, nos passos seguintes, agrupamentos parecidos vão sendo fundidos até que exista um único agrupamento contendo todos os N objetos. Nos métodos divisivos todas as N observações iniciam-se como um único agrupamento, sendo sucessivamente divididos até que cada objeto seja um agrupamento unitário. Para uma melhor visualização do funcionamento do procedimento hierárquico, examinemos o método aglomerativo, que segue um processo simples e repetitivo (Gonçalves, et al., 2008):

1. Começar com N agrupamentos C_1, C_2, \dots, C_N cada um contendo exatamente um único objeto.
2. Determinar o par de agrupamentos distintos (C_i, C_j) de modo que:

$$d(C_i, C_j) = \min_{1 \leq k, l \leq N} d(C_k, C_l), \quad \forall k \neq l. \quad (1.27)$$

onde $d(*, *)$ é o coeficiente de parecença.

3. Formar um novo agrupamento $C_q = C_i \cup C_j$.
4. Calcular as novas distâncias, $d(C_q, C_l)$ entre o novo agrupamento C_q e todos os outros restantes.
5. Repetir os passos 2, 3 e 4 até que os objetos estejam em um único agrupamento.

Anteriormente foi visto algumas maneiras de se determinar o coeficiente de parecença $d(*, *)$ para dois objetos. Contudo é importante definir de que maneira medir a parecença entre agrupamentos com mais de um objeto no processo. Para tal, existem algumas metodologias, as quais são comumente chamadas de algoritmos de agrupamento. Abaixo, serão apresentado alguns algoritmos de agrupamento aglomerativos (Bussab et al., 1990).

- **Ligação Simples.** O método de ligação simples define a parecença entre dois agrupamentos como a menor distância de qualquer objeto de um agrupamento a qualquer objeto no outro. Assim, dados os agrupamentos C_k e C_l do conjunto de objetos X , a distância entre eles será definida por:

$$d(C_k, C_l) = \min d(\mathbf{x}_i, \mathbf{x}_j) : \mathbf{x}_i \in C_k \text{ e } \mathbf{x}_j \in C_l. \quad (1.28)$$

- **Ligação Completa.** O método da ligação completa define a parecença entre dois agrupamentos baseando-se na distância máxima entre objetos em cada grupo. Desta forma, a distância entre os agrupamentos C_k e C_l será:

$$d(C_k, C_l) = \max d(\mathbf{x}_i, \mathbf{x}_j) : \mathbf{x}_i \in C_k \text{ e } \mathbf{x}_j \in C_l. \quad (1.29)$$

- **Ligação média.** No método de ligação média a parecença entre dois agrupamentos é definida pela média de todos os objetos em um agrupamento com todos os objetos em outro. Desta forma, a distância $d(C_k, C_l)$ entre os agrupamentos C_k e C_l é expressada pela equação:

$$d(C_k, C_l) = \frac{\sum_{i,j} d(\mathbf{x}_i, \mathbf{x}_j)}{n_k n_l}, \quad \mathbf{x}_i \in C_k \text{ e } \mathbf{x}_j \in C_l, \quad (1.30)$$

onde n_k e n_l representam os números de objetos contidos em C_k e C_l , respectivamente.

- **Método centróide.** O método centróide, substitui cada fusão de objetos em um único ponto, o qual é representado pelas coordenadas de seu centro. A distância entre grupos é definida pela distância entre os centros.

$$d(C_k, C_l) = d(c_k, c_l), \quad (1.31)$$

$$\text{onde } c_k = \frac{1}{n_k} \sum_i x_i, \quad x_i \in C_k.$$

com n_k igual ao número de objetos em C_k .

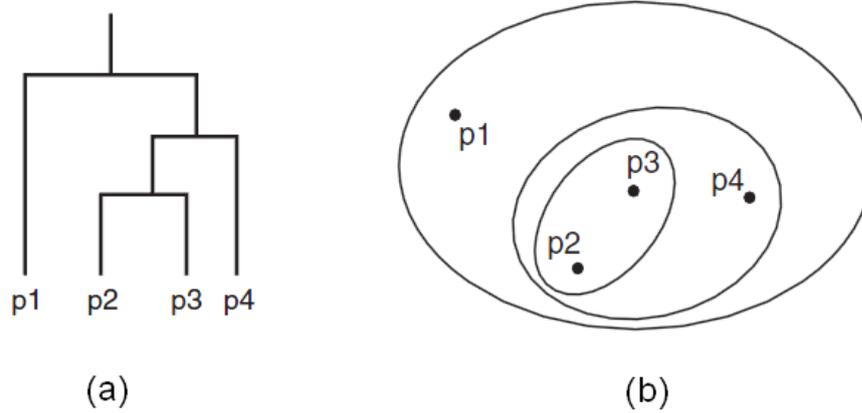


Figura 1.4: (a) Dendrograma; (b) Diagrama de agrupamentos.

Sendo assim, os métodos hierárquicos são uma combinação de um processo repetitivo para agregar com um algoritmo de agrupamento, para definir a parença entre agregados com múltiplos objetos. O resultado gera um gráfico em forma de um diagrama bidimensional conhecido como dendrograma ou árvore de agrupamento (Figura 1.4a). Contudo, surge uma questão: onde cortar o dendrograma de modo a se obter uma quantidade ótima de agrupamentos? Infezivelmente, esse passo fundamental da análise de agrupamentos não está completamente resolvido. Desta forma, no presente estudo foi utilizado a sensibilidade do pesquisador para se determinar uma quantidade de agrupamentos satisfatórios.

1.6 Entropia de Shannon

Algumas vezes é necessário eliminar algumas posições do vetor \mathbf{x} dos dados de entrada da RNA. O truncamento dos dados deve conservar ao máximo as informações capazes de caracterizar os padrões adequadamente. Sendo assim, o objetivo da entropia é eliminar do conjunto de vetores de entrada, aqueles vetores que representem características de menor relevância a partir de um certo limiar a ser estabelecido.

A entropia de Shannon foi estabelecida em 1948 como parte da medida de informação ou incerteza (Shannon, 1948). A teoria da informação lidava com a recém-nascida ciência da comunicação de dados. Para Shannon, a questão fundamental da comunicação é a capacidade de se reproduzir remotamente uma mensagem, com o grau de fidelidade desejado. Esta proposta baseia-se na teoria das probabilidades, onde a quantidade de informação de uma mensagem é medida em função de sua previsibilidade de ocorrência, ou seja, se essa probabilidade de ocorrência for pequena, a mensagem contém muita informação, caso contrário, se ela for previsível, há pouca informação.

Considere, então, uma variável aleatória discreta X em que cada apresentação sua pode ser vista como uma mensagem. Associe o vetor de probabilidades $(p_1, \dots, p_n, \dots, p_N)$ como grau de incerteza da ocorrência de cada mensagem. A entropia, segundo Shannon, é então calculada da seguinte forma (Cover & Thomas, 2006):

$$H(X) = - \sum_{i=1}^n p_i \cdot \log_2 p_i \quad (1.32)$$

Na utilização de logaritmo na base 2 a entropia é expressada em *bits*. Por exemplo, a entropia no lançamento de uma moeda é de 1 bit. Por convenção, adota-se $0 \log 0 = 0$, o que é facilmente justificado desde que $x \log x \rightarrow 0$ para $x \rightarrow 0$. Da definição de entropia, pode-se concluir que:

$$0 \leq H(\mathbf{X}) \leq \log_2(N) \quad (1.33)$$

Além disso, podemos fazer as seguintes afirmações:

1. $H(\mathbf{X}) = 0$ se, e somente se, $p_n = 1$ para algum n , e todas as demais forem iguais a 0. Neste ponto não existe nenhuma incerteza e conseqüentemente a entropia é mínima.
2. $H(\mathbf{X}) = \log_2(N)$ se, e somente se, as probabilidades de todos os eventos forem equiprováveis, isto é, $p_n = \frac{1}{N}$ para todo n . Correspondendo a entropia máxima.

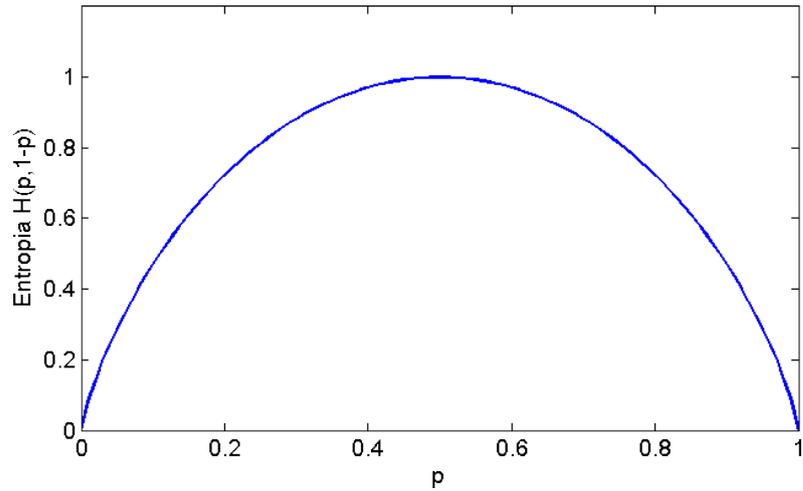


Figura 1.5: Entropia no caso de duas possibilidades com probabilidades p e $1 - p$

O gráfico da entropia no caso de duas possibilidades com probabilidades p e $q = 1 - p$, $H(p, 1 - p)$ pode ser visto na Figura (1.5). Analisando a entropia de Shannon, verifica-se que ela atinge um máximo quando os estados estão igualmente distribuídos e um mínimo quando os estados têm uma distribuição concentrada em qualquer um dos estados.

Capítulo 2

Imagem Funcional por Ressonância Magnética

2.1 Introdução

A *imagem por ressonância magnética* é uma avançada tecnologia que vem sendo cada vez mais utilizada em diagnósticos médicos por imagem. O exame permite retratar, com alta qualidade, os tecidos do interior do corpo humano, permitindo diferenciar os saudáveis dos patológicos.

A técnica explora a interação entre partículas sub-atômicas carregadas, particularmente os prótons, que por sua vez possuem um momento angular freqüentemente referido como *spin* nuclear. Quando este momento angular é não nulo, gera-se um momento magnético não nulo, o que os torna suscetíveis a reagir diante de um campo magnético externo.

O corpo humano é composto por muitos átomos, dentre eles o hidrogênio, que está presente em abundância nos tecidos, os quais são o alvo da ressonância magnética. Quando submetido a um intenso campo magnético, o *spin* de alguns prótons do hidrogênio tendem a assumir um estado de alta energia (alinhando-se em direção oposta ao campo), outros adquirem um estado de baixa energia (alinhando-se na mesma direção do campo). O estado de baixa energia é ligeiramente predominante, o que produz uma pequena magnetização em direção ao campo magnético principal.

Emite-se energia em forma de pulsos de radiofrequência (RF) ao sistema, o que excita os *spins* em estado de baixa energia, deslocando-os para o estado de energia maior. Ao se interromper o

pulso, os *spins* excitados *relaxam*¹ retornando ao seu estado anterior, liberando energia, a qual é detectada e interpretada pelo equipamento para formar as imagens.

2.2 Aquisição das imagens

Existem muitas técnicas para a aquisição de imagens MR tais como seqüência de pulsos RF, como descrito de maneira sucinta anteriormente. Em fMRI, a seqüência mais utilizada é a eco-planar (EPI - *echo-planar imaging*), originalmente proposta por Peter Mansfield em 1977, é atualmente a mais rápida técnica disponível (Hashemi et al., 2004 *apud* Lazar 2008). Com EPI, uma imagem inteira pode ser obtida a partir um único pulso RF inicial. Normalmente, são feitas aquisições de imagens tridimensionais em um intervalo de 1 a 3 segundos. Este valor pode variar de experimento para experimento, e é conhecido como *tempo de repetição* (TR).

As imagens podem ser adquiridas em diferentes cortes: axial, coronal e sagital. As fatias axiais são perpendiculares ao eixo longitudinal do corpo. Coronal, ou frontal, são fatias paralelas a frente do corpo. As fatias sagitais são paralelas a linha que divide o corpo. A Figura 2.1 ilustra os três diferentes cortes.

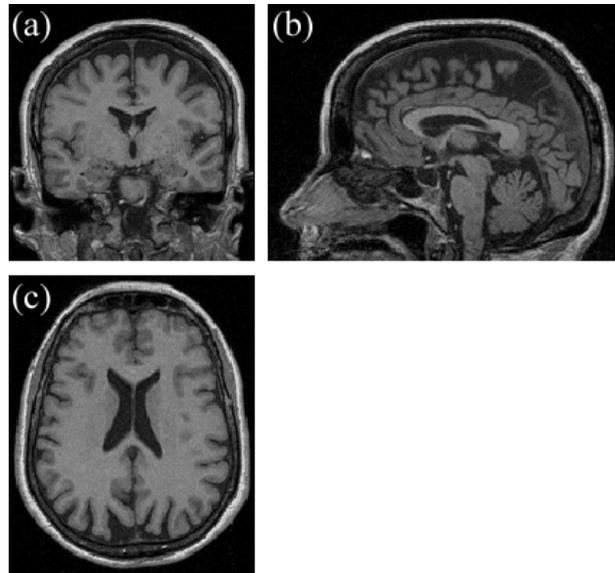


Figura 2.1: Exemplo de orientação de três diferente cortes: (a) Fatia Coronal; (b) Fatia Sagital; (c) Fatia Axial.

Durante o processo de aquisição, alguns parâmetros se fazem importante em virtude de

¹Existem três tempos de relaxação relevantes na ressonância magnética. São eles: tempo de relaxação longitudinal T_1 , tempo de relaxação transversal T_2 e o tempo de relaxação T_2^* associado ao fenômeno decaimento livre da indução (FID - *free induction decay*) (Lazar, 2008).

regular a resolução espacial e temporal dos dados, sobretudo, influenciando na qualidade das imagens.

Pode-se dizer que a resolução espacial de uma imagem MR é determinada pelo tamanho dos elementos da imagem, elementos estes conhecidos como *voxels*². O tamanho do voxel é função de três parâmetros. O primeiro, é o *campo de visão* (FOV - *field of view*), o qual indica o tamanho físico da imagem, isto é, o alcance espacial da imagem de uma borda a outra, e é medido em *mm*². Em segundo o *tamanho da matriz*, que representa o tamanho da grade na qual o FOV é dividido para cada fatia, freqüentemente apresenta-se nas dimensões 64×64 e 128×128. Por fim, a *espessura da fatia*, que como nome já diz é a espessura de uma única fatia, geralmente, medida em *mm*. Sendo assim, as dimensões de um voxel para um FOV de 220 *mm* em uma direção, tamanho da matriz 128×128 e espessura da fatia de 5 *mm* é de 1,72 *mm* × 1,72 *mm* × 5 *mm* (FOV e tamanho da matriz definem duas dimensões $220/128 = 1,72$ *mm* cada).

2.3 Efeito BOLD

O efeito BOLD (*Blood Oxigenation Level Dependent*) tem sido amplamente usado nas recentes técnicas de fMRI. O fenômeno, que foi primeiramente descrito por Seiji Ogawa em 1990 em seu estudo com ratos, tem origem nas propriedades magnéticas de oxigenação e desoxigenação do sangue (Ogawa, et al., 1990). Em 1936, Paulling e Coryell haviam mostrado que a associação oxigênio e hemoglobina, denominada oxi-hemoglobina, é diamagnética, enquanto que a hemoglobina quando não está ligada ao oxigênio, desoxi-hemoglobina, é paramagnética (Paulling & Coryell, 1936).

Durante uma atividade cerebral a taxa de fluxo sanguíneo é aumentada em virtude da maior necessidade de glicose e oxigênio. Com a alteração do metabolismo local, a quantidade de oxigênio torna-se maior que a demanda, aumentando os níveis de oxigênio nos vasos sanguíneos e diminuindo a concentração de sangue desoxigenado. A diminuição da concentração de desoxi-hemoglobina, sendo ela paramagnética, gera mudanças na susceptibilidade magnética local, a qual varia linearmente com a oxigenação do sangue (Weisskoff & Kiihne, 1992 *apud* Buxton, 2009), provocando perturbações no sinal da ressonância magnética com alteração dos tempos de relaxação T_2 e T_2^* (Thulborn et al., 1982; Ogawa et al., 1990). Estas mudanças são detectadas nas imagens MR através de pequenas mudanças na intensidade de brilho dos voxels ativos.

A Figura (2.2) mostra a evolução temporal de uma ativação neural, associada a uma tarefa ou atividade cerebral espontânea, conhecida como função resposta hemodinâmica (HRF - *Hemo-*

²voxel (volume elements), um termo generalizado da palavra pixel (picture elements).

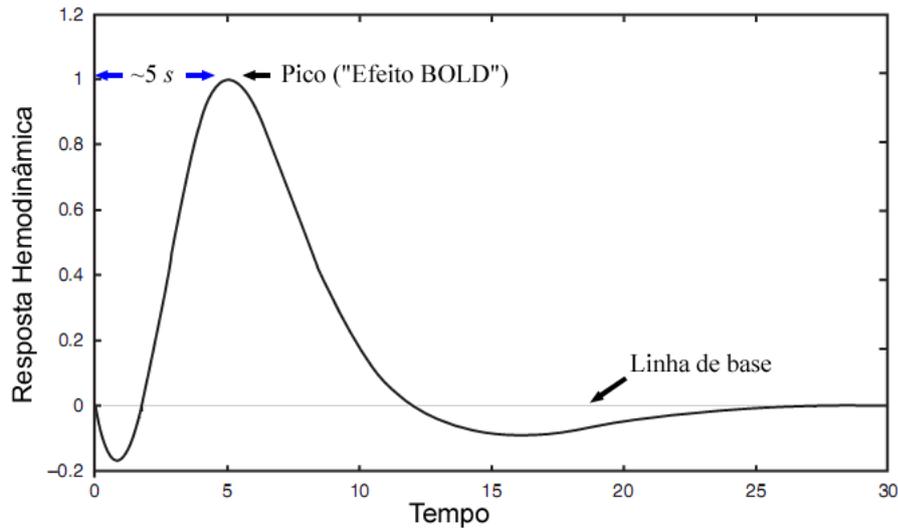


Figura 2.2: Função Resposta Hemodinâmica para um estímulo de curta duração. Imagem retirada e adaptada a partir de Mulert & Lemieux, 2010

dynamic Response Function). Pode-se, inicialmente, verificar um breve decréscimo do sinal, o qual nem sempre pode ser observado. Posteriormente, começa uma etapa de resposta positiva BOLD mantendo-se de 5-10s, consistindo de uma ascensão do sinal até alcançar o valor máximo. Em seguida, há uma queda no sinal, a qual estende-se até ficar abaixo da linha de base, para só então, retornar a valores comparáveis a linha de base inicial. Contudo, a forma da HRF varia de acordo com as propriedades da tarefa executada e com a atividade cerebral inerente. Se a taxa de atividade neural for aumentada, a amplitude da HRF também aumenta e se esta atividade for prolongada a largura da HRF é estendida. Ainda de acordo com o grau de complexidade da tarefa, pode haver HRF's diferentes em uma mesma área (Dale & Buckner, 1997).

2.4 Experimentos em fMRI

A fim de que se possa observar o comportamento da resposta BOLD resultante de uma atividade neural são elaboradas tarefas, as quais seguem uma estrutura organizacional temporal, a serem realizadas pelo sujeito durante o processo de aquisição de imagens em um exame fMRI. As diferentes formas de se apresentar estas tarefas são comumente chamadas de *desenhos experimentais* ou *paradigmas*. O seu planejamento é fundamental na obtenção de bons resultados, que refletem a atividade de regiões específicas de interesse (de Araújo, 2002). Em fMRI, os principais tipos de paradigmas são bloco e evento-relacionado. A Figura 2.3 mostra as diferenças entre os paradigmas.

O *paradigma em bloco* é o mais comum experimento utilizado em fMRI. Nele, longos períodos,

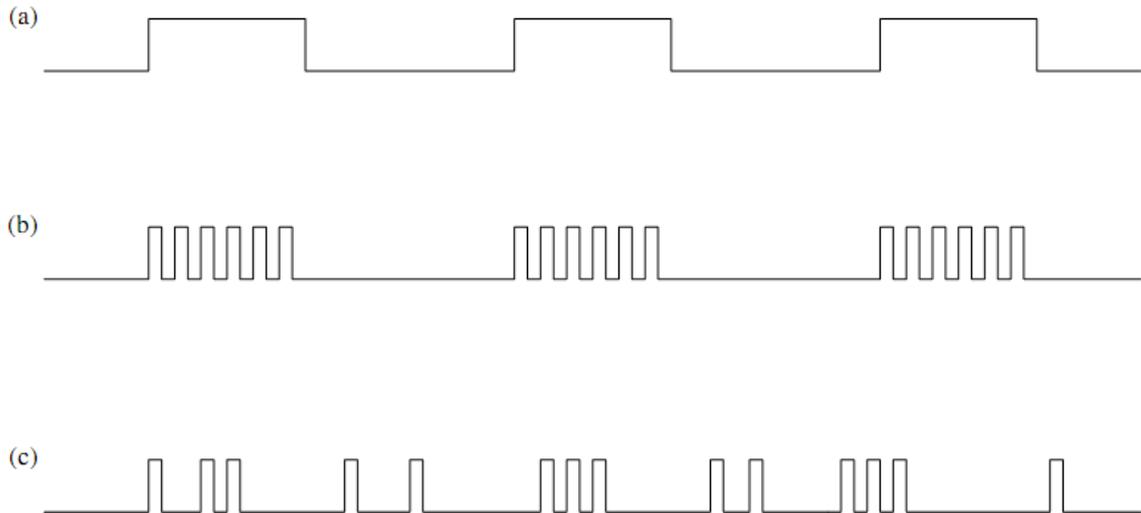


Figura 2.3: Experimentos em fMRI: (a) Paradigma em Bloco, quatro períodos de repouso intercalados com três de ativação; (b) Um paradigma misto, paradigma em bloco interpretado como evento-relacionado; (c) um paradigma evento-relacionado aleatório. Imagem retirada de Sarty, 2007.

ou blocos, de repouso são alternados com períodos igualmente longos de estímulos, ou ainda, pode-se alternar períodos de tarefas distintas. Um ciclo de blocos de ativação e repouso é conhecido como época. No *paradigma evento-relacionado* os estímulos são apresentados por um período curto, seguido por um período relativamente longo de repouso. Cada um tem suas vantagens e desvantagens, por exemplo, o paradigma em bloco, geralmente, é mais sensível para detectar ativações, enquanto que o evento relacionado pode ser melhor para caracterizar a resposta BOLD. Uma estratégia mista pode ser uma alternativa por reunir as vantagens das duas técnicas, porém sua análise pode ser mais complicada, podendo comprometer o estudo.

2.5 Pré-processamento das Imagens

A etapa do pré-processamento se faz necessária para corrigir ou minimizar uma variedade de aspectos não desejáveis, porém freqüentes após a aquisição de dados fMRI. Fatores como diferenças no tempo de aquisição além de movimento involuntário da cabeça, batimentos cardíacos e respiração, entre outros processos fisiológicos, geram ruídos, o que dificulta a análise e interpretação dos dados, e se não forem tratados podem invalidar análises e conclusões.

2.5.1 Realinhamento

Enquanto o equipamento MR captura as imagens, o sujeito é instruído a permanecer parado, porém pequenos movimentos podem ser inevitáveis. Se ele move a cabeça, por mínimo que seja, a posição do cérebro dentro das imagens funcionais irão variar com o tempo, isto é, a série temporal dos voxels já não se referem ao mesmo ponto do cérebro. Portanto, é comum o uso de recursos para prevenir o movimento da cabeça. Além disso, processos fisiológicos tais como respiração e batimentos cardíacos provocam movimentos intracranianos, e ainda expansão cerebral (Hu et al, 1995). Uma estratégia de correção de movimento é alinhar todas as séries temporais das imagens tomando como referência uma imagem, a qual pode ser a primeira (Friston et al., 1996) ou a média das imagens (Bullmore et al., 1999).

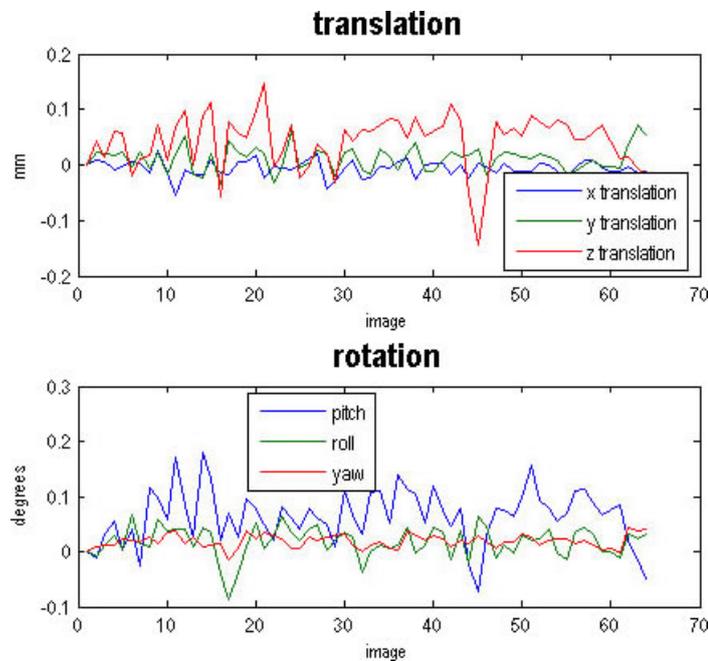


Figura 2.4: Realinhamento espacial de um experimento auditivo. Na legenda, os 6 parâmetros de corpo rígido, para translação e rotação. Figura obtida após a etapa de realinhamento realizada no SPM8 (Statistical Parametric Mapping).

O processo de alinhamento utiliza seis parâmetros que definem translação e rotação, aplicando-se transformações de corpo rígido³. Translações e rotações, em cada direção, são então estimadas a fim de minimizar as distancias entre as imagens e a imagem referência (Figura 2.4). Uma vez determinado um conjunto de parâmetros de realinhamento, os dados originais são re-amostrados por interpolação espacial.

³em inglês *rigid body registration*, onde considera-se que cabeça e cérebro não mudam sua forma, somente sua posição e orientação.

Além disso, nos exames de fMRI, as imagens (ou volumes) são formadas por fatias, as quais são adquiridas uma por vez. O problema é que, em geral, os métodos de análises de dados de fMRI assumem que todas as fatias foram adquiridas ao mesmo tempo. Isto pode levar a resultados não ótimos. Uma estratégia, é corrigir este defasamento por reajuste da série temporal de cada voxel de uma dada fatia, o que pode ser feito via transformada de Fourier.

2.5.2 Filtros Espaciais e Temporais

O ruído, qualquer que seja sua origem, reduz a significância de detectar ativações. O processamento das imagens funcionais pode envolver a aplicação de filtros espaciais e temporais na tentativa de melhorar a qualidade dos dados fMRI, aumentando a relação sinal ruído (SNR - *signal to noise ratio*).

O método mais comum de realizar a filtragem espacial é com a utilização de um filtro de perfil *gaussiano* convoluindo com cada volume de imagens (Jezzard et al., 2001). A forma usual de caracterização do filtro é através dos parâmetros de largura a meia altura (FWHM - *full width half maximum*) que determina a forma de atuação do filtro sobre a imagem. Além de melhorar a SNR, a aplicação do filtro espacial se faz importante para posteriores análises estatísticas, fazendo com que a distribuição dos erros se aproxime de uma distribuição normal. O principal papel dos filtros temporais é remover componentes indesejáveis da série temporal sem comprometer o sinal de interesse. Com filtros passa-alta, busca-se atenuar apenas baixas frequências do sinal, as quais podem estar relacionadas a alguns efeitos fisiológicos ou associados a instabilidades do hardware. Filtros passa-baixa tentam reduzir as altas frequências da série temporal de cada voxel. É comum usar filtros passa-baixa de perfil gaussiano de maneira similar aos filtros espaciais.

2.6 Análise dos dados fMRI

Em um conjunto de dados fMRI cada posição do cérebro está ligada a um voxel em uma imagem tridimensional. Enquanto o sujeito realiza distintas tarefas, uma seqüência destas imagens vão sendo adquiridas. Podemos então associar uma série temporal a cada voxel da imagem, em que cada elemento da série tem o valor do voxel das imagens funcionais adquiridas no tempo. E, a partir da análise dos dados, espera-se encontrar evidências da resposta BOLD nas séries temporais dos voxels pertencentes a área de interesse (ver Figura 2.5). Isto pode vir a ser um desafio, haja vista que o efeito BOLD tem uma baixa SNR, promovendo diferenças de contraste muito pequenas, além da abundância de informações a serem analisadas. Contudo, muitos

métodos tem sido introduzidos, alguns originalmente testados em técnicas mais antigas, como a tomografia por emissão de pósitrons (PET - *Positron Emission Tomography*).

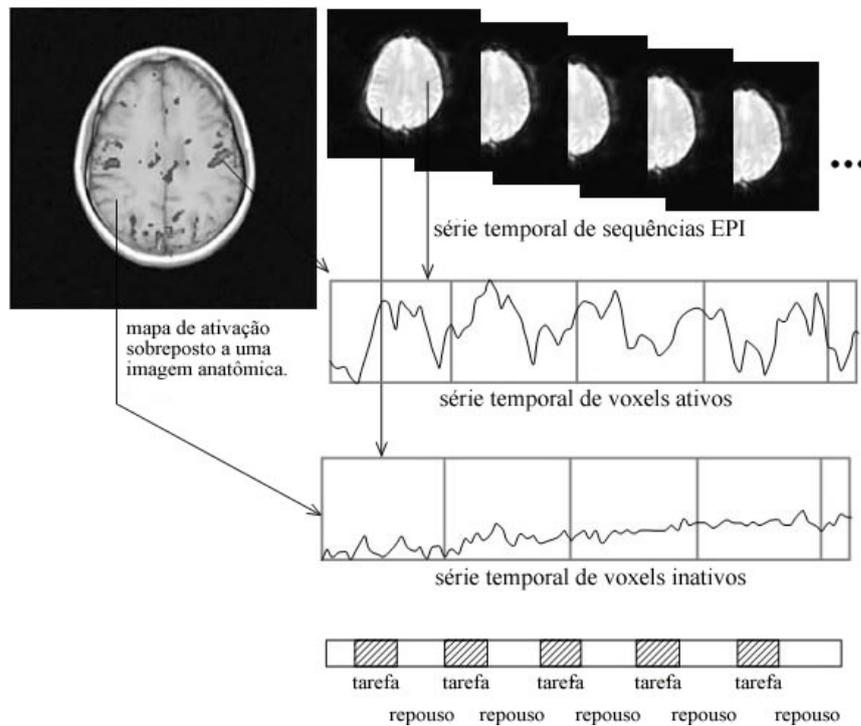


Figura 2.5: Série temporal de voxels ativos e não ativos de uma única fatia dos volumes de imagens. Imagem retirada e adaptada a partir de Sarty, 2007.

Os primeiros métodos de análise consistiam na subtração de imagens adquiridas em períodos de atividade das imagens adquiridas em períodos de repouso. Atualmente, a maioria das metodologias utilizam de técnicas estatísticas para detecção de regiões ativas, incluindo, teste-t de *Student* (Huettel et al., 2004), correlação cruzada (Rabe-Hesketh et al., 1997) e variações do Modelo linear geral (GLM - *General Linear Model*) (Friston, et al., 1995). Outros métodos abordam perspectivas multivariadas de componentes e análise de correlação, tais como análise de componentes independentes (ICA - *Independent Component Analysis*) (McKeown et al., 1998; Biswal e Ulmer, 1999) e análise de componente principal (PCA - *Principal Component Analysis*) (Friston et al., 1993). Além disso, métodos que exploram a teoria da informação também tem sido empregados em fMRI, tais como, entropia de Shannon e Tsallis dependentes do tempo (de Araújo, et al. 2003; Sturzbecher, 2006). Nas subseções seguintes é feito um apanhado geral do GLM, um método muito popular entre os pesquisadores além de um breve comentário a respeito dos métodos entrópicos.

2.6.1 O modelo linear geral

No modelo linear geral, o comportamento de uma dado voxel é representando por um vetor y , ao longo de N observações (scans). Cada observação é modelada como combinação linear de um conjunto de variáveis explanatórias x , somados a um termo de erro ϵ :

$$y_i = \beta_1 x_{i1} + \dots + \beta_j x_{ij} + \dots + \beta_K x_{iK} + \epsilon_i, \quad i = 1, \dots, N \quad (2.1)$$

onde os parâmetros β_j são as amplitudes que indicam o quanto cada uma das K variáveis x contribui para os dados. Expandindo-se a Equação (2.1) para N observações, de forma matricial, tem-se:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1K} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & \dots & x_{ij} & \dots & x_{iK} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{N1} & \dots & x_{Nj} & \dots & x_{NK} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_j \\ \vdots \\ \beta_K \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_i \\ \vdots \\ \epsilon_N \end{pmatrix} \quad (2.2)$$

ou de maneira simplificada:

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon \quad (2.3)$$

onde \mathbf{X} é denominada matriz experimental⁴. Cada coluna desta matriz, ou conjunto de colunas, modela uma característica particular da série temporal observada. A análise consiste em estimar as amplitudes β_j com intuito de fornecer o melhor ajuste do modelo aos dados fMRI, ou seja, calcular qual a melhor combinação de parâmetros β_j que minimiza o termo de erro. O problema de encontrar a melhor solução para β pode ser resolvido pelo método dos mínimos quadrados, resultando em:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (2.4)$$

Os fatores que descrevem a matriz \mathbf{X} envolvem desde o desenho experimental, como também predições da resposta hemodinâmica, a qual pode ser modelada como uma simples função de

⁴Em inglês *design matrix*

distribuição gama (Lange & Zeger, 1997) ou Poisson (Friston et al., 1994). Contudo, o modelo fisiológico mais plausível propõe a diferença de duas funções gama, capturando o pequeno decréscimo da resposta BOLD após o término do estímulo, antes do retorno a linha de base (Friston et al., 1998).

$$h(t) = \left(\frac{t}{d_1}\right)^{a_1} \exp\left(-\frac{(t-d_1)}{b_1}\right) - c \left(\frac{t}{d_2}\right)^{a_2} \exp\left(-\frac{(t-d_2)}{b_2}\right) \quad (2.5)$$

onde $d_j = a_j b_j$ correspondem ao tempo até o pico de cada uma das funções gama, e $a_1 = 6$, $a_2 = 12$, $b_1 = b_2 = 0.9s$, e $c = 0.35$ (Glover, 1999).

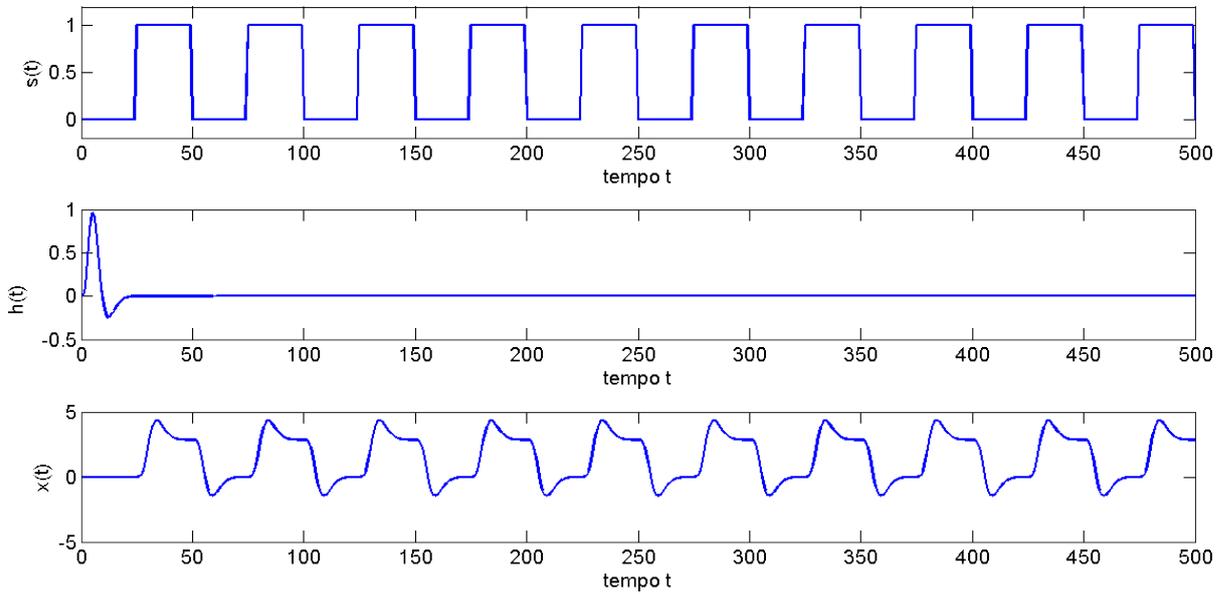


Figura 2.6: (a) Representação de um paradigma em bloco, o valor 0 indica o repouso e 1 a ativação; (b) HRF modelada segundo a proposta de Friston (1998), usando parâmetros de Glover (1999); (c) Convolução da HRF com o paradigma em bloco.

Esta particular escolha da função HRF pode ser vista na Figura (2.6b). O modelamento conjunto da função de estímulo $s(t)$ (Figura 2.6a) com a resposta hemodinâmica esperada $h(t)$ é realizado com a convolução de ambas as funções no domínio de tempo (Figura 2.6c), conforme:

$$x(t) = \int_0^{\infty} h(u)s(t-u)du. \quad (2.6)$$

O resultado $x(t)$ é utilizado na composição da matriz experimental \mathbf{X} . Desse modo, a necessidade de informar a forma da HRF constitui a principal desvantagem do GLM, pois se esta não for precisamente modelada, o método perderá sensibilidade de detecção (Sturzbecher, 2006).

2.6.2 Método Entrópico

Nos estudos desenvolvidos por Capurro (1999), de Araújo (2003) e Sturzbecher (2006), a entropia de um sinal é calculada no domínio do tempo, definindo estados acessíveis do sistema como a amplitude do sinal dentro de um intervalo específico (janela), centrado em um instante t . No contexto do fMRI, a evolução temporal de um voxel ao longo de uma época foi denotada por $s(t)$. Onde cada posição discreta da curva $s(t)$ corresponde ao valor de intensidade do voxel em função do tempo. Para o cálculo da entropia, dividiu-se o sinal $s(t)$ em duas janelas (W_1 e W_2) e o intervalo de tempo total (W) foi dividido em um número discreto de (k) níveis de intensidade do sinal (I_k) (ver Figura 2.7). Desta maneira, a probabilidade de um nível de intensidade (I_k) conter uma porção do sinal ($s(t_k)$) foi dada por (ver de Araújo, et al., 2001; Sturzbecher 2006):

$$p^m(I_k) = \frac{\text{n.º de valores de } s(t_k) \in W_m \text{ no intervalo } I_k}{\text{n.º de valores de } s(t_k) \in W_m}. \quad (2.7)$$

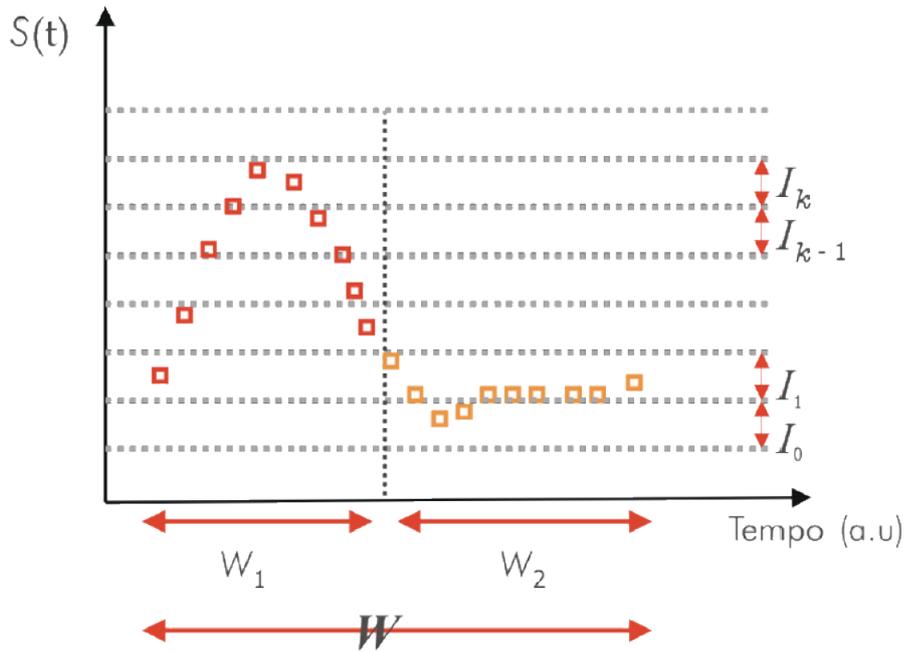


Figura 2.7: Diagrama descritivo do cálculo da entropia de Shannon. As probabilidades podem ser obtidas sobre cada janela separadamente (W_1 e W_2) e para cada nível I_k . Imagem retirada de Araújo et al. 2003.

De posse das probabilidades, o cálculo da entropia de Shannon para cada uma das janelas é obtido, independentemente, de acordo com a Eq. (1.32). Observou-se então, que a evolução temporal da entropia nas regiões onde haviam um predomínio de atividade devido a determinada tarefa, se assemelha a uma função “dente de serra”. E, o algoritmo era concluído aplicando-se uma correlação cruzada entre uma função dente de serra simulada e os valores de evolução

temporal das entropias de Shannon calculadas. Sendo possível, então, gerar um mapa estatístico com valores de coeficientes de correlação que se referem às áreas cerebrais com maior chance de estarem ativas.

Capítulo 3

Materiais e métodos

3.1 Dados Simulados

Para avaliar a qualidade ou sensibilidade do método em estudo, foi criado um experimento fMRI (64×64) artificialmente (Figura 3.1), acrescentando sinal BOLD simulado ao longo de 120 fatias. A HRF seguiu a proposta dada por Friston modelada como a diferença de duas funções de distribuição gama (Eq. 2.5). A função de estímulo foi modelada a partir de um paradigma em bloco, consistindo de 6 blocos de ativação e 6 blocos de repouso, com 20 pontos temporais cada época.

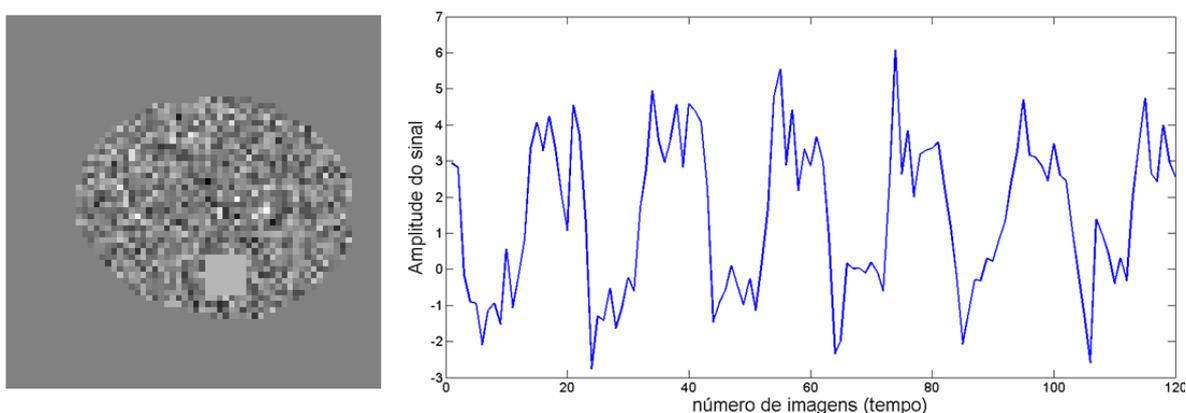


Figura 3.1: Esquema ilustrativo da distribuição espacial de voxels ativos e a intensidade destes voxels ao longo do tempo em dados sintéticos com $\text{SNR} = 5\text{dB}$

A área considerada ativa corresponde a 49 voxels, enquanto que 1349 voxels designam a massa cinzenta restante, isto é, contém apenas ruído. Os outros 2698, correspondem ao fundo da imagem, e não variam com o tempo. Para tornar o modelo artificial mais realístico, foi adicionado ruído às séries temporais dos voxels ativos. O ruído adicionado descreve um padrão

gaussiano, em que os valores do desvio-padrão do ruído determinarão as características da SNR, a qual é medida em *decibéis* (dB), sendo calculada pela seguinte expressão:

$$\text{SNR} = 10 \log \left(\frac{\sigma_S^2}{\sigma_R^2} \right) = 20 \log \left(\frac{\sigma_S}{\sigma_R} \right) \quad (3.1)$$

onde σ_S e σ_R são o desvio-padrão do sinal e do ruído, respectivamente. A Figura (3.2) apresenta séries temporais com 6 épocas para alguns valores de SNR. É possível visualizar que quanto maior o ruído menor é a SNR, conseqüentemente mais difícil de se identificar a resposta BOLD.

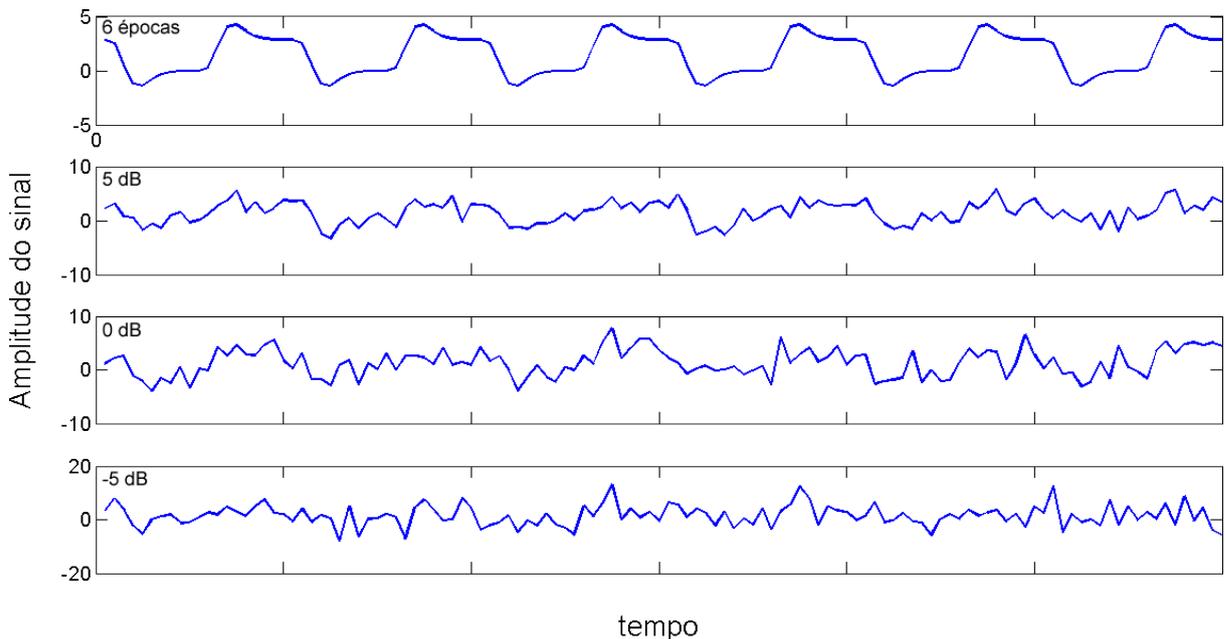


Figura 3.2: Séries temporais para diferentes valores de SNR. Para valores altos de SNR, a HRF é mais evidente.

3.2 Dados Reais

O experimento fMRI foi realizado em um scanner Siemens de 1.5 T (Magnetom Vision, Erlangen, Germany), usando-se seqüências EPI (echo-planar imaging) com os seguintes parâmetros: TE = 60 ms, TR = 4.6 s, FA = 90°, FOV = 220 mm, espessura das fatias de 6.25 mm, e dimensão do voxel = 1.72 × 1.72 × 6.25 mm. Foram adquiridos 64 volumes cerebrais com 16 fatias cada, e tamanho da matriz 128 × 128.

Durante o procedimento, o indivíduo foi submetido a um estímulo auditivo com paradigma em bloco, sendo 6 blocos em repouso (27.5 s cada) intercalados com 5 de atividade (27.5s cada). Durante a tarefa o paciente ouviu passivamente uma estória complexa com uma estrutura

narrativa padrão. Imediatamente após o exame, o sujeito teve que informar o conteúdo da narrativa.

As imagens adquiridas foram devidamente pré-processadas, usando-se o software SPM8 (Statistical Parametric Mapping)¹. Nesta etapa foi realizada a correção de movimento, a correção do tempo entre fatias, a suavização espacial (através de um Filtro Gaussiano com FWHM de 4 mm) e também a filtragem temporal. As análises dos dados simulados e pós-processamento das imagens reais, através dos métodos descritos anteriormente, foram realizadas utilizando algoritmos próprios implementados em MATLAB 7.6.

3.3 Análise de Entropia

Alguns autores recomendam a redução do conjunto de voxels na aplicação dos algoritmos, devido ao fato que a proporção de voxels ativos no cérebro é relativamente pequena (Gibbons et al., 2004). Então, para aumentar a eficiência da análise dos dados, é aplicado a cada voxel do conjunto de características uma análise de entropia, como sugerido na Seção 1.6, com o objetivo de suprimir algumas posições do vetor dos dados de entrada, sendo eliminados os voxels com nível de entropia abaixo de certo limiar empiricamente determinado.

No presente estudo, o cálculo da entropia do sinal foi baseado nas pesquisas anteriores (ver Seção 2.6.2). Desta vez, $s(t)$ denota o comportamento de toda série temporal de um voxel. Este sinal $s(t)$ é dividido em dois intervalos de intensidade ($I_l, l = 1, 2$). Essa divisão é realizada calculando-se os valores máximo e mínimo da série temporal. Matematicamente, tem-se:

$$s_0 = \min_k \{s(t_k), k = 1, \dots, K\}, \quad (3.2)$$

$$s_1 = \max_k \{s(t_k), k = 1, \dots, K\}, \quad (3.3)$$

Desse modo, define-se os intervalos de intensidade, I_1 e I_2 , como:

$$I_1 = [s_0, s_M), \quad (3.4)$$

$$I_2 = [s_M, s_1], \quad (3.5)$$

em que s_M é a média aritmética de s_1 e s_2 .

¹<http://www.fil.ion.ucl.ac.uk/spm/>

E daí, a probabilidade de cada intervalo, simplificando a Eq. (2.7), é dada por:

$$p(I_l) = \frac{\text{n.º de valores de } s(t_k) \text{ no intervalo } I_l}{\text{n.º total de valores de } s(t_k)}. \quad (3.6)$$

E a entropia sendo, por fim, calculada segundo a Eq. (1.32). Conforme analisamos a Figura (3.3), é possível dizer que a entropia é um parâmetro intuitivo no sentido de que é possível distinguir, visualmente, um sinal regular de um sinal irregular. Na Figura (3.3a) é possível observar que para um sinal correspondente a um voxel não ativo, a presença de picos torna a distribuição de probabilidade estreita, conseqüentemente, um baixo valor de entropia. Por outro lado, na Figura (3.3b) apresenta a série temporal de um provável voxel ativo. O elevado valor da entropia, neste caso, é associado a ampla distribuição de probabilidade.

3.4 Procedimentos

Como etapa de preparação, calculou-se o valor da entropia de todos os voxels dos dados fMRI, conforme descrito antes. Aqueles voxels que apresentaram valores de entropia menores que o limiar escolhido foram eliminados do conjunto de dados de entrada referentes a etapa do SOM juntamente com aqueles voxels pertencentes a parte externa do cérebro (fundo da imagem).

Para configuração dos parâmetros de configuração do SOM usou-se informações obtidas em estudos anteriores (Peltier, et al., 2003; Liao, et al., 2008) tanto nos dados simulados quanto nos dados reais. Desta forma, utilizou-se uma grade bidimensional retangular 10×10 . Inicializou-se a taxa de aprendizagem (Eq. 1.13) em $\alpha(0) = 0.1$, e o parâmetro largura efetiva (Eq. 1.5) $\sigma(0) = 7$. O número de iterações do SOM foi regulado conforme a estabilização do erro (Eq. 1.17), contudo Peltier, et al., (2003) sugere que para imagens de dimensões 64×64 , 100 iterações seja o suficiente. E, a seleção do neurônio vencedor é dada pela máxima correlação entre dado de entrada e neurônios da grade, em alternativa a mínima distância euclideana (Eq. 1.3):

$$m_c(t) = \arg \max_i \{corr(\mathbf{x}(t), \mathbf{m}_i(t))\}, \quad i = 1, 2, \dots, k. \quad (3.7)$$

Após o treinamento do SOM, com a finalidade de reconhecer nos dados de entrada aqueles padrões que representam atividade, aplicou-se agrupamento hierárquico aglomerativo, adotando o método de ligação simples (Eq. 1.28), e tomando o coeficiente de correlação de Pearson como coeficiente de parença (Eq. 1.26). Contudo só foram agrupados neurônios vizinhos no mapa, isto é, para que dois neurônios ou agrupamentos fossem agregados, a distância em relação a

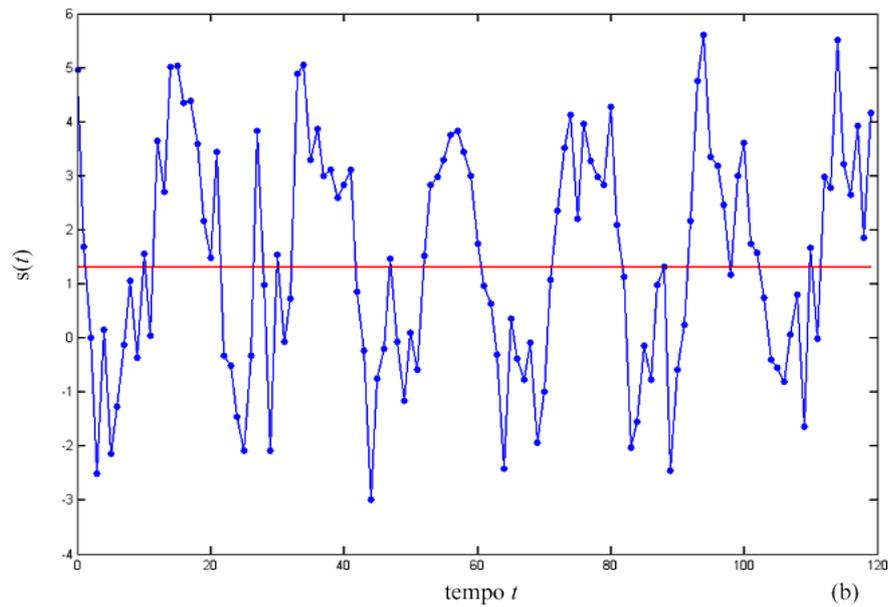
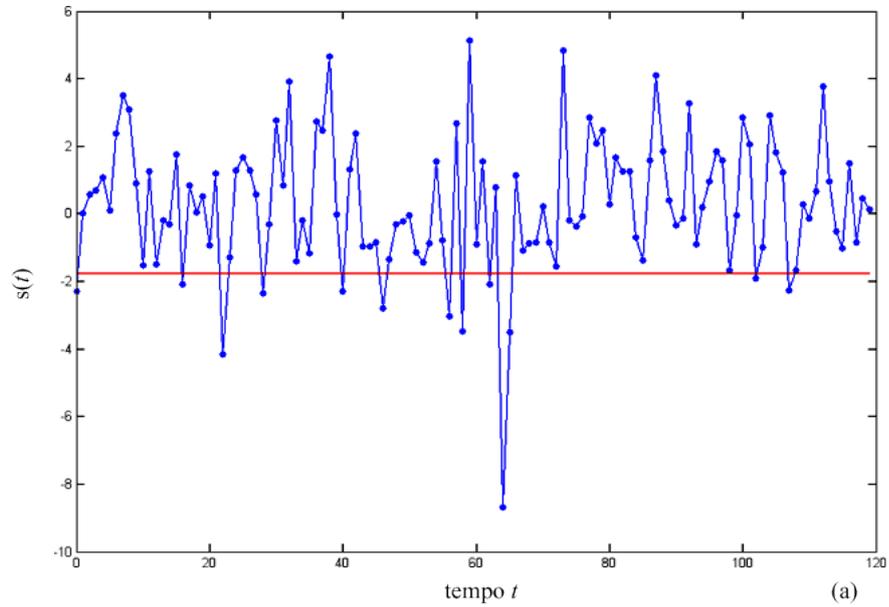


Figura 3.3: (a) Sinal apresentando baixa entropia, os pontos concentram-se no intervalo superior, com as seguintes distribuição de probabilidade de cada intervalo, $p(I_1) = 0.1083$ e $p(I_2) = 0.8917$ e entropia $H = 0.4949$; (b) Sinal apresentando alta entropia, com distribuição de probabilidade para os intervalos $p(I_1) = 0.4667$ e $p(I_2) = 0.5333$ e entropia $H = 0.9968$.

posição discreta entre eles não deveria ser superior a $\sqrt{2}$. Além disso, determinou-se *U-matrix* para auxiliar na procura da quantidade de agrupamentos a serem formados.

Tendo definido quais agrupamentos apresentam evidências de resposta BOLD, então, correlacionou-se a média dos pesos dos neurônios pertencentes ao agrupamento encontrado com todos os voxel das imagens. Esta correlação (Eq. 1.26) visa medir o quanto os dois sinais estão relacionados linearmente, que em fMRI possibilita inferir se um voxel está ativo ou não, dependendo do coeficiente de correlação (Cabella, 2008)(por conveniência, adotamos a partir daqui a notação CC, aos invés de r). Feito isso, aqueles voxels que apresentaram um CC com a média dos pesos, superior a um limiar a ser escolhido são destacados em uma máscara com a finalidade de se gerar o mapa de ativações.

Capítulo 4

Resultados

4.1 Dados Simulados

A primeira etapa do processo foi calcular a entropia de Shannon para cada voxel dos dados simulados. Os valores de entropia para os 1398 voxels contidos no interior do cérebro artificial variaram de 0.4949 à 1, e como já esperado, os 49 voxels com sinal de ativação apresentaram um alto valor de entropia (Figura 4.1a). Após isto, cerca de 6% dos 1398 voxels foram eliminados dos dados de entrada do SOM, estes voxels apresentaram um valor de entropia $H < 0.85$ (Figura 4.1b).

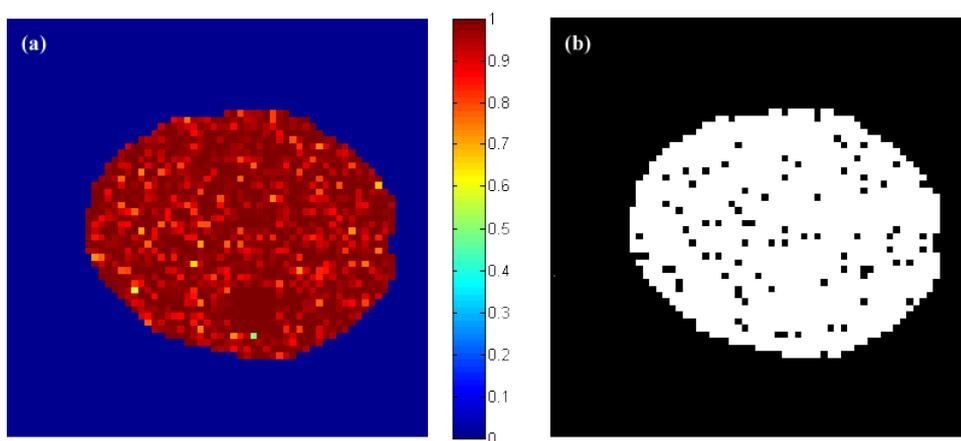


Figura 4.1: (a) Mapa entrópico dos dados artificiais; (b) Análise entrópica dos dados sintéticos. Os pontos pretos na imagem foram eliminados, o equivalente a 6% dos dados de entrada.

Posteriormente, iniciou-se a etapa de treinamento do SOM em um total de 100 iterações. Feito isso, o passo seguinte foi descobrir os agrupamentos que apresentassem padrões de sinal BOLD nos neurônios do mapa através do método de agrupamento hierárquico. A Figura (4.2a) mostra o resultado do algoritmo de agrupamento para os dados artificiais. Nele é possível ver

a formação de 2 grandes grupos (canto superior esquerdo e canto inferior direito) o que pode ser confirmado a partir da visualização da U -matrix em tons de azul escuro e em baixo relevo (Figura(4.2b)), resultado este que serviu para segmentar o mapa, destacando somente os neurônios candidatos aos padrões de ativação (Figura 4.3).

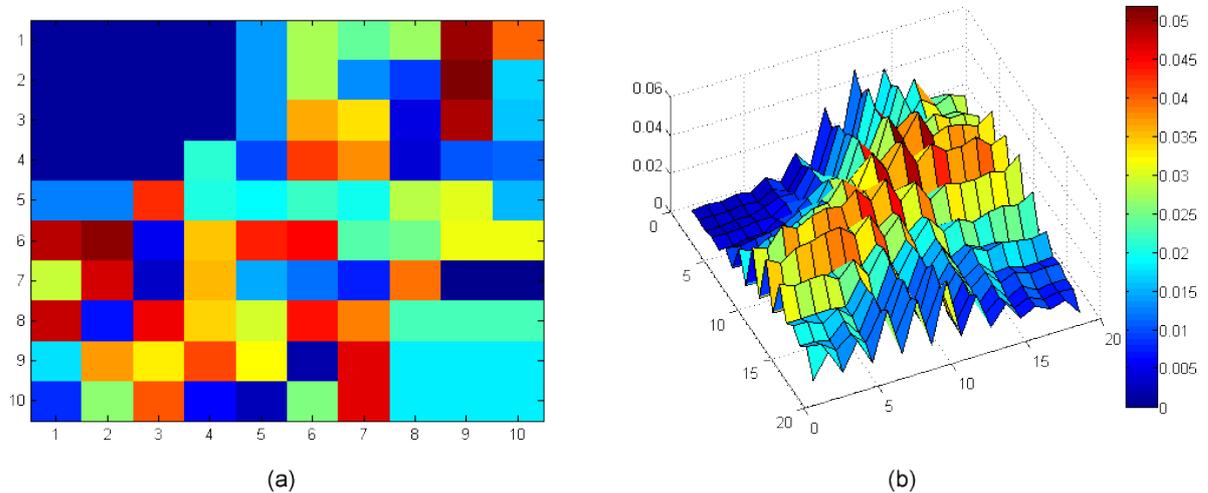


Figura 4.2: (a) Resultado do método de agrupamento hierárquico aplicado aos pesos do SOM, 72 agrupamentos foram formados; (b) Visualização tridimensional da U -matrix dos dados sintéticos.

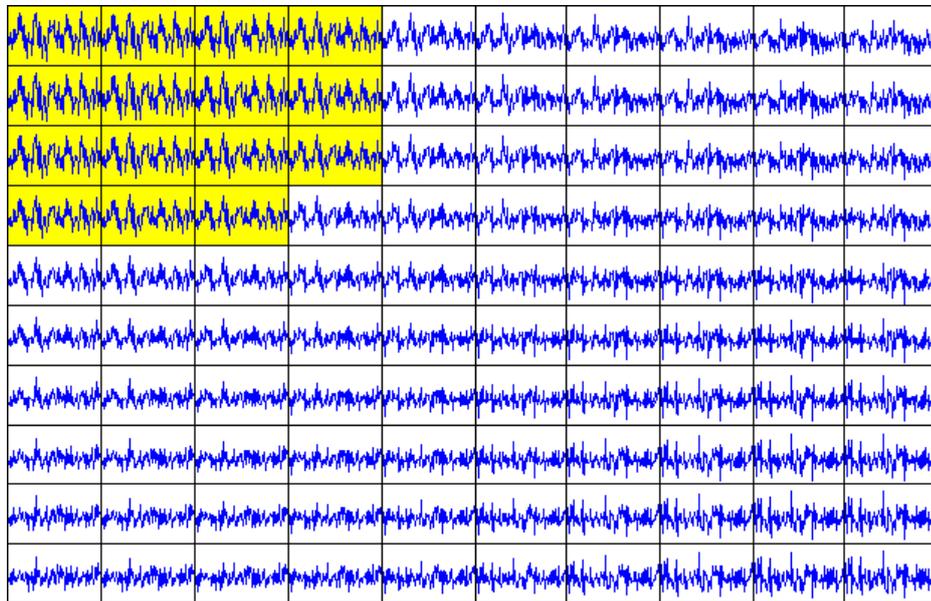


Figura 4.3: Configuração final do SOM, grade 10×10 para os dados sintéticos, o neurônios em destaque representam os padrões ativos dos dados de entrada.

Para gerar o mapa de ativações (Figura 4.5), é calculado um coeficiente de correlação entre todos os voxels das áreas correspondentes ao cérebro e a média dos pesos pertencentes ao agrupamento dos neurônios candidatos. A Figura (4.4) faz uma comparação entre média dos neurônios e a HRF atribuída aos dados simulados. O coeficiente de correlação entre estes dois

sinais é de 0.7056, valor este que foi usado como limiar para encontrar os voxels com maiores possibilidades de estarem ativos.

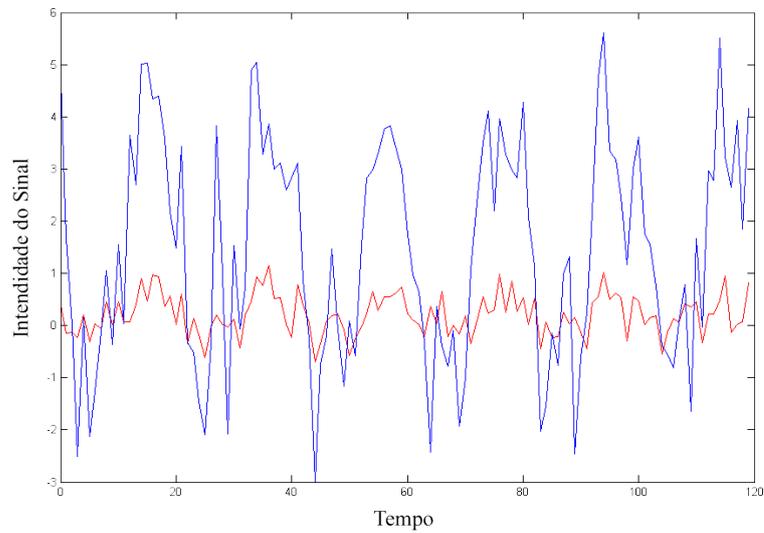


Figura 4.4: O sinal em azul corresponde a HRF com SNR de 5dB atribuída aos dados sintéticos. O sinal em vermelhos representa a média da série temporal dos neurônios candidatos a ativação.

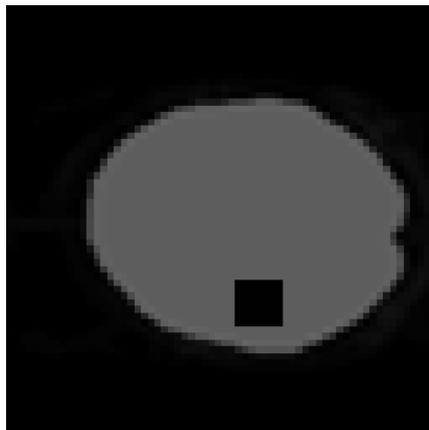


Figura 4.5: As regiões pretas no interior do cérebro correspondem às regiões ativas computadas pelo SOM com 100 iterações para um $CC > 0.7056$.

4.2 Dados Reais

Utilizamos o SOM para analisar os dados reais corresponde a um estímulo auditivo, como proposto na Seção 3.2. A Figura (4.6a) mostra a análise entrópica para os voxels da oitava fatia que variaram de 0.1161 a 1. Do total de voxels interiores ao cérebro, 8% foram eliminados do posterior treinamento do SOM (Figura 4.6b). Estes voxels apresentavam valor de entropia $H < 0.8$.

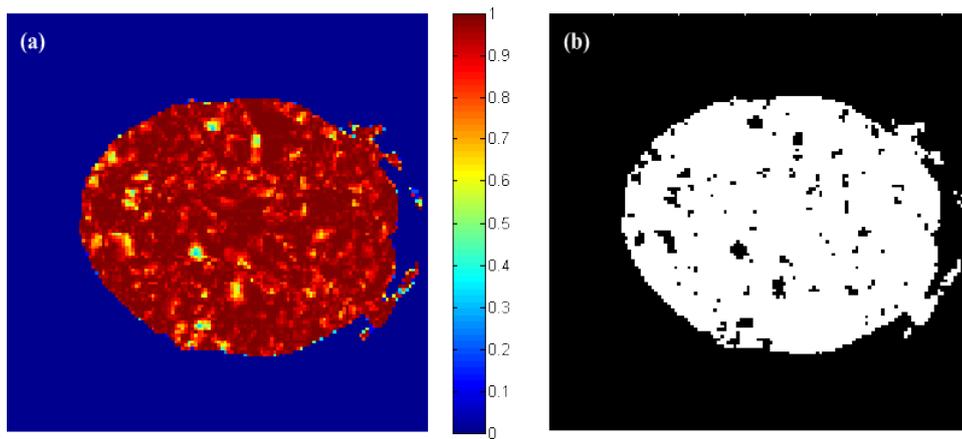


Figura 4.6: (a) Mapa entrópico dos dados reais da oitava fatia; (b) Análise entrópica dos dados reais. Os pontos pretos na imagem foram eliminados, o equivalente a 8% dos dados de entrada.

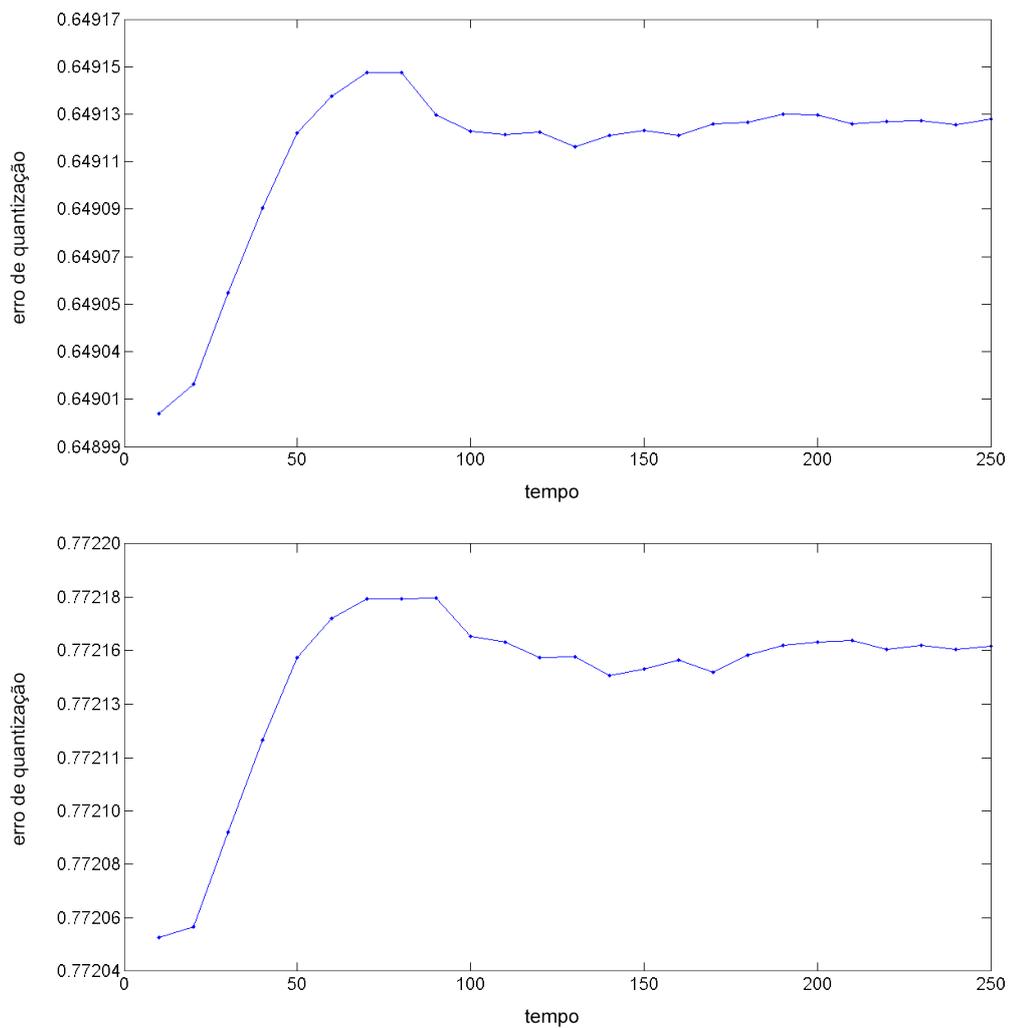


Figura 4.7: Gráfico do erro de quantização para um total de 250 iterações. (a) erro de quantização com aplicação de entropia; (b) erro de quantização sem entropia.

Diferente dos dados simulados, nos dados reais o conjunto de padrões de entrada é maior, embora a grade do SOM apresente a mesma quantidade de neurônios. Deduz-se, portanto, que o número de iterações na etapa do treinamento seja também maior para que o processo de aprendizado não seja comprometido. Assim sendo, adotou-se o erro de quantização (Eq. 1.17) para estimar a quantidade de passos do algoritmo e também atuar como controlador de qualidade do aprendizado. A Figura (4.7) mostra a evolução do erro a cada 10 iterações. Porém, ainda que o erro tenha começado a se estabilizar por volta das 100 repetições o treinamento estendeu-se até as 250 iterações com a finalidade de se realizar uma sintonia fina do mapa de características e assim produzir uma quantização estatística precisa do espaço de entrada. Ainda analisando a mesma figura, é possível verificar que a amplitude do erro para a situação em que foi aplicada a análise entrópica precedendo o SOM (Fig. 4.7a), apresenta-se menor além de ter começado a se estabilizar um pouco antes do que o caso em que não utilizou-se a entropia de Shannon (Fig. 4.7b).

Finalizado o processo de treinamento, pode-se observar o mapa de características referentes aos dados da oitava fatia (Figura 4.9) já segmentado após aplicação do agrupamento hierárquico (Figura 4.8a) e *U-matrix* (Figura 4.8b).

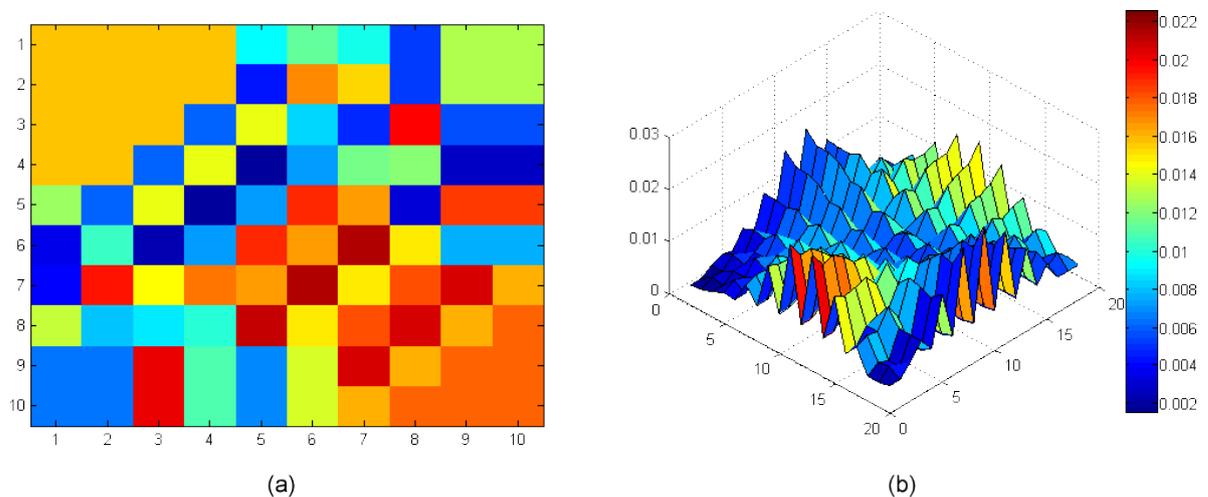


Figura 4.8: (a) Resultado do método de agrupamento hierárquico aplicado aos pesos do SOM, 49 agrupamentos foram formados; (b) Visualização tridimensional da *U-matrix* do experimento auditivo.

O mapa de ativação construído a partir do CC entre a série temporal de cada voxel do interior do cérebro nas imagens e a média dos pesos dos neurônios, que também é ilustrado, pode ser visto na Figura (4.10). Valores com CC superior a 0.6 foram destacados no mapa cerebral. Nele, é possível ver duas principais regiões, como resultado da tarefa auditiva, localizadas no lobo temporal. Segundo Huettel et al., (2004) o lobo temporal além de estar associado com

estímulos auditivos, também pode estar envolvido com memória e fala.

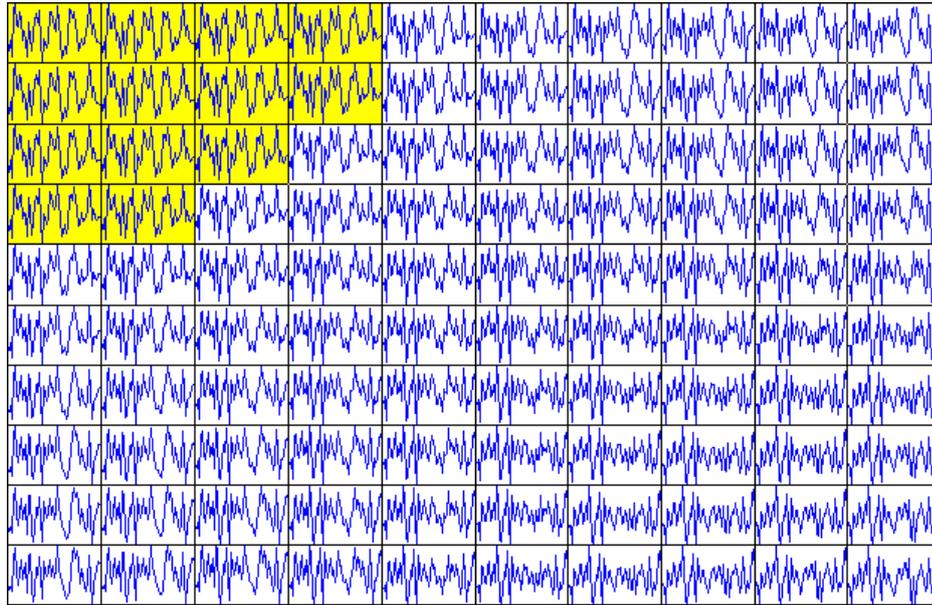


Figura 4.9: Configuração final do SOM (10×10) para os dados reais, o neurônios em destaque representam os padrões ativos dos dados de entrada.

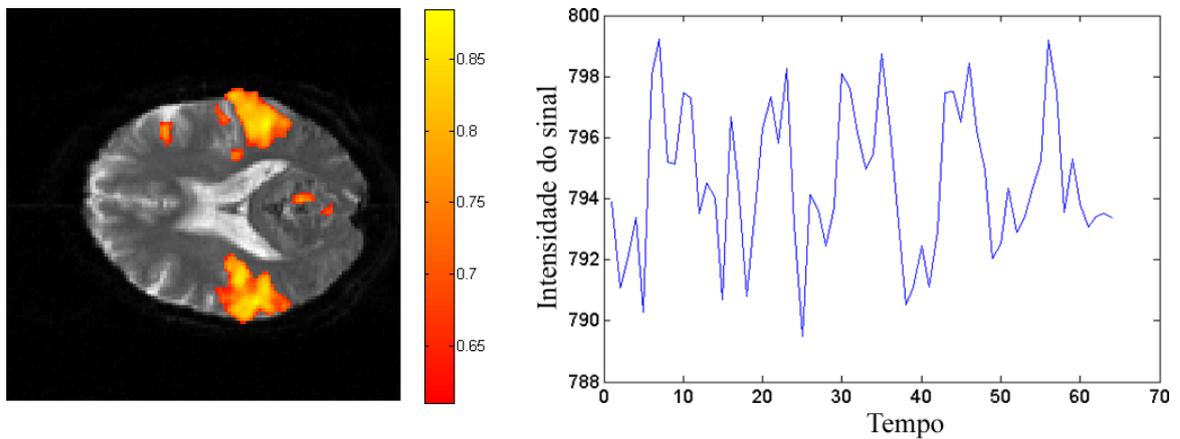


Figura 4.10: À esquerda, resultado de uma estimulação auditiva no córtex cerebral segundo análise do SOM ($CC > 0.6$). À direita, média da série temporal dos pesos dos neurônios ativos.

As figuras seguintes (Fig. 4.11) mostram o mapa de ativação, após a execução do SOM, para as fatias 9 e 10 dos dados reais. A Figura (4.12) mostra o resultado do mesmo conjunto de dados segundo análise do SPM, o qual é baseado no GLM.

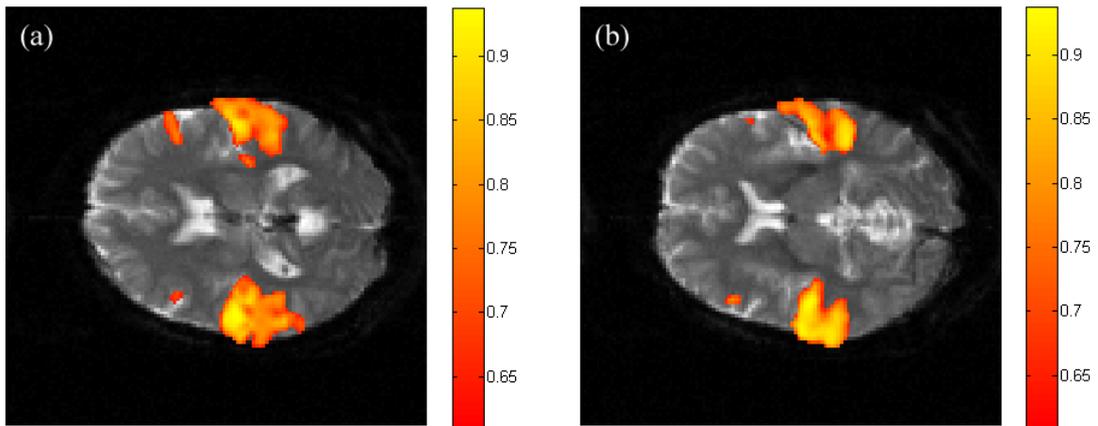


Figura 4.11: Resultado de uma estimulação auditiva no córtex cerebral segundo análise do SOM, fatias 9 e 10 ($CC > 0.6$).

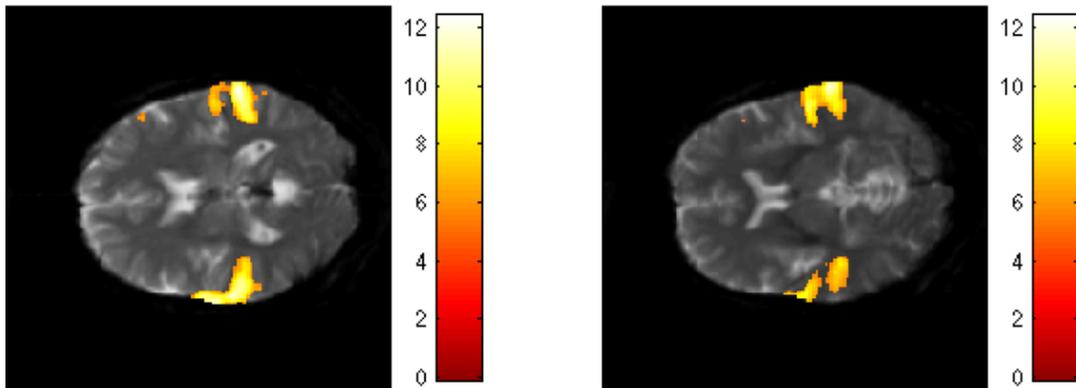


Figura 4.12: Resultado de uma estimulação auditiva no córtex cerebral segundo análise do SPM.

Conclusão

No presente estudo foi usado como ferramenta de análise de dados fMRI o mapa auto-organizável de Kohonen, um tipo de rede neural artificial com treinamento baseado na aprendizagem competitiva e que ordena topologicamente os padrões de dados incidentes em um mapa discreto. Também, com o intuito de segmentar padrões do mapa de características provenientes de diferentes fontes dos dados de entrada, utilizou-se agrupamento hierárquico auxiliado pela *U-matrix* na decisão do número de agrupamentos. Além disso, com o propósito de aumentar a eficiência do método de análises, foi proposta a entropia de Shannon, a qual eliminou uma faixa de 5-10% do conjunto de dados de entrada, que por sua vez estavam relacionados a voxels que apresentavam picos e estavam distribuídos de forma irregular. Como se pôde ver em dados simulados, a entropia relacionada a série temporal de voxels que continham evidências da resposta BOLD tende a ser máxima. Portanto, a configuração dos dados após a análise entrópica permite maiores chances de se encontrar grupos de neurônios ativos na grade do SOM com um menor número de iterações. Ademais, na evolução temporal do erro de quantização do SOM, pode-se verificar que a análise entrópica diminuiu a amplitude deste erro e admitiu sua estabilização sutilmente mais rápida.

Os resultados do SOM, tanto para dados simulados, quanto para dados reais, reafirmaram que ele pode ser usado como ferramenta de interpretação de dados fMRI. Além disso, como vantagem, ele não considera a forma da resposta hemodinâmica, não fazendo uso de uma HRF modelada matematicamente.

Como próximo passo para o desenvolvimento do método, pretende-se aplicar uma técnica que possa determinar, de maneira automática, o limiar da análise entrópica, permitindo decidir, com mais segurança, na eliminação de possíveis voxels irrelevantes. Além disso, pretende-se apresentar métodos estatísticos que possam mensurar o poder de detecção do método como o todo e validar os resultados obtidos.

Referências Bibliográficas

- [1] BALSLEV, D. NIELSEN, F. A., FRUTIGER, S.A., SIDTIS, J. J., CHRISTIANSEN, T. B., SVARER, C., STROTHER, S. C., ROTTENBERG, D. A., HANSEN, L. K., PAULSON, O. B., e LAW, I., 2002. **Cluster analysis of activity-time series in motor learning.** Human Brain Mapping, 15, 135-145.
- [2] BAUMGARTNER, R., WINDISCHBERGER, C. e MOSER, E., 1998. **Quantification in functional magnetic resonance imaging: Fuzzy clustering vs. correlation analysis.** Magnetic Resonance Imaging, 16, 115-125.
- [3] BISWAL, B. B., AND ULMER J. L., 1999. **Blind source separation of multiple signal sources of fMRI data sets using independent component analysis.** Journal of Computer Assisted Tomography, 23:265-271.
- [4] BULLMORE, E.T., BRAMMER, M.J., RABE-HESKETH, S., CURTIS, V.A., MORRIS, R.G., WILLIAMS, S.C.R., SHARMA, T., e McGUIRE, P.K., 1999. **Methods for diagnosis and treatment of stimulus-correlated motion in generic brain activation studies using fMRI.** Human Brain Mapping, 7, 38-48.
- [5] BUXTON, R. B., 2009. **Introduction to Functional Magnetic Resonance Imaging.** Principles and Techniques. Cambridge University Press, New York, USA.
- [6] BUSSAB, W. O., MIAZAKI, E. S., ANDRADE, D. F. **Introdução à análise de Agrupamento.** Associação Brasileira de Estatística. ABE 9º Simpósio Nacional de Probabilidade e Estatística. São Paulo.
- [7] CABELLA, B. C. T., 2008. **Inferência Estatística em Métodos de Análise de Resonância Magnética Funcional.** Dissertação de Mestrado, Universidade de São Paulo.
- [8] CAPURRO A., DIAMBRA L., LORENZO D., MACADAR O., MATRIN M.T., MOSTACCIO C., PLASTINO A., PÉREZ J., ROFMAN E., TORRES, M.E., VELLUTI J., 1999.

- Human brain dynamics: the analysis of EEG signals with Tsallis information measure.** *Physica A* 265: 235-254.
- [9] CHEN, H., YUAN, H. YAO, D., CHEN, L. e CHEN, W., 2006. **An integrated neighborhood correlation and hierarchical clustering approach of functional MRI**, *IEEE Transactions on Biomedical Engineering*, vol. 53, pp. 452-458, 2006.
- [10] COSTA, J.A.F. and NETTO, M.L.A., 2001. **A new tree-structured self-organizing map for data analysis**. In: *Proc. of the Intl. Joint Conf. on Neural Networks*, Washington, DC, pp. 1931-1936.
- [11] COSTA, J.A.F., e NETTO, M.L.A., 2007. **Segmentação de mapas auto-organizáveis com espaço de saída 3-D**. *Revista Controle e Automação* vol. 18 (2), pp. 151-162.
- [12] COVER, T. M., THOMAS J. A., 2006. **Elements of information theory** - 2 ed. A Wiley-Interscience publication. New Jersey, Canadá.
- [13] DALE, A. M e BUCKNER, R. L., 1997. **Selective averaging of rapidly presented individual trials using fMRI**. *Human Brain Mapping*, 5, 329-340.
- [14] DE ARAUJO, D.B., 2002. **Sobre neuroimagens funcionais por magnetoencefalografia e ressonância magnética: novos métodos e aplicações**. Tese (Doutorado). Universidade de São Paulo. FFCLRP.
- [15] DE ARAUJO D. B., TEDESCHI W, SANTOS AC, ELIAS J JR, NEVES UP, BAFFA O., 2003. **Shannon entropy applied to the analysis of event-related fMRI time series**. *Neuroimage*. 2003 Sep;20(1):311-7.
- [16] FADILI, M. J., RUAN, S., BLOYET, D., e MAZOYER, B., 2000. **A multistep unsupervised fuzzy clustering analysis of fMRI time series**. *Human Brain Mapping*, 10, 160-178.
- [17] FOUNTURA, et al., 2008. **Identificação de Regiões Cerebrais de Linguagem: Estudo de Ressonância Magnética Funcional em Pacientes com Epilepsia Refratária de Lobo Temporal**. *Journal of Epilepsy and Clinical Neurophysiology*, 14(1):7-10.
- [18] FISCHER, H., HENNING, J., 1999. **Neural-network based analysis of MR time series**. *Magnetic Resonance in Medicine* 41:124-131.

- [19] FRISTON, K. J., FRITH, C. D., LIDDLE, P. F., AND FRACKOWIAK, R. S. J., 1993. **Functional connectivity: the principal component analysis of large (PET) datasets.** *Journal of Cerebral Blood Flow and Metabolism*, 13:5-14.
- [20] FRISTON, K. J., JEZZARD, P., e TURNER, R., 1994. **Analysis of functional MRI time series.** *Human Brain Mapping*, 1:151-171.
- [21] FRISTON, K. J., HOLMES, A. P., WORSLEY, K. J., POLINE, J. P., FRITH, C. D., e FRACKOWIAK, R. S. J., 1995. **Statistical parametric maps in functional imaging: A general linear approach.** *Hum. Brain Mapping* 2: 189-210.
- [22] FRISTON, K.J., WILLIAMS, S.C.R., HOWARD, R., FRACKOWIAK, R.S.J., e TURNER, R., 1996. **Movement-related effects in fMRI time series.** *Magnetic Resonance in Medicine*, 35, 346-55.
- [23] FRISTON, K. J., FLETCHER, P., JOSEPHS, O. HOLMES, A., RUGG, M. D., e TURNER, R. 1998. **Event-related fMRI: Characterizing differential responses.** *NeuroImage*, 7, 30-40.
- [24] GLOVER, G. H. Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, 9:416-429,1999.
- [25] GIBBONS, R. D., LAZAR, N. A., BHAUMIK, D. K., SCLOVE, S. L., CHEN, H. Y., THULBORN, K. R., SWEENEY, J. A., HUR, K., AND PATTERSON, D., 2004. **Estimation and classification of fMRI hemodynamic response patterns.** *NeuroImage*, 22, 804-814.
- [26] GONÇALVES, M. L. et al., 2008. **Classificação não-supervisionada de imagens de sensores remotos utilizando redes neurais auto-organizáveis e métodos de agrupamentos hierárquicos.** In: *Revista Brasileira de Cartografia* No 60/01 (ISSN 1808-0936).
- [27] HAIR, J. F. et al., Tradução: SANT'ANNA, A. S., 2009. **Análise Multivariada de dados.** 6. ed. - Porto Alegre, Bookman.
- [28] HAYKIN, S., 2001. **Redes Neurais Princípios e prática.** 2 edição, Bookman, São Paulo.
- [29] HEBB, D. O., 1949. **The Organization of Behavior: A neuropsychological Theory.** New York: Wiley.
- [30] HU, X., LE, T. H., PARRISH, T., e ERHARD, P., 1995. **Retrospective estimation and correction of physiological fluctuation in functional MRI.** *Magnetic Resonance in Medicine*, 34, 201-212.

- [31] HUETTEL, S. A., SONG, A. W., e MCCARTHY, G., 2004. **Functional Magnetic Resonance Imaging**. Sinauer Associates, 1 edição.
- [32] JEZZARD, P., PAUL M. M. STEPHEN, S. M., 2001. **Functional MRI: An Introduction to Methods**. Oxford University Press Inc., New York, USA.
- [33] KOHONEN, T., 1982. **Self-Organizing formation of topologically correct feature maps**. *Biological Cybernetis*, vol. 43, pp. 59-69.
- [34] KOHONEN, T., 1990. **Self-Organizing Maps. Proceedings of the Institute of Electrical e electronics Engineers**, vol. 78, pp. 1464-1480.
- [35] KOHONEN, T., 2001. **Self-Organizing Maps**. 3 edição; Springer: Berlin.
- [36] LAZAR, N. A., 2008. **The Statistical Analysis of Funcional MRI Data**. Springer Science: USA.
- [37] LANGE, N. A. E ZEGER, S. L., 1997. **Non-linear Fourier time series analysis for human brian mapping by functional magnetic resonance imaging**. *Applied Statistics*, 46, 1-29.
- [38] LIAO, W., CHEN, H., YANG, Q., LEI, X., 2008. **Analysis of fMRI Data Using Improved Self-Organizing Mapping and Spatio-Temporal Metric Hierarchical Clustering**. *IEEE Transactions on Medical Imaging*, vol. 27, pp. 1472-1483.
- [39] LOGOTHETIS, N.K., 2002. **The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal**. *Philos. Trans. R. Soc. Lond B. Biol. Sci.* 357 (1424): 1003-1037.
- [40] MCKEOWN, M.J., MAKEIG, S., BROWN, G. G., JUNG, T. P., KINDERMANN, S. S., BELL, A. J., e SEJNOWSKI, T. J., 1998. **Analysis of fMRI data by blind separation into independent spatial components**. *Human Brain Mapping*, 6:160-168.
- [41] MULERT, C. e LEMIEUX, L., 2010. **EEG-fMRI Physiological Basis, Technique and Applications**. Springer-Verlag Berlin Heidelberg.
- [42] NGAN, S. C., YACOUB E. S., AUFFERNANN W. F, e HU X., 2002. **Node merging in Kohonen's self-organizing mapping of fMRI data**. *Artificial Intelligence in Medicine*, vol. 25, pp. 19-33.

- [43] OGAWA, S., LEE, T. M., KAY, A. R., e TANK, D. W., 1990. **Brain magnetic resonance imaging with contrast dependent on blood oxygenation**. Proceedings of the National Academy of Sciences of USA, 87:9868-9872.
- [44] PAULING, L. e CORYELL, C., 1936. **The magnetic properties and structure of hemoglobin, oxyhemoglobin, and carbon monoxyhemoglobin**. Proceedings of the National Academy of Sciences of USA, 22, 210-216.
- [45] PELTIER, S. J., POLK T. A., NOLL D.C., 2003. **Detecting low-frequency functional connectivity in fMRI using a self-organizing map (SOM) algorithm**. Hum. Brain Mapp., vol. 20, pp. 220-226.
- [46] RABE-HESKETH, S., BULLMORE, E., AND BRAMMER, M., 1997. **The analysis of functional magnetic resonance images**. Stat. Methods Med. Res. 6: 215-237.
- [47] SARTY, G. E., 2007. **Computing brain activity maps from fMRI time-series images**. Cambridge University Press, New York.
- [48] SAMARASINGHE, S., 2006. **Nerual Networks for applied sciences and engineering: from fundamentals to complex pattern recognition**. Taylor e Francis Group, LLC. New York.
- [49] STANBERRY, L., NANDY, R., e CORDES, D., 2003. **Cluster analysis of FMRI data using dendrogram shaperning**. Human Brain Mapping, 20, 201-219.
- [50] SHANNON, C. E., 1948. **A mathematical theory of communication**. Bell system technical journal 27, 379-423, 623-656.
- [51] SHEPHERD, G.,M., e KOCH, C., 1990. **Introduction to synaptic circuits**. In: The Synaptic Organization of the brain, G. M. Shepherd, ed., pp.3-31. New York: Oxfod University Press.
- [52] STURZBECHER, M. J., 2006. **Detecção e caracterização da resposta hemodinâmica pelo desenvolvimento de novos métodos de processamento de Imagens Funcionais por Ressonância Magnética**. Dissertação de Mestrado, Universidade de São Paulo.
- [53] THULBORN, K, R., WATERTON, J. C., MATTHEWS, P. M., e RADDI, G. K., 1982. **Oxygenation dependence of the tranverse relaxation time of water protons in whole blood at high field**. Biochimica et Biophysica Acta, 714, 265-270.

- [54] ULTSCH, A., 1993. **Self-Organizing Neural Networks for Visualization e Classification**. In: O. Opitz et al. (Eds). Information and Classification. Springer, Berlim, pp. 307-313.
- [55] VESANTO, J. e ALHONIEMI, E., 2000. **Clustering of the Self-Organizing Map**. IEEE Trans. on Neural Networks, vol. 11, (3), pp. 586-602.