



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

Liliane Ribeiro da Silva

**Aplicação da Decomposição em Valores Singulares e Análise
de Componentes Independentes em dados de fMRI**

Orientador: Prof. Dr. Marcus Pinto da Costa da Rocha

BELÉM- PA

2011

Liliane Ribeiro da Silva

Aplicação da Decomposição em Valores Singulares e Análise de Componentes Independentes em dados de fMRI

Dissertação apresentada do Programa de Pós-Graduação em Matemática e Estatística - UFPA, como requisito parcial para a obtenção do grau de Mestre em Matemática.

Área de Concentração: **Processamentos de Imagens Funcionais**

Orientador: **Prof. Dr. Marcus Pinto da Costa da Rocha**

BELÉM - PA

2011

Liliane Ribeiro da Silva

Esta Dissertação foi julgada pelo Corpo Docente do Programa de Pós-Graduação em Matemática e Estatística - PPGME da Universidade Federal do Pará, para a obtenção do grau de Mestre em Matemática.

Belém, 11 de Março de 2011

Prof. Dr. Giovany de J. M. Figueiredo
(Coordenador do PPGME - UFPA)

Banca Examinadora

Prof. Dr. Marcus Pinto da Costa da Rocha (**Orientador**)
Universidade Federal do Pará- UFPA-PPGME

Prof. Dr. Valcir João da Cunha Farias (**Co-orientador, Membro**)
Universidade Federal do Pará- UFPA-PPGME

Prof. Dr. Héilton Ribeiro Tavares(**Membro**)
Universidade Federal do Pará- UFPA-PPGME

Prof. Dr. Regivan Hugo Nunes Santiago(**Membro**)
Universidade Federal do Rio Grande do Norte- UFRN-DIMAP

Dedicatória

À minha querida mãe, mãe de sangue e de alma, minha referência de força e de coragem, minha base. À mulher mais guerreira que já conheci sobre a Terra, cujo caráter é meu exemplo. Meu colo, minha palmatória, minha amiga, meu espelho ferino, minha veia exposta, minha eterna companheira em todas as jornadas. À sua contrariedade muda e apoio irrestrito. Para que nunca se esqueça de meu amor sobejo e inexorável.

Agradecimento

A Deus que é onipresente, onisciente e onipotente;

Ao meu tio-pai, Manoel do Carmo, por seu incentivo e crédito, quando tudo era apenas um sonho. A sua dedicação e preocupação com meu futuro e principalmente comigo e companheirismo incansável.

Ao meu tio-pai, José Ribeiro, pelo apoio incondicional junto à minha mãe para que nada tira-se minha atenção dos estudos e por todo seu apoio material e sentimental.

Ao meu orientador-pai-amigo, Marcus Rocha (Marcão), por seu irretocável empenho em construir um laboratório de excelência num cenário de recursos escassos e objeções políticas. Por sua amizade e dedicação, não somente comigo, mas com todos os colaboradores do Laboratório de Avaliações e Medidas (LAM).

Ao meu amigo-orientador-padaastro, Valcir João, brilhante em inteligência e retidão, que me enche de orgulho por sua orientação e, orgulho por sua amizade. Obrigado pela amizade e apoio nos momentos mais difíceis e por me ensinar a olhar sempre com esperança e confiança para todas as situações.

À minha amiga, Lucélia Lima, sempre por perto para colar os caquinhos do meu frágil coração. Por sua admirável dedicação aos estudos e por sempre acreditar que posso ir mais além. agradeço por tornar meus dias mais leves com sua alegria e amizade infinita.

À minha amiga, Lidiane Carvalho, por seu apoio emocional e amizade incondicional e por alegrar todos os meus dias com suas mensagens de incentivo e perseverança.

Aos meus amigos Anderson Ramos, João Rodrigues, Lindalva, Marcel e Walter, que fizeram parte do meu crescimento acadêmico e pela amizade construída e solidificada durante esses dois anos de convívio diário.

Ao meu amigo, Edinaldo Alencar, pelo seu apoio, suporte e prontidão, sempre recheado com idéias notáveis, base de sua incansável dedicação à programação e aos amigos.

Ao meu irmão, Luiz Augusto, por entender que mesmo na minha ausência nunca deixei de me

preocupar com sua formação.

Aos meus amigos do LAM, pelo apoio e companheirismo durante os quatro anos que faço parte dessa Família LAM.

Aos meus familiares, em especial a minha prima Cleide, a minha tia Fátima e meu primo-irmão Élido, por sempre acreditarem em meus sonhos e demonstrarem preocupação e disponibilidade em todos os momentos.

Aos professores do ICEN, por todo o aprendizado e amizade adquiridos durante esses 7 anos de trabalho e estudo nessa instituição.

Aos professores da banca, Regivan Nunes e Héilton Ribeiro, pelo interesse e colaboração em diversas situações.

Aos meus queridos Luiz Ranyer e Michelle Lúcia, mesmo a quilômetros de distância, pelo companheirismo e incentivo para a finalização dessa etapa em minha vida.

A todos os meus amigos, pelo carinho apesar de minha ausência.

Ao CNPq, pelo apoio financeiro durante esse curso.

"A dúvida permite extrair um núcleo de certeza, que cresce à medida que ela se radicaliza; é indubitável que, se duvido, penso."

Descartes

Sumário

Resumo	9
1 Introdução	10
2 Imagem Funcional por Ressonância Magnética	12
2.1 Aquisição das Imagens	12
2.2 Paradigmas de Aquisições das Imagens	14
2.2.1 Paradigma em Blocos	14
2.2.2 Paradigma Evento-Relacionado	14
2.3 Pré-processamento das Imagens	15
2.3.1 Realinhamento	16
2.3.2 Normalização Espacial	16
2.3.3 Suavização	17
2.3.4 Segmentação	18
3 Análise de Componentes Independentes	19
3.1 Levantamento Histórico	19
3.2 Modelo e Definição de ICA	22
3.3 Restrições do Método ICA	23
3.3.1 Independência Estatística	23
3.3.2 Não Gaussianidade das Variáveis	25
3.4 Técnicas de Pré-Processamento	26
3.4.1 Centralização	26
3.4.2 Branqueamento	27
3.5 Princípios para Estimativa de Componentes Independentes	28
3.5.1 Maximização da Não Gaussianidade	29

3.5.2	Decomposição em Valores Singulares	32
3.5.3	Algoritmo Rápido de Ponto Fixo (FastICA)	35
4	Resultados	37
4.1	Dados Sintéticos	37
4.2	Dados Reais	46
5	Considerações Finais	48
	Referências	49

Resumo

No Brasil o desenvolvimento de técnicas de processamento de sinais funcionais vem crescendo consideravelmente. A técnica de imagens funcionais por ressonância magnética (fMRI) é usada para localizar e analisar diferentes

aspectos da atividade cerebral humana. Durante uma experiência de fMRI, uma seqüência de imagens é adquirida enquanto o paciente desempenha tarefas específicas. No entanto, as alterações no contraste das imagens não são grandes, o que impossibilita uma inspeção visual direta, tornando necessária a utilização de algoritmos computacionais de identificação dessas áreas. Assim, neste trabalho, nossos esforços são direcionados para detectar níveis de atividade em regiões cerebrais. Nesse sentido, estudou o método de Análise de Componentes Independentes (ICA), com o diferencial da aplicação da Decomposição em Valores Singulares (SVD) nos dados já pré-processados utilizando dados artificiais, para avaliar a sensibilidade e a especificidade do método. As curvas ROC indicam que o método tiveram um melhor desempenho após a utilização da SVD. Além disso, ICA foi aplicada a um conjunto de dados reais de fMRI onde um indivíduo normal realizou um experimento pré-definido. As áreas identificadas corresponderam a audição e foram consistentemente associadas à tarefa. O método ICA não precisa de parâmetros pré-acordados o que melhora o seu desenvolvimento para as estimativas do seus componentes independentes deixando livre a quantidade de componentes a serem estimados.

Palavras-chaves: Imagens ressonância magnética funcional, decomposição de valores singulares, análise de componentes independentes.

Capítulo 1

Introdução

Em diversas situações práticas são utilizados sensores para coletar informações, geralmente há o problema de misturas das fontes (sinais) desejadas na captação dos sinais.

Na maioria das situações práticas, não há como observar as fontes diretamente, e também não é conhecida a forma como as quais foram misturados. Esse tipo de problema é conhecido, na literatura, como o problema da Separação Cega de Fontes (BSS- Blind Source Separation). A BSS tem sido alvo de muitos trabalhos em diversas áreas como a apresentada nesta dissertação.

Um dos métodos mais desenvolvido da BSS é a Análise em Componentes Independentes (ICA). A ICA é um método que é utilizado para recuperar um conjunto de fontes a partir de dados multi-dimensional sendo essas estatisticamente independentes e não-gaussianas.

A fMRI é um método de aquisição desse tipo de dados. Onde existe a possibilidade de localizar, registrar e estudar os diferentes aspectos da atividade cerebral de forma não invasiva (uma das principais vantagens da fMRI).

A fMRI tem se tornado ao longo dos anos uma das principais ferramentas para mapear o funcionamento do cérebro e seus padrões de ativação. Nas primeiras experiências com a técnica de fMRI foram usados agentes de contraste externo, mas, com o desenvolvimento dos procedimentos, o próprio sangue passou a ser usado como um agente de contraste.

Após a aquisição das imagens funcionais é feito o pré-processamentos. A partir disso, inicia-se as análise estáticas para aferir quais regiões estão sendo ativadas durante uma tarefa pré-definida, executada pelo paciente. Depois são realizadas as análises para detectar as regiões do cérebro responsáveis pela execução da tarefa, neste trabalho a técnica utilizada foi a técnica de ICA, com o diferencial de aplicar a SVD, para otimizar o processo computacional e assim melhorar a qualidade dos resultados.

No capítulo 2 descreve-se a ressonância magnética bem como o processo de aquisição das imagens, os dois tipos de arranjos experimentais. O pré-processamento das imagens após a etapa de aquisição.

No capítulo 3 é feita a apresentação histórica do ICA. É apresentado o modelo e definição de ICA, bem como as suas principais restrições. Além disso, apresenta-se técnica de pré-processamento em ICA. Neste capítulo é exposto os princípios para estimativa das componentes independentes, assim como algoritmo utilizado, FastICA.

No capítulo 4, são apresentados os principais resultados, indo da análise das curvas ROC (Receiver Operating Characteristic) obtidos por simulações para medir o desempenho do método de ICA, como a análise dos resultados obtidos com dados artificiais e reais.

O capítulo 5, contém as considerações finais, bem como as perspectivas futuras dos trabalhos desenvolvidos neste.

Capítulo 2

Imagem Funcional por Ressonância Magnética

A imagem funcional por ressonância magnética (fMRI) 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 [31]. 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 [11].

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 dificultam a inspeção visual direta das áreas de atividade cerebral, tornando necessária a utilização de métodos computacionais de análises.

2.1 Aquisição das Imagens

O efeito BOLD (Blood Oxigenation Level Dependent) tem sido amplamente usado nas recentes técnicas de fMRI. O fenômeno, que foi primeiramente descrito em um estudo com ratos, tem origem

nas propriedades magnéticas de oxigenação e desoxigenação do sangue [32]. Em [34] foi 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.

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 desoxihemoglobina, sendo ela paramagnética, gera mudanças na susceptibilidade magnética local, a qual varia linearmente com a oxigenação do sangue [6], provocando perturbações no sinal da ressonância magnética com alteração dos tempos de relaxação T_2 e T_2^* [37]. Estas mudanças são detectadas nas imagens de ressonância magnética através de pequenas mudanças na intensidade de brilho dos voxels ativos [7]

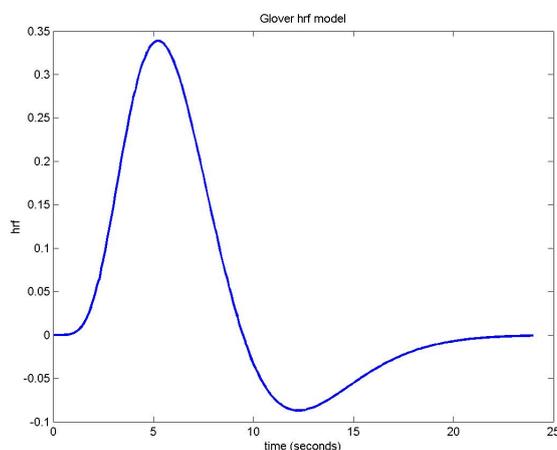


Figura 2.1: Função Resposta Hemodinâmica para um estímulo de curta duração. Imagem retirada e adaptada a partir de Mulert & Lemieux, 2010

A Figura (2.1) mostra a evolução temporal de uma ativação neural, associada a uma tarefa ou atividade cerebral espontânea, (HRF - Hemodynamic Response Function). É possível, dependendo do experimento, verificar um breve decréscimo do sinal logo no início. Posteriormente, começa uma etapa de resposta positiva BOLD mantendo-se de 5-10s, em alguns experimentos, 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.

O processo de aquisição das imagens depende de um paradigma selecionado previamente para realizar a aquisição dos dados durante o exame, o que é detalhado na seção seguinte deste trabalho.

2.2 Paradigmas de Aquisições das Imagens

Um paradigma é uma sucessão de eventos que consiste em intercalar períodos de estímulo e ausência de estímulo, provenientes da aplicação, periódica ou não, de tarefas específicas executadas pelo paciente ou voluntário durante a sessão de aquisição das imagens funcionais. Ele pode ser de vários tipos como por exemplo: em blocos, evento relacionado ou paramétrico. Neste contexto, preferiu-se abordar apenas os paradigmas em blocos e evento relacionado, que são mais comumente estudados.

2.2.1 Paradigma em Blocos

Em um arranjo experimental em blocos, uma certa quantidade de imagens são adquiridas em blocos ou períodos alternados de atividade e de repouso. Geralmente, propõe-se que a aquisição das imagens em cada bloco dure algo em torno de 25s a 30s, de modo que a duração destes períodos seja longa o suficiente para acomodar a resposta hemodinâmica [1]. De acordo com o tipo de estímulo realizado pelo paciente/voluntário e observando o curso temporal das imagens, espera-se que as alterações no contraste resultante da alternância entre estímulo e ausência dele apareçam de forma mais evidente nas regiões efetivamente ativadas durante a tarefa executada, evitando a contaminação por outros processos cerebrais. Naturalmente, existem outras funções cerebrais involuntárias, tais como atenção, intenção de fazer ou agir, e emoções, que podem interferir no sinal de interesse.



Figura 2.2: Modelo de um experimento em blocos alternado ausência de estímulo e estímulo.

A Figura (2.2) mostra um modelo de um paradigma em blocos alternando quatro período de repouso e três de atividades, começando com repouso, [12].

2.2.2 Paradigma Evento-Relacionado

O paradigma de evento-relacionado (ER) é pouco utilizado, no entanto, nos últimos anos vem ganhando espaço e se mostrando uma ferramenta muito eficiente para detectar a resposta BOLD à

atividade neuronal. Diferentemente do arranjo em blocos, onde os estímulos são longos e se repetem periodicamente, neste tipo de arranjo estímulos ou eventos são executados em períodos curtos de tempo (geralmente de 1s a 2s) em seqüências arbitrárias. Cada evento é separado no tempo dos eventos anteriores com um intervalo entre os estímulos que pode variar de aproximadamente 2s a 20s dependendo dos objetivos da experiência, [1].

A aplicação de ER-fMRI tem a vantagem de possibilitar a mistura aleatória de tarefas de diferentes tipos em uma única série temporal, além da menor contaminação dos dados por artefato cognitivo. Por conseguinte, esse tipo de arranjo experimental vem crescendo bastante nos últimos anos no contexto de experimentos cognitivos. A figura (2.3) mostra o modelo evento relacionado:

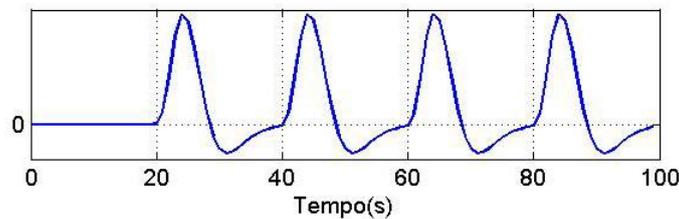


Figura 2.3: Modelo de um experimento evento relacionado alternado longos períodos ausência de estímulo e curtos períodos de estímulo.

2.3 Pré-processamento das Imagens

A técnica de fMRI é suscetível a várias fontes de ruído, causando nos dados do experimento imprecisões temporais e espaciais, como já mencionado neste texto. Dentre as fontes que causam essas imprecisões, podemos citar os seguintes ruídos: movimento de cabeça de forma involuntária, batimento cardíaco, respiração ou outros processos fisiológicos, diferenças no tempo de aquisição das imagens, imperfeições do hardware, interferência de outros estímulos não relacionados à tarefa de interesse. Existe também o ruído térmico que é proveniente da movimentação dos elétrons tanto no sujeito como nos componentes eletrônicos do scanner. Uma maior movimentação dos elétrons implica em mais colisões e conseqüentemente um sinal mais distorcido.

A partir das técnicas de pré-processamentos podemos suavizar a maioria dos ruídos citados acima, existem diversas etapas para realizar o pré-processamento das imagens, de tal forma que ao aplicarmos as técnicas para aferir os diagnósticos dos dados analisados tenhamos resultados com um acentuado grau de representatividade. Problemas de ruído podem até invalidar os resultados de todo o processo.

A análise de neuroimagens geralmente inicia com uma série de transformações que visam reduzir

a variação indesejada nos componentes dos voxels de uma série de imagens, que ocorrem por motivos já citados. O primeiro passo é o realinhamento das imagens para eliminar os efeitos de eventuais movimentos da cabeça do paciente/voluntário durante a aquisição das imagens.

2.3.1 Realinhamento

Variações na intensidade do sinal de qualquer voxel podem indicar a movimentação da cabeça do paciente, o que representa um sério fator de confusão (especialmente em estudos de fMRI). Apesar da restrição do espaço de movimentação da cabeça na aquisição das imagens, a co-operatividade dos sujeitos podem provocar o deslocamento de vários centímetros. O processo de realinhamento de uma série consiste de duas etapas. A primeira delas é a estimação de um conjunto de seis parâmetros de transformação de um corpo rígido (três rotações e três translações). Para cada imagem, este processo é feito através do cálculo dos parâmetros que minimizam a diferença quadrática média entre ela e uma imagem de referência. A segunda etapa é a aplicação da transformação por re-amostragem dos dados usando interpolação tri-linear, sinc ou spline.

Para a maioria das modalidades de aquisição de imagens este procedimento é suficiente, entretanto, para fMRI, mesmo após o realinhamento, a presença de movimentos no sinal podem persistir. Isto pode ser resolvido com o ajuste dos dados para efeitos residuais de movimento. [15].

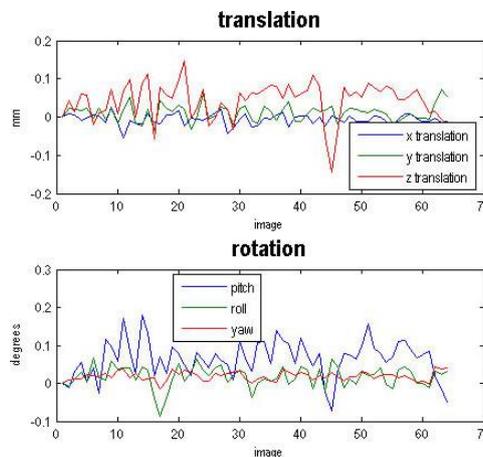


Figura 2.4: Realinhamento espacial de um experimento auditivo. Figura obtida após a etapa de realinhamento realizada no SPM8 (Statistical Parametric Mapping), [7].

2.3.2 Normalização Espacial

Após o realinhamento das imagens, uma imagem média da série, ou alguma da própria série pode ser usada para estimar alguns parâmetros de deformação da imagem através de um modelo

definido em um espaço anatômico padrão. Essa estimação utiliza uma variedade de modelos para o mapeamento, incluindo: (1) uma transformação afim com 12 parâmetros que constituem uma matriz de transformação espacial, (2) funções espaciais base de baixa frequência (geralmente um conjunto discreto de senos ou polinômios), onde os parâmetros são os coeficientes das funções base empregadas, e (3) um vetor especificando o mapeamento para cada ponto controle.

Uma consideração especial é a normalização espacial de cérebros que têm alguma lesão anatômica grosseira. Esta lesão pode ser de duas espécies: (1) variações quantitativas na estrutura de um tecido específico (por exemplo, atrofia cortical) ou (2) variações qualitativas na anatomia envolvendo a inserção ou exclusão de estruturas de tecidos normais (por exemplo, displasia cortical). O primeiro caso geralmente não é problemático, uma vez que as variações entre o tecido cortical não afetarão sua localização espacial quando referenciado por algum modelo (e mesmo que isso aconteça, um modelo da doença específica é facilmente construído). A segunda espécie de patologia pode introduzir erros substanciais na normalização ao menos que precauções especiais sejam tomadas, que geralmente envolvem restrições na deformação para assegurar que a patologia não seja um “viés” na deformação do tecido não danificado. Uma estratégia alternativa é usar outra modalidade que seja menos sensível à patologia como a base do procedimento de normalização espacial ou para simplesmente remover a região danificada da estimativa por uma máscara, [15].

2.3.3 Suavização

A suavização é um processo pelo qual são calculadas as médias dos pontos com seus respectivos vizinhos em uma série, tal como em uma série temporal, ou uma imagem. Geralmente isto tem o efeito de borrar as bordas. A suavização algumas vezes é referida como sendo um filtro, pois tem o efeito de suprimir o sinal de alta frequência e realçar os pontos de baixa frequência espacial. Há diferentes métodos de suavização, porém o SPM utiliza apenas a suavização através de um Kernel Gaussiano.

Um kernel para suavização define a forma da função que será usada para calcular a média dos pontos vizinhos. O Kernel Gaussiano é um kernel com a forma da curva Gaussiana (distribuição normal).

Na estatística padrão é definido que a largura da Gaussiana está relacionado ao termo σ , entretanto quando a Gaussiana é utilizada para suavização é usual descrever a largura da Gaussiana com outra medida relacionada, conhecida como Full Width at Half Maximum - FWHM, ou largura máxima à meia altura.

2.3.4 Segmentação

Por fim, a segmentação é um processo que consiste na subdivisão de uma imagem de ressonância magnética em regiões distintas, levando-se em consideração a descontinuidade e similaridade dos tecidos cerebrais . O objetivo principal é que cada região obtida esteja diretamente relacionada com um objeto real contido na imagem, como por exemplo um tumor. Em fMRI a segmentação tem basicamente o objetivo de fazer a diferenciação entre o que é vaso sanguíneo e o que é tecido ativo. Afinal, vaso sanguíneo interpretado como região ativa, constitui-se num tipo de artefato muito comum com grande potencial de produzir atividade indevida (falsos positivos), [1].

Capítulo 3

Análise de Componentes Independentes

O principal objetivo da ICA é encontrar uma representação linear de dados não gaussianos, minimizando a dependência estatística entre eles de forma que os componentes resultantes sejam estatisticamente independentes, ou tão independentes quanto possível. O que diferencia a ICA das outras técnicas existentes é justamente o fato dela trabalhar com componentes que são ao mesmo tempo não gaussianos e estatisticamente independentes. Esta representação pode ser usada para capturar a estrutura das componentes, ou dados, em várias aplicações como extração de características ou separação de sinais.

Neste capítulo será realizada uma revisão do estado da arte dos princípios da Análise de Componentes Independentes, bem como uma revisão dos principais algoritmos que utilizam a ICA com o objetivo de realizar a separação cega de sinais. É apresentada ainda uma revisão de conceitos estatísticos indispensáveis para a compreensão do problema da ICA.

3.1 Levantamento Histórico

As pesquisas na área de processamento cego de fontes (BSP - Blind Source Processing) tiveram início na década de 80, principalmente na França onde surgiram os primeiros trabalhos na área. Em abril de 1986, no encontro sobre Redes Neurais para Computação realizado em Snowbird (Utah, EUA) Jeanny Herault e Christian Jutten apresentaram um artigo intitulado Space or time adaptive signal processing by neural network models, que descrevia uma rede neural recorrente e um algoritmo de treinamento baseado nas regras de Hebb, que segundo os autores era apta para realizar a separação cega de sinais independentes. Como exemplo, foi apresentada a separação de dois sinais misturados e foi mencionada a possibilidade de separação de sinais de estereoscópios com quatro sinais misturados. Aquele trabalho abriu as portas para uma nova área dentro do processamento de sinais, [27].

Ao longo dos anos 80, as técnicas para ICA, estudadas principalmente por pesquisadores franceses, tiveram pouca influência no cenário internacional. Os poucos trabalhos apresentados em congressos de Redes Neurais em meados dos anos 80 foram ofuscados pelo grande número de trabalhos relacionados a redes de retropropagação, redes de Hopfield e Mapas Auto-Organizáveis de Kohonen.

No final da década de 80, foram apresentados os primeiros artigos relacionados a ICA [8,10], que foram de grande importância para o avanço das técnicas de separação cega de fontes utilizando estatísticas de ordem superior, em especial cumulantes de quarta ordem, aliadas aos princípios da Teoria da Informação.

Desta forma, pode-se dizer que até meados dos anos 90 poucos trabalhos foram apresentados na área de ICA e os apresentados não atraíram de forma significativa o interesse da comunidade científica. Por isso, em muitos casos, o problema da separação cega foi considerado sem solução.

Em meados dos anos 90, os pesquisadores da área de Processamento de Sinais voltaram sua atenção para os algoritmos não supervisionados. Em 1994, a estrutura geral foi revisada e aperfeiçoada por [10]. Neste trabalho, Comon desenvolveu uma teoria matemática sobre o problema da separação cega de fontes utilizando Análise de Componentes Independentes. Em [10], prova que o critério de independência estatística entre fontes originais é suficiente para a obtenção da separação das fontes sujeitas a misturas lineares. A definição de Análise de Componentes Independentes apresentada neste artigo depende de uma função custo.

Embora a aplicação da ICA à separação cega de fontes apresente algumas limitações, como, por exemplo, há necessidade de conhecer ou estimar a função densidade de probabilidade das fontes originais, o trabalho de Comon consagrou a Análise de Componentes Independentes como uma importante ferramenta para a separação cega de fontes. Bell e Sejnowski, [5], publicaram um artigo onde foi apresentado um algoritmo baseado no princípio de Maximização da Informação (INFOMAX) que atraiu as atenções para a área de ICA por colocar o problema da separação cega de fontes dentro de uma estrutura neural baseada em princípios de Teoria da Informação.

Bell (1996) colaborou a distância com Lee (1998) em seu trabalho para separação cega de fontes acústicas gravadas levando em consideração atrasos de tempo.

Cardoso e Laheld (1996a), uma melhoria no algoritmo desenvolvido por Bell e Sejnowski (1995), permitindo que este algoritmo se tornasse prático com aplicações em vários problemas reais. O trabalho de Cardoso e Laheld (1996a) introduziu uma classe de algoritmos adaptativos, que implementam uma versão da estimativa de equivariância e foi denominado EASI (Equivariant Adaptive Separation via Independence). Esta classe de algoritmos é baseada na idéia de atualização serial (serial updating) da matriz de separação, que geram algoritmos com estruturas simples. Outro importante

ponto deste algoritmo é que o seu desempenho independe da matriz de mistura, dependendo somente da distribuição normalizada das fontes.

Amari(1996), de forma independente, refinaram o algoritmo desenvolvido por Bell e Sejnowski(1995) utilizando gradiente natural, o que aumentou a velocidade de convergência do algoritmo, pois eliminou a inversão de matriz.

Um ano mais tarde, Hyvärinen e outros pesquisadores, [27], da Universidade de Tecnologia Helsinque na Finlândia, apresentaram o algoritmo de ponto fixo para ICA ou FastICA, que contribuiu para a utilização da ICA em diversas aplicações devido a sua eficiência computacional (Hyvärinen, 2000; Hyvärinen, 2001, Hyvärinen, 1997). Este algoritmo foi proposto com base nas técnicas desenvolvidas pelos pesquisadores para estimar de forma simples a f.d.p. dos sinais originais a partir das amostras recebidas pelos sensores.

Outro algoritmo também de grande importância foi proposto por Lee et al. (1997), que desenvolveram uma versão generalizada ou estendida do INFOMAX, denominado Extendend-Infomax (ou E-Infomax), que estimava os momentos das fontes originais e chaveava o algoritmo de acordo com o tipo de fonte super ou sub-gaussiana. As principais aplicações deste algoritmo encontram-se na área biomédica, especialmente para o processamento de sinais de eletroencefalograma (EEG), onde o algoritmo tem se mostrado adequado para separação dos sinais de EEG de artefatos.

Cardoso(1998), [8], apresentou uma importante revisão do ponto de vista estatístico, das principais técnicas que haviam sido desenvolvidas até o momento para solucionar o problema da separação cega de sinais utilizando ICA. Este trabalho teve grande repercussão e serve como uma boa introdução matemática ao problema. Outra boa introdução ao problema da ICA foi apresentada por Hyvärinen e Oja (2000). Este trabalho descreve de forma simples os principais conceitos da ICA, bem como suas principais aplicações.

Em 1999, foi criado na Europa um projeto, não governamental, de pesquisa denominado BLISS (Blind Source Separation and Applications) (BLISS, 2003). Como o próprio nome sugere, o projeto é voltado para pesquisas em BSS e ICA com o objetivo de desenvolver trabalhos considerando misturas instantâneas e convolutivas, presença de ruído, métodos adaptativos e/ou não lineares. Os trabalhos desenvolvidos visam, em geral, aplicações nas áreas de telecomunicações e engenharia biomédica. Este projeto, ou consórcio, é composto atualmente por cinco países incluindo Portugal (INESC), Finlândia (Helsinki University of Technology), França (Institute National Polytechnique de Grenoble), Alemanha (GMD) e Canadá (McMaster University).

Portanto, desde meados dos anos 90, tem crescido o número de publicações e congressos destinados a ICA. O primeiro congresso internacional sobre ICA aconteceu em Janeiro de 1999 em Aussois,

França e o segundo em Junho de 2000 em Helsinki, Finlândia. Estes congressos contaram com mais de 100 pesquisadores que trabalham em ICA e BSS e que contribuíram para transformar ICA em uma área de pesquisa bem definida.

Em 2003, a Conferência Internacional sobre Análise de Componentes Independentes e Separação Cega (ICAZ2003) foi realizado de 1-4 de abril em Nara, Japão, atraindo pesquisadores de diferentes áreas de pesquisa, entre as quais se destacaram as áreas de engenharia biomédica, processamento de voz e comunicações móveis. Além das aplicações, foram apresentadas dezenas de novos algoritmos ou versões modificadas dos algoritmos existentes (ICA, 2003).

Em junho de 2003, foi organizada em Berlim, pelo projeto BLISS, uma escola de verão que discutiu a teoria e as aplicações da ICA e BSS (ICA-SS, 2003). A escola contou com a participação de importantes pesquisadores na área como Luis Almeida (INESC, Portugal), J. F. Cardoso (ENST, França), Simon Haykin (McMaster, Canadá), Christian Jutten (INPG, França), T-W. Lee (Salk Institute, EUA), J. Karhunen e E. Oja (HUT, Finlândia) entre outros.

Deste modo, a ICA vem ganhando força no cenário de pesquisas internacionais. Atualmente sua utilização em dados de fMRI é uma das áreas de aplicação.

3.2 Modelo e Definição de ICA

Com o objetivo de se estimar quantidades desconhecidas, sejam elas parâmetros ou seqüências de dados ou sinais, a partir de medidas conhecidas e de alguma restrição estatística sobre os dados originais, deve-se utilizar um modelo de representação que seja capaz de expressar a relação entre as grandezas desconhecidas.

O modelo mais utilizado é o modelo linear expresso, por:

$$x(k) = As(k) \quad (3.1)$$

Onde $x(k)$ é o vetor das medidas realizadas pelos sensores e A é a matriz de mistura e o vetor $s(k)$ é aquele que contém as fontes sinais ou dados originais. A matriz A corresponde aos parâmetros de combinação dos sinais ou dados, sendo que os coeficientes ou pesos desta matriz são considerados constantes e desconhecidos, uma vez que não é possível determiná-los sem o conhecimento das propriedades do sistema de mistura.

Assim, partindo do modelo linear (3.1) e do princípio de que as fontes originais são estatisticamente independentes, a ICA procura estimar um conjunto de fontes $y(k)$ também independentes, através de um sistema de separação W quando somente as medidas dos sensores $x(k)$ são conhecidas

(Hyvarinen, 2001; Kofidis, 2001; Comon, 1994). Para isso, a restrição estatística utilizada é que a função densidade de probabilidade das fontes sejam não-gaussianas, ou que no máximo uma das fontes tenha f.d.p. gaussiana. Esta restrição é utilizada, visto que, conforme descreve o Teorema do Limite Central (Papoulis, 1991), a soma de variáveis aleatórias fornece uma distribuição de probabilidade conjunta gaussiana, o que inviabiliza qualquer inferência sobre as fontes originais a partir das observações dos sensores, ou seja das misturas.

O modelo de separação utilizado na ICA pode ser escrito como:

$$y(k) = \hat{s}(k) = Wx(k) \quad (3.2)$$

Na realidade, as formulações das equações (3.1) e (3.2) não são diferentes, uma vez que W pode ser considerada a inversa de A .

A partir da equação (3.2) o problema passa a ser a determinação da matriz de separação W , tal que os componentes de $y(k)$ sejam independentes ou tão independentes quanto possível.

Mais uma vez vale a pena frisar que este problema só será resolvido se, e somente se, as fontes originais forem consideradas não-gaussianas, pois só assim a restrição de independência estatística será suficiente para a determinação dos coeficientes de W e dos componentes $y(k)$. Desta forma, pode-se dizer que a não-gaussianidade das fontes originais é o ponto de partida para a elaboração de métodos para a solução do problema da ICA, que tem por objetivo encontrar componentes independentes para o caso especial em que as fontes são não-gaussianas.

Para melhor compreensão de como são feitas medidas de não-gaussianidade e como estas medidas são utilizadas para a separação de fontes estatisticamente independentes serão apresentados, ao longo deste capítulo, os princípios matemáticos da ICA e os principais algoritmos desenvolvidos com base nestes princípios.

3.3 Restrições do Método ICA

Nas seções seguintes serão apresentados as principais considerações estatísticas para solucionar o problema da Análise de Componentes Independentes.

3.3.1 Independência Estatística

Para que se possa utilizar ICA é necessário que os sinais originais sejam estatisticamente independentes. Assim sendo, a matriz de separação W é determinada de tal forma que os componentes de $y(k)$ também sejam estatisticamente independentes. Isto significa que o valor de um componente

não fornece nenhuma informação sobre o valor dos outros componentes. Em termos estatísticos, as variáveis são mutuamente independentes se a f.d.p. conjunta de $y(k)$ puder ser fatorada no produto de suas f.d.p.'s marginais, como mostra a equação (3.3) [33].

$$p(y_1, \dots, y_n) = \prod_{i=1}^n p_{y_i}(y_i) \quad (3.3)$$

A equação (3.3) pode ser utilizada para definir a propriedade mais importante das variáveis aleatórias independentes, ou seja, [27]:

$$E[g(y_1), \dots, g(y_n)] = E[g(y_1)]E[g(y_2)] \cdots E[g(y_n)] \quad (3.4)$$

Onde $g(\cdot)$ são funções não lineares quaisquer.

Através da equação (3.4) pode-se verificar que se uma determinada v.a. independente sofrer uma transformação não linear a condição de independência não é afetada.

Uma forma fraca de independência é a não correlação. Um conjunto de n v.a. (y_1, \dots, y_n) é dito ser não correlacionado se sua covariância é igual a zero; como definido pela equação (3.5).

$$E[y_1, \dots, y_n] - E[y_1]E[y_2] \cdots E[y_n] = 0 \text{ para } y_1 \neq y_2 \neq \dots \neq y_n \quad (3.5)$$

Como base em (3.3) e (3.5) pode-se comprovar que se as v.a. são estatisticamente independentes, elas são não correlacionadas, mas a não correlação não implica em independência. O único caso em que não correlação implica em independência estatística ocorre quando as v.a. são gaussianas, uma vez que estas variáveis são completamente descritas por suas estatísticas de segunda ordem (Papoulis, 1991). Como a não correlação não garante a independência, ela também não pode garantir a separação em componentes independentes. Assim, para garantir a separação é necessário recorrer a estatísticas de ordem superior. A utilização de estatísticas de ordem superior, por sua vez, faz com que a separação, utilizando ICA, somente seja garantida para no máximo uma fonte gaussiana.

Não obstante, como a independência estatística implica na não correlação, a maioria dos métodos de ICA restringe o problema da estimativa de fontes originais de tal forma que os componentes independentes estimados sejam sempre não correlacionados. O procedimento de não correlação reduz o número de parâmetros livres a serem determinados na separação, simplificando o problema.

3.3.2 Não Gaussianidade das Variáveis

Uma das principais restrições para a realização da ICA é que os componentes independentes sejam não gaussianos ou que no máximo um dos componentes seja gaussiano. Nesta seção são apresentadas as principais razões que justificam esta restrição.

Considere, por exemplo, um sistema de mistura (2×2) ortogonal e duas fontes gaussianas ($s_1e s_2$). Assim, os sinais misturados resultantes (x_1 e x_2) são gaussianos, não correlacionados e tem variância unitária. Como a f.d.p. conjunta de (x_1 e x_2) é simétrica, ela não contém nenhuma informação sobre a matriz de mistura A . Desta forma, pode-se dizer que a matriz A não pode ser estimada, se mais do que uma das fontes originais for gaussiana [27]. De forma mais rigorosa, pode-se comprovar que a distribuição das variáveis gaussianas não é afetada por qualquer transformação ortogonal e, que as variáveis são independentes. Assim, para variáveis gaussianas, pode-se apenas estimar o modelo para ICA a menos de uma transformação ortogonal. Isto confirma e reforça a afirmação de que a separação em componentes independentes só é possível se no máximo uma fonte original for gaussiana. Além disso, é importante reforçar a idéia de que a separação somente é possível através da utilização de estatísticas de ordem superior.

Por outro lado, para variáveis não gaussianas, a independência estatística é a principal garantia de que as fontes estimadas serão distintas e com base neste critério são definidas a maioria das estratégias para a separação cega utilizando ICA.

A seguir são apresentadas mais informações sobre as f.d.p.s gaussianas e não gaussianas para melhor compreensão das definições.

A função densidade de probabilidade para a distribuição gaussiana ou normal é definida em (3.6).

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (3.6)$$

Onde μ e σ são, respectivamente, a média e o desvio padrão.

Uma densidade de probabilidade sub-gaussiana ou Platykurt é tipicamente mais plana que a distribuição gaussiana. Um exemplo é a f.d.p. uniforme.

Já a f.d.p. super-gaussiana ou Leptokurtic tem tipicamente extremidades mais prolongadas e picos mais proeminentes que os da distribuição gaussiana. Um exemplo é distribuição Laplaciana. Sinais de voz e de música apresentam distribuição super-gaussiana.

A Figura (3.1), mostra exemplos de variáveis sub, super e gaussianas.

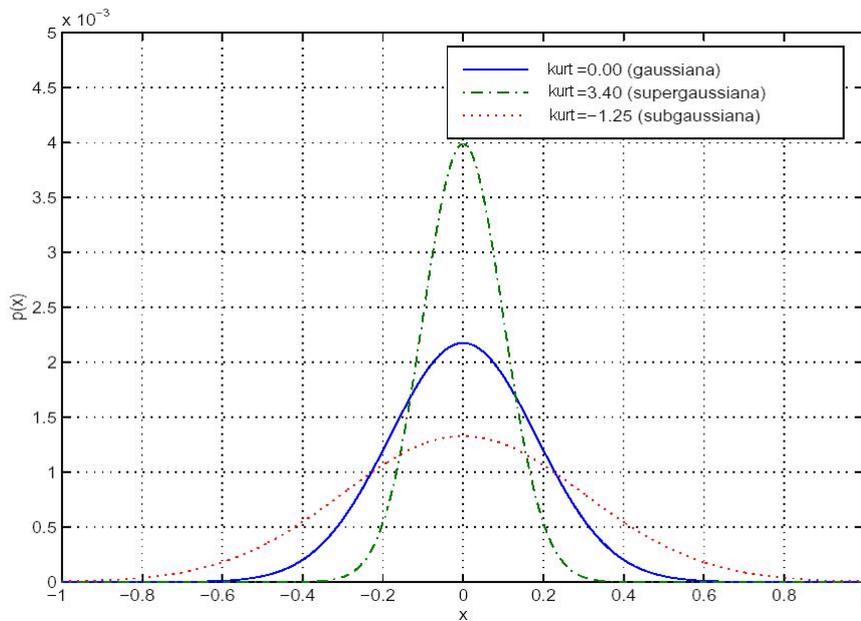


Figura 3.1: Exemplo de distribuições gaussianas, super-gaussianas e sub-gaussianas

Assim, vários métodos foram desenvolvidos, baseados no princípio de que para variáveis não gaussianas a independência é suficiente para garantir que as fontes estimadas serão distintas, validando desta forma a separação das fontes.

Na seção seguinte, serão apresentados os principais métodos utilizados para medidas de independência que permitem a separação dos sinais recebidos pelos sensores em componentes independentes.

3.4 Técnicas de Pré-Processamento

A maioria dos métodos para ICA necessita que os dados passem por pré-processamento. Os principais métodos de pré-processamento empregados pelas técnicas de análise de componentes independentes são a Centralização (Centering) e o Branqueamento (Whitening). Outros métodos como, por exemplo, a filtragem tipo passa faixa [27], também podem ser empregados, mas não serão discutidos neste trabalho, apenas apresentaremos a centralização e o processo de branqueamento dos dados.

3.4.1 Centralização

Este tipo de pré-processamento não é obrigatório, no entanto simplifica os dados de entrada para o algoritmos. O processamento inicial é feito através da centralização do vetor $x(k)$. esse processo é feito subtraindo a média $E[x(k)]$ de cada elemento de $x(k)$, tornando $x(k)$, desta forma, uma variável

de média zero. Logo a matriz das componentes $s(k)$ também terá média zero.

Após a estimativa da matriz de mistura W , com os dados centralizados, pode-se completar a estimativa adicionando vetor média de $s(k)$ de volta ao vetor $s(k)$ centralizado.

3.4.2 Branqueamento

As informações estatísticas de segunda ordem não permitem a separação em componentes independentes. Mas, por outro lado, a não correlação espacial ou independência de segunda ordem permite expressar o espaço dos sinais medidos em um sistema ortogonal, ou branco, que ainda não implica na independência estatística destas medidas, mas que irá simplificar o problema [8].

No branqueamento, também denominado esferamento, o vetor de observações $x(k)$ é linearmente transformado em um novo vetor $z(k)$ que é branco e é definido como: $z(k) = Qx(k)$, onde Q é uma matriz de transformação linear.

O branqueamento faz com que os componentes de $z(k)$ sejam não correlacionados e que sua matriz de covariância seja unitária, ou seja, igual à identidade.

Portanto, se Q é uma matriz de branqueamento para $x(k)$, então $z(k) = Qx(k)$, é espacialmente branco, seus componentes são não correlacionados e sua variância é unitária.

O branqueamento é aplicado, portanto, após a centralização e antes do algoritmo principal e sua aplicação é sempre possível.

Um dos métodos mais simples para branqueamento e pode ser empregado com algoritmos de ICA - utiliza a decomposição em autovetores (EVD) da matriz de covariância dos dados medidos, ou seja,

$$E[x(k)x(k)^T] = MDM^T \quad (3.7)$$

Onde M é a matriz ortogonal de autovetores de $E[x(k)x(k)^T]$ e D é a matriz diagonal de seus autovalores.

Os dados branqueados $z(k)$ são então obtidos, tais que:

$$z(k) = MD^{-\frac{1}{2}}Mx(k) \quad (3.8)$$

Onde a matriz $D^{-\frac{1}{2}}$, é calculada como $D^{-\frac{1}{2}} = \text{diag}(d_1^{-\frac{1}{2}}, \dots, d_n^{-\frac{1}{2}})$. É fácil, então, verificar que $E[z(k)z(k)^T] = I$, onde I é a matriz identidade [27].

O branqueamento também transforma a matriz de mistura em uma nova matriz A_b , tal que:

$$\begin{aligned}
z(k) &= MD^{-\frac{1}{2}}Mx(k) \\
z(k) &= MD^{-\frac{1}{2}}MA_s(k) \\
z(k) &= A_b s(k)
\end{aligned} \tag{3.9}$$

Sendo que, a matriz branqueada A_b é uma matriz ortogonal. Por isso, muitas vezes o branqueamento é descrito como uma transformação ou decomposição ortogonal [8].

O fato de A_b ser ortogonal reduz o número de variáveis desconhecidas na matriz de mistura A de n^2 para $\frac{n(n-1)}{2}$. Muitos algoritmos se beneficiam desta afirmação e estimam a matriz de mistura A_b ao invés da matriz de mistura A . Em sistemas de grandes dimensões, uma matriz ortogonal contém aproximadamente metade do número de parâmetros de uma matriz arbitrária. Para grandes dimensões, pode-se dizer que o branqueamento reduz metade do problema da ICA.

Uma ilustração gráfica do efeito do branqueamento pode ser visto na Figura (3.2). Na Figura (3.2a) tem-se a função de distribuição de probabilidade conjunta dos componentes independentes $s_1(k)$ e $s_2(k)$, em (3.2b) pode-se ver a distribuição conjunta dos sinais medidos $x_1(k)$ e $x_2(k)$ e finalmente em (3.2c) tem-se a distribuição conjunta dos sinais $z_1(k)$ e $z_2(k)$, após o branqueamento.

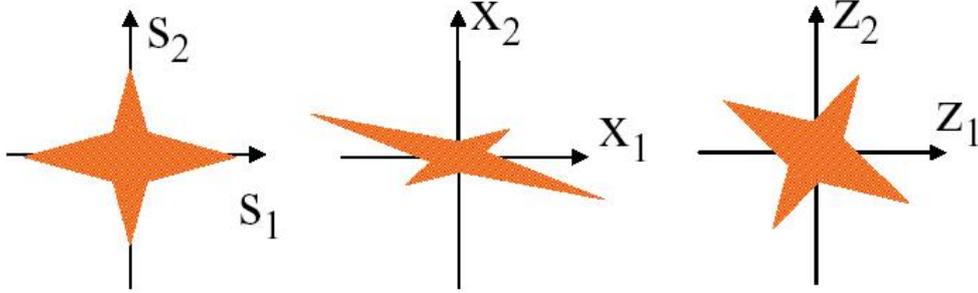


Figura 3.2: Efeito do branqueamento: representação da f.d.p. para $n = 2$, sendo: (a) fontes originais, (b) medidas dos sensores, (c) dados branqueados (Fuentes, 2001).

3.5 Princípios para Estimativa de Componentes Independentes

A estimativa dos dados independentes é realizada, na maioria dos casos, pela minimização ou maximização de uma dada função custo, de contraste ou objeto [10]. Neste trabalho, será utilizado o

termo função custo para definir qualquer função cuja otimização permite a estimativa de componentes independentes. Além disso, não será considerada a presença de ruído aditivo.

Em [27], como neste trabalho, é feita uma distinção entre a função custo e o algoritmo para implementar a ICA, onde a relação entre eles pode ser dada pela figura (3.3):

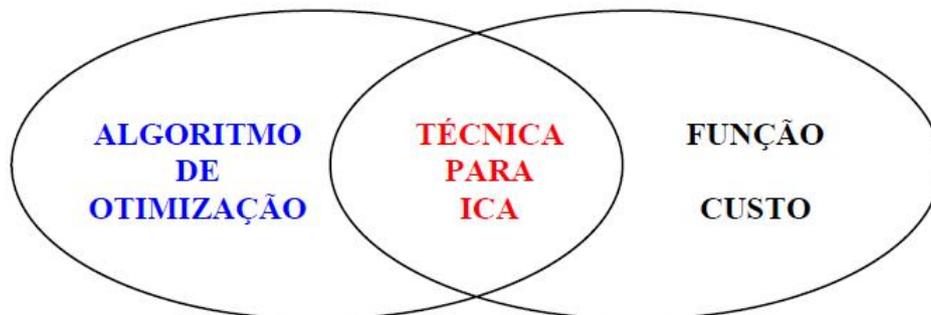


Figura 3.3: Relação entre algoritmo, função custo e modelo para ICA

Ou seja, o modelo para ICA depende de ambos, algoritmo de otimização e função custo, em outras palavras, as propriedades estatísticas- como consistência e robustez- do modelo de ICA dependem da escolha da função custo e as propriedades computacionais- como velocidade de convergência, requisitos de memória e estabilidade- dependem da escolha do algoritmo de otimização sendo que algoritmo e função custo são independentes.

As principais funções custo para estimativa dos componentes independentes são baseadas na medida e maximização da não-gaussianidade das variáveis medidas, minimização da não-gaussianidade da informação mútua e estimativa da máxima verossimilhança.

3.5.1 Maximização da Não Gaussianidade

Uma das formas de estimar os componentes independentes é através da medida da não gaussianidade dos componentes, com as seguintes considerações:

- Sejam centralizados;
- Tenham variância unitária.

A medida de não gaussianidade pode ser feita por meio do Kurtosis ou achatamento da curva da distribuição de probabilidade dos componentes sob estudo, ou usando o princípio da Negentropia.

Medida de Não Gaussianidade por Kurtosis

O Kurtosis ($Kurt(.)$) de uma variável aleatória $y(k)$ é definido como:

$$Kurt(y(k)) = E[y(k)^4] - 3(E[y(k)^2])^2 \quad (3.10)$$

como $y(k)$ tem variância unitária, podemos reescrever a equação (3.10):

$$Kurt(y(k)) = E[y(k)^4] - 3 \quad (3.11)$$

Logo, por (3.11) o kurtosis é simplesmente uma normalização do momento de quarta ordem, ou seja, $E[y(k)^4]$.

A justificativa de utilizar o kurtosis vem do fato que existe uma propriedade muito importante em relação ao kurtosis que para uma variável gaussiana, o momento de quarta ordem é dado por $3E[[S^2(k)]^2]$, como a variância $E[S^2(k)] = 1$, conseqüentemente, o kurtosis é nulo, o que reforça a idéia de que variáveis gaussianas são proibidas para ICA.

O kurtosis de uma variável aleatória pode ser positivo, negativo ou nulo, sendo que:

- $k > 0$ a variável é sub-gaussiana é um exemplo de função densidade de probabilidade é a plana, uniforme;
- $k < 0$ a variável é super-gaussiana é um exemplo de função densidade de probabilidade é a Laplaciana.
- $k = 0$ a variável é gaaussiana é um exemplo de função densidade de probabilidade gaussiana ou normal.

A não gaussianidade pode ser medida através do valor absoluto do kurtosis, sendo que:

- $|Kurt| = 0$ variáveis gaussianas;
- $|kurt| > 0$ variáveis não- gaussianas.

Os componentes independentes são encontrados maximizando ou minimizando o kurtosis, de acordo com o algoritmo de otimização empregado.

Na prática, para se encontrar a matriz de separação W que fornecerá os componentes independentes $y(k)$ com base na medida dos kurtosis das fontes, deve-se:

- Branquear o sinal medido $x(k)$, de forma a se obter $z(k)$;
- Inicializar W com um valor qualquer;
- Calcular a direção em que kurtosis de $y(k)$ está crescendo ou decrescendo mais fortemente;

- Baseado nas medidas disponíveis $x(k)$ para encontrar um novo vetor W ;
- Repete o processo descrito acima até que todas as fontes sejam estimadas.

No entanto os algoritmos que utilizam kurtosis para solucionar o problema da ICA apresentam algumas desvantagens:

- Implementação computacional complexa;
- Não é uma medida robusta de não gaussianidade;
- É sensível a outliers;
- Como se deve estimar os kurtosis a partir de dados observados, os valores podem variar bastante em função do número de amostras e alterar a estimativa das fontes.

Negentropia

A entropia é um conceito fundamental da Teoria de Informação e mede o grau de informação que pode ser obtida através da observação de uma variável. Neste sentido, quanto mais imprevisível for a variável observada, maior será sua entropia. Se uma variável possui uma tendência, ou seja, se há uma grande probabilidade de um evento específico ocorrer, o ganho informacional não é tão grande em cada observação, pois já se espera que este evento seja altamente provável.

As variáveis gaussianas possuem a maior entropia entre variáveis de mesma variância, [27]. Dessa forma, podemos usar a entropia como uma medida de não-gaussianidade.

Portanto a entropia H para uma dada variável aleatória define o grau de informação dessa variável aleatória:

$$H(y) = - \sum P(y = a_i) \log(P(y = a_i)) \quad (3.12)$$

onde $P(y = a_i)$ é a probabilidade de $y(k) = y$ ser igual a a_i .

A negentropia (J) é dada:

$$J(y) = H(y_{gauss}) - H(y) \quad (3.13)$$

onde y_{gauss} é uma variável aleatória gaussiana com a mesma matriz de covariância de y .

Desta forma obtém-se uma medida que é a sempre positiva e mede a diferença de Entropia existente entre a variável gaussiana com os mesmos parâmetros. Esta medida recebe o nome de negentropia, e nada mais é do que uma versão normalizada da Entropia e que pode ser utilizada pelo algoritmo do modelo ICA para encontrar os Componentes Independentes baseando-se na não-gaussianidade da distribuição.

O cálculo da entropia, por conseguinte da negentropia é uma medida que exige muito tempo computacional. Devido a isso, geralmente são utilizadas aproximações para se estimar a negentropia.

O cálculo da negentropia aproximada é através de momentos de ordem superior, dado por:

$$J(y) = \frac{1}{12}E[y^3]^2 + \frac{1}{48}Kurtosis(y)^2 \quad (3.14)$$

Como y tem média zero e variância unitária e a validade desta aproximação é limitada, devido utilizar a *kurtosis*, acaba sofrendo a não robustez dela, chegando algumas vezes a ser imprecisa e em muitos casos sensíveis a “outliers”.

Por esse motivo, novas técnicas foram propostas. No caso mais simples, utiliza-se uma função não-quadrática, $G(\cdot)$, resultando na seguinte aproximação:

$$J_G(y) = \alpha[E[G(y)] - E[G(b)]]^2 \quad (3.15)$$

é válida para praticamente qualquer função $G(\cdot)$ não quadrática e considerando que b é uma variável aleatória gaussiana com média zero e variância unitária.

A escolha adequada de $G(\cdot)$, faz com que (3.15) seja uma aproximação conceitualmente simples, computacionalmente rápida e com propriedade estatísticas interessantes, como a robustez.

As funções normalizadas usadas são:

$$G_1(u) = u^3, \quad G_2(u) = \tanh(a_1 \cdot u), \quad G_3(u) = u \cdot \exp(-a_2 \cdot u^2/2), \quad G_4(u) = u^2 \quad (3.16)$$

Um dos principais algoritmos que utilizam a aproximação da negentropia para medida de não gaussianidade é o FastICA[27].

Existem diversos algoritmos que são utilizados para a implementação do princípio estatístico, nesse momento, já determinado. Algoritmo de [28] trabalha com matrizes de dados quadradas e ainda exigem que as entradas tenham média zero, no entanto, quando sua matriz de mistura for mal condicionada ele trabalha de forma imprecisa. Além disso, o número de fontes que o método pode separar é limitado. O algoritmo rápido de ponto fixo (FastICA) é um dos mais utilizados e que apresentam uma melhor estimativa das componentes independentes.

3.5.2 Decomposição em Valores Singulares

Antes de aplicarmos os algoritmos para estimar as componentes independentes, vamos aplicar a decomposição em valores singulares como uma etapa de redução da dimensão dos dados, objetivando uma redução significativa no custo computacional.

A grande dimensão dos dados de fMRI, normalmente no domínio do espaço, além disso os algoritmos de ICA são, na maioria das vezes, de computação intensiva. Por esse motivo antes de aplicar ICA vamos realizar uma redução de dimensão dos dados usando o SVD. Para isso as matrizes ortogonais desempenham um papel importante nesta seção.

Definição 3.1 Dizemos que uma matriz U é ortogonal se

$$U^T U = I, \tag{3.17}$$

onde I é a matriz identidade.

Observe que, se uma matriz é ortogonal, sua inversa é calculada pela simples operação de transposição.

O embasamento teórico do método SVD está no resultado que enunciaremos a seguir, cuja demonstração pode ser facilmente encontrada, por exemplo, em [18].

Teorema 3.1 Seja $X_{m \times n}$, com $m \geq n$. Então existem matrizes ortogonais $U_{m \times m}$ e $V_{n \times n}$ tais que

$$X = U D V^T \tag{3.18}$$

onde

$$D = \begin{bmatrix} \sigma_1 & & & & & \\ & \sigma_2 & & & & \\ & & \sigma_3 & & & \\ & & & \ddots & & \\ & & & & \sigma_n & \\ & 0 & \dots & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & 0 & \dots & \dots & \dots & 0 \end{bmatrix} \tag{3.19}$$

com $p = n$ e $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \dots \geq \sigma_p > \sigma_{p+1} = \sigma_n = 0$. Os números σ_i são chamados valores singulares de X .

Onde as matrizes U e V são matrizes ortogonais e D é uma matriz diagonal que contém os autovalores de X .

O produto de XX^T fornece os autovalores da matriz X e as colunas de V , enquanto o produto $X^T X$ fornece os mesmos autovalores e as colunas de U . Os valores singulares são a raiz quadrada dos autovalores obtidos nestes produtos, em ordem decrescente.

As colunas das matrizes U e V , respectivamente são os chamados autovetores à esquerda e à direita de X . Aqui U e V podem ser vistos como os vetores de base que abrangem os padrões espaciais e seqüências temporais, respectivamente.

A utilização desse método nesse trabalho se deu pela ampla utilização que o SVD tem no processo de compressão de imagens, principalmente na recuperação das informações da mesma.

Veamos um exemplo, bem conhecido da utilização da svd na compressão de imagens:

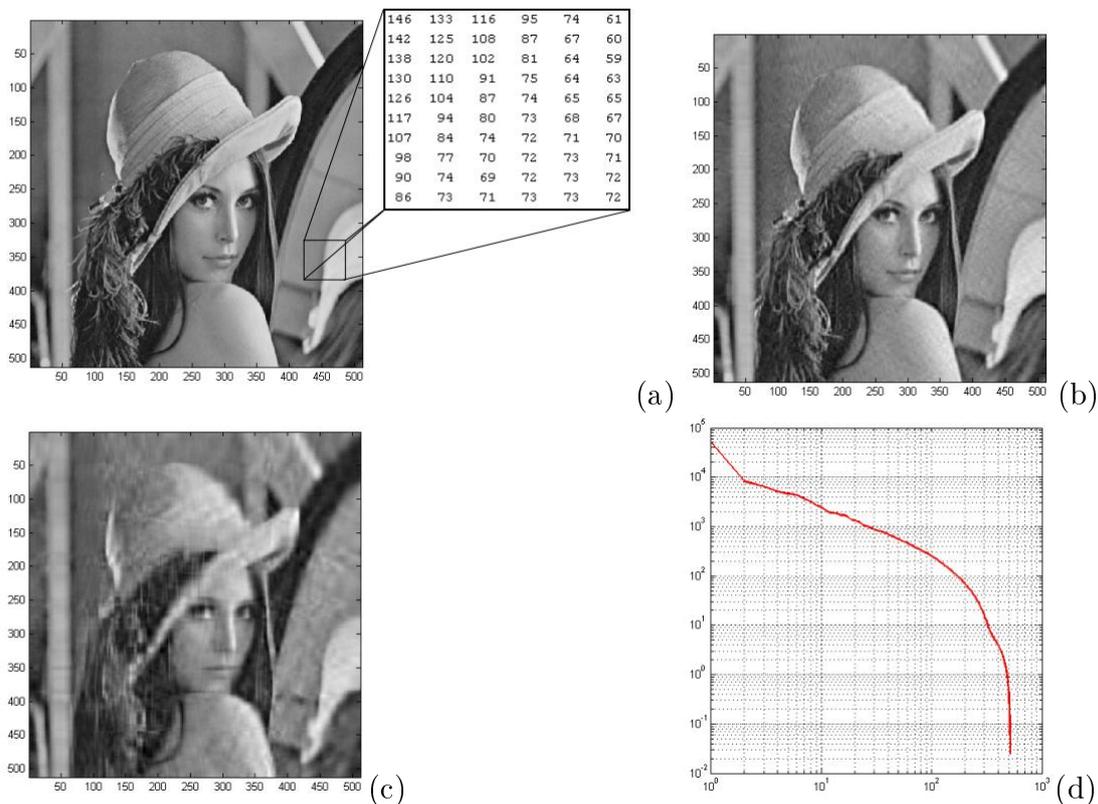


Figura 3.4: As figuras (a), (b) e (c) mostram as imagens resultantes se consideramos o total das componentes espectrais de 512, 50 e 20, respectivamente e a imagem (d) com a dispersão dos autovalores

Consideremos a imagem com 512×512 pixels que se reproduz na figura a). Cada um dos pixels da imagem corresponde a um número inteiro entre 0 e 255, relativo a 1 de 256 níveis de cinza, constituindo toda a imagem uma matriz de $512 \text{ times } 512 = 262144$ números. Decompondo a matriz em valores singulares,

$$X = UDV^T, \tag{3.20}$$

podemos verificar que os seus 512 autovalores, σ_i , se distribuem por 10^6 ordens de grandeza, sendo o maior dos valores $\sigma_1 \approx 5 \times 10^4$, e o menor $\sigma_{512} \approx 2 \times 10^{-2}$, como podemos ver na figura d) (note a escala logarítmica).

Se fizermos a decomposição espectral da matriz,

$$X = \sum_{i=1}^{512} \sigma_i u_i v_i^T, \quad (3.21)$$

e sendo os coeficientes das matrizes $u_i v_i^T$ todos, em módulo, menores que 1, dado que os autovalores à esquerda e à direita, u_i e v_i , são vetores normalizados, resulta que cada uma das parcelas, $\sigma_i u_i v_i^T$, contribui de modo muito diferente para a matriz final.

As figuras de a) a c) mostram as imagens resultantes de se considerar o total das componentes espectrais, e apenas as 50 e 20 mais importantes, isto é, associadas aos 50 e 20 valores singulares de maior valor. Note que, sendo necessário $512 \times 512 = 262144$ números para reproduzir a imagem original, a imagem que utiliza apenas 50 componentes espectrais necessita de apenas 50 autovalores mais 50 autovetores u e 50 autovetores v , cada um de dimensão 512, ou seja $50 \times (1 + 512 + 512) = 51250$ números, cerca de 5 vezes menos informação.

A decomposição de imagens em valores singulares está na base de alguns dos mais eficientes algoritmos de compressão de imagem. O armazenamento e transmissão de dados têm um papel extremamente importante no atual panorama tecnológico, e o estudo das técnicas que permitam a sua execução com o mínimo espaço de armazenamento e o menor tempo de processamento.

A idéia da etapa de redução dos dados antes do ICA é reduzir a dimensão das matrizes a serem utilizados como insumos para algoritmos de ICA. considerando-se o fato de que a maioria dos experimentos fMRI tem muito menos pontos no tempo do que voxels, esta etapa de pré-processamento aumenta muito a eficiência computacional.

3.5.3 Algoritmo Rápido de Ponto Fixo (FastICA)

Os algoritmos baseados em iterações de ponto fixo denominado FastICA [27] são diferenciados pela abordagem e pela função custo utilizada, eles visam encontrar componentes independentes através da maximização da negentropia.

Foram desenvolvidas duas versões do FastICA; a primeira permite a recuperação de todas as fontes (multi-unit) e a segunda, encontra um dos componentes (one-unit). Ambas podem utilizar qualquer função custo não paramétrica para estimar a negentropia das fontes.

O algoritmo FastICA procura encontrar uma direção, isto é, um vetor cuja projeção $W^T x(k)$ maximiza a função $J_G(W)$, ou seja, uma aproximação da negentropia dada por:

$$J_G(W) = [E[G(W^T X(k))] - E[G(b)]]^2 \quad (3.22)$$

Onde $G(\cdot)$ é qualquer função não quadrática e b é uma variável gaussiana de média zero e variância unitária. Na prática, as esperanças $E[\cdot]$ são substituídas por médias amostrais.

A regra de treinamento do algoritmo consiste de um método que estima os componentes um a um (one-unit), sendo que para dados branqueados, o algoritmo FastICA one-unit pode ser equacionado da seguinte forma:

1. Inicializa-se aleatoriamente uma matriz W ;
2. $W_{n+1} = E[X \cdot G(W^T \cdot X)] - E[G'(W^T \cdot X)] \cdot W$;
3. $W = \frac{W_{n+1}}{\|W_{n+1}\|}$;
4. Repetir o passo 2 até que haja convergência.

As esperanças $E[\cdot]$ são estimadas através da média amostral de um determinado número de amostras dos dados de entrada para garantir a precisão.

As mais relevantes características dos algoritmos FastICA são listradas abaixo:

- O FastICA não necessitar de ajuste no passo de adaptação comparado com algoritmos baseados em gradiente, se torna mais simples.
- Pelo fato de ter uma estrutura paralela e distribuída, mas não adaptativa, pode ser considerado um algoritmo neural;
- A baixa quantidade de interações, de 5 a 10, que são necessárias na maioria dos casos faz com que o método tenha uma máxima precisão com um pequeno numero de interações.
- O FastICA utiliza as médias amostrais calculadas sobre uma amostra grande de dados, ao invés de utilizar todos os dados imediatamente para treinamento;
- O algoritmo encontra diretamente as fontes para praticamente qualquer função não gaussiana, ou seja, não é necessário que seja feita uma distinção entre as fontes com kurtosis positivas ou negativas;
- O desempenho do algoritmo pode ser melhorado de acordo com a escolha da função não linear utilizada;
- Pode ser utilizado para identificar os componentes independentes um a um.

Capítulo 4

Resultados

O método de ICA foi aplicado a dados simulados e reais. Para verificar a robustez da ICA feita a análise das curvas ROC utilizando os dados simulados, onde pode ser encontrado uma abordagem profunda em [1]. Foram analisados as relações existentes entre erros entre falsos positivos, fração de pixels inativos marcados como ativos, e falso negativos, fração de pixels ativos classificados como inativos.

4.1 Dados Sintéticos

Os dados artificiais foram construídos simulando um experimento em bloco, com as seguintes características: 60 imagens (64×64), intercalando entre três períodos de repouso e três períodos de atividade começando com um período de repouso com 10 imagens cada.

As áreas consideradas ativas estão em três grupos com 250, 187 e 106 pixels ativos, e 3553 pixels não ativos. Para obter dados mais realistas foram utilizados ruídos com padrão gaussiano, onde a relação sinal-ruído (SNR) é dada por:

$$SNR = 10 \log \left(\frac{\sigma_S^2}{\sigma_R^2} \right) = 20 \log \left(\frac{\sigma_S}{\sigma_R} \right) \quad (4.1)$$

onde σ_S e σ_R são o desvio padrão do sinal e do ruído, respectivamente.

Os dados sintéticos foram gerados contendo 3 regiões de atividades, de forma que apenas o centro dessas regiões são fixadas.

Análise ROC

A robustez dos métodos de análise de dados de fMRI é testada principalmente pela capacidade de detectar o sinal (sensibilidade - SENS) e o ruído (especificidade ou specificity - SPEC) nas imagens

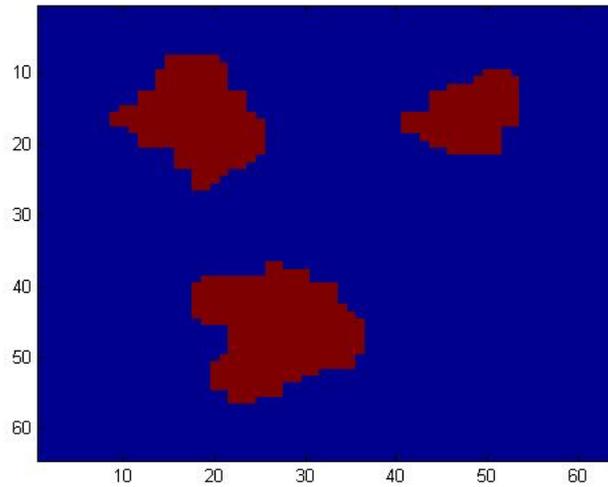


Figura 4.1: Mapa dos dados sintéticos.

funcionais. Os resultados destas análises são apresentados através da construção de curvas características de operação de um receptor conhecidas como Receiver Operating Characteristic (ROC). As curvas ROC descrevem a fração de verdadeiros positivos (FVP) ou fração de pixels ativos classificados corretamente, em função da fração de falsos positivos (FFP) ou fração de pixels inativos classificados incorretamente. De outro modo, a detecção do ruído é obtida por curvas $(1-SENS) \times (SPEC)$ que agora descrevem a fração de verdadeiros negativos (FVN) ou fração de pixels inativos classificados corretamente, em função da fração de falsos negativos (FFN) ou fração de pixels ativos classificados incorretamente.

Em uma experiência simulada de fMRI típica, uma curva ROC é uma demonstração bidimensional da performance do método de análise utilizado para produzir os mapas proporcionais de ativação [20]. Cada mapa de atividade corresponde a uma curva ROC distinta parametrizada por uma sequência de limiares ou faixas de corte. Cada limiar, por sua vez, corresponde a um ponto da curva. Esta sequência de limiares é obtida dentro de um intervalo I que compreende os valores numéricos do respectivo mapa proporcional.

Pode-se obter os limiares usando várias estratégias como por exemplo, tomando um passo constante em I ou condicionando os limiares de modo que os pontos da curva fiquem uniformemente espaçados entre si. Além disso, sabe-se que quanto mais pontos se concentrarem no canto superior esquerdo da curva (ponto $(0; 1)$), melhor poderá ser a avaliação da robustez do método, dependendo das características próprias do problema. Nesse sentido, a estratégia de tomar um passo constante em I é recomendada. Em geral, costuma-se utilizar alguns parâmetros para definir valores escalares

que expressem o desempenho descrito por uma curva ROC, um deles é a área abaixo desta. A área ROC dentro do espaço ROC unitário varia de 0,5 a 1. Dessa forma, um valor de área igual a 0,5 corresponde a uma performance sem importância, e área ROC igual a 1 representa um teste perfeito, [3] e [4].

Na tabela abaixo mostra uma análise comparativa das áreas ROC considerando diferentes funções não-quadráticas, $G(\cdot)$, que inferem na diminuição da quantidade de informações obtidas dos dados analisados, onde o nível de ruído usado foi $\sigma = 1$.

Componentes	G_1	G_2	G_3	G_4	Componentes	G_1	G_2	G_3	G_4
1	0.7290	0.6812	0.7244	0.8212	21	0.8541	0.6713	0.8140	0.5536
2	0.7344	0.7473	0.8416	0.6677	22	0.8011	0.8346	0.7164	0.7778
3	0.6204	0.6295	0.7230	0.7388	23	0.8946	0.7317	0.6202	0.6146
4	0.6434	0.6914	0.6658	0.6336	24	0.7051	0.7263	0.8613	0.5176
5	0.8408	0.8789	0.7405	0.5783	25	0.7097	0.6352	0.8340	0.7098
6	0.6809	0.7511	0.8851	0.5052	26	0.9011	0.8330	0.6286	0.7678
7	0.8921	0.7338	0.6616	0.9723	27	0.7942	0.7598	0.7172	0.6148
8	0.6269	0.7281	0.8745	0.9046	28	0.6593	0.8806	0.7482	0.8629
9	0.7045	0.8411	0.6649	0.8219	29	0.7058	0.7265	0.8910	0.5325
10	0.6499	0.7588	0.8515	0.8026	30	0.6897	0.7191	0.7298	0.5742
11	0.7334	0.8976	0.7612	0.8216	31	0.6907	0.7224	0.7622	0.8015
12	0.6384	0.9018	0.7704	0.8469	32	0.7754	0.6467	0.7541	0.7721
13	0.8433	0.8121	0.7864	0.7811	33	0.6899	0.8767	0.7018	0.5811
14	0.8934	0.8780	0.7949	0.5924	34	0.7175	0.8974	0.8291	0.7465
15	0.8815	0.7339	0.8912	0.7876	35	0.6462	0.7302	0.7418	0.9325
16	0.7709	0.7814	0.6902	0.5725	36	0.7714	0.7285	0.7467	0.6075
17	0.8898	0.7068	0.8559	0.9282	37	0.7741	0.7697	0.6692	0.5013
18	0.7862	0.7596	0.7214	0.9551	38	0.7757	0.7773	0.7012	0.9264
19	0.8443	0.7320	0.7016	0.6887	39	0.8419	0.7395	0.7675	0.5391
20	0.7015	0.6693	0.7409	0.7005	40	0.8452	0.7534	0.7378	0.8601

A escolha adequada de $G(\cdot)$, faz com que a aproximação da negentropia seja uma aproximação mais simples, computacionalmente mais rápida, porém mesmo utilizando a função que de melhor desempenho, G_4 , o custo computacional só foi reduzido consideravelmente ao aplicar o SVD antes do algoritmo principal.

Para obter as componentes independentes realizou-se a correlação das componentes com as series

temporais.

As figuras abaixo mostram os mapas das componentes das quatro funções utilizadas, em destaque temos as componentes mais representativas:

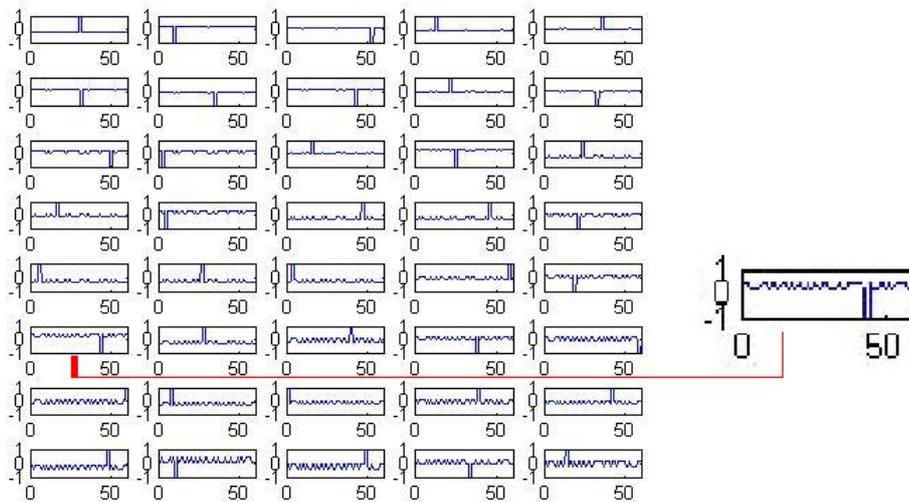


Figura 4.2: Mapa das componentes usando a função G_1 , em destaque a componente 26

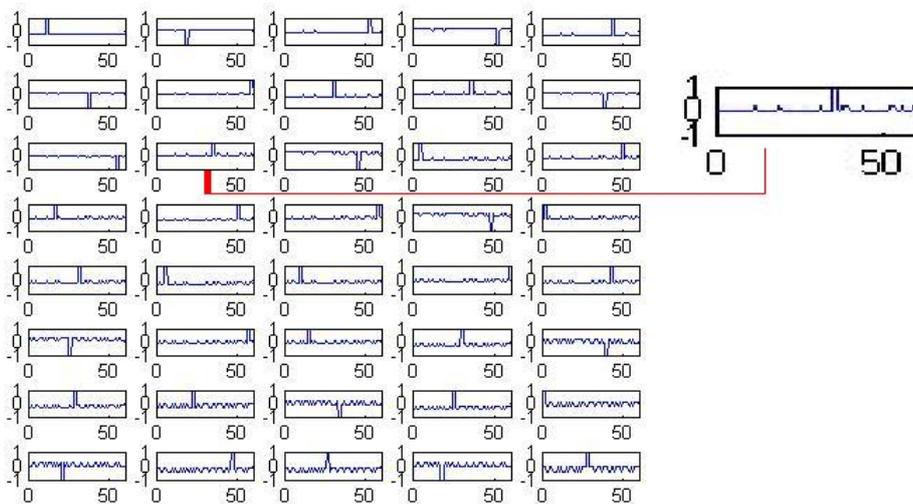


Figura 4.3: Mapa das componentes usando a função G_2 , em destaque a componente 12

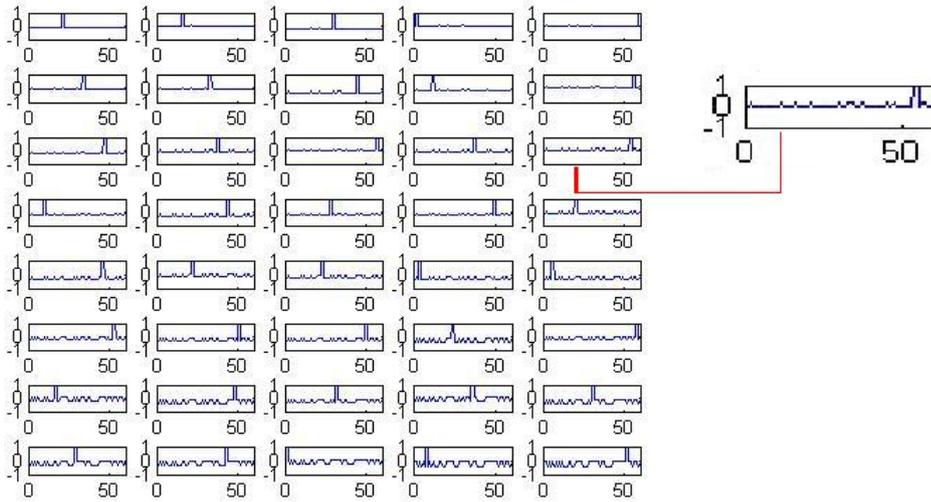


Figura 4.4: Mapa das componentes usando a função G_3 , em destaque a componente 15

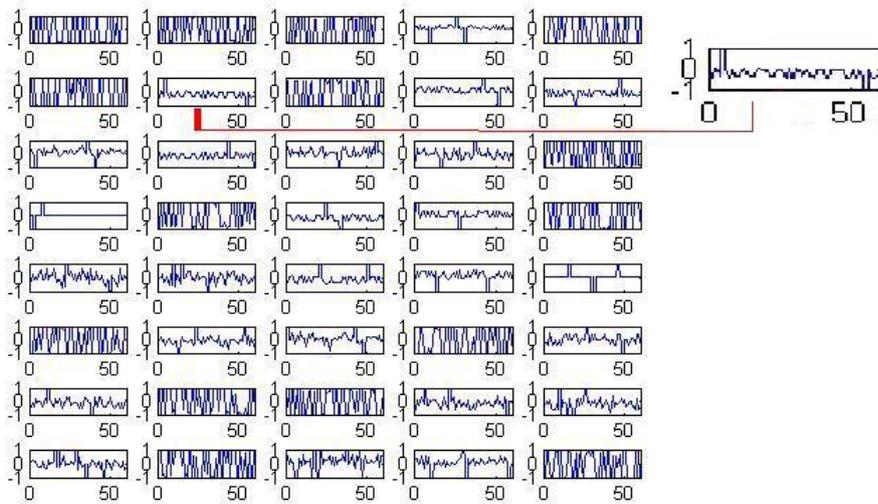


Figura 4.5: Mapa das componentes usando a função G_4 , em destaque a componente 7

Através dos mapas e da tabela com os valores da área de cada componentes, podemos exibir as componentes com maior representatividade e suas respectivas curvas ROC.

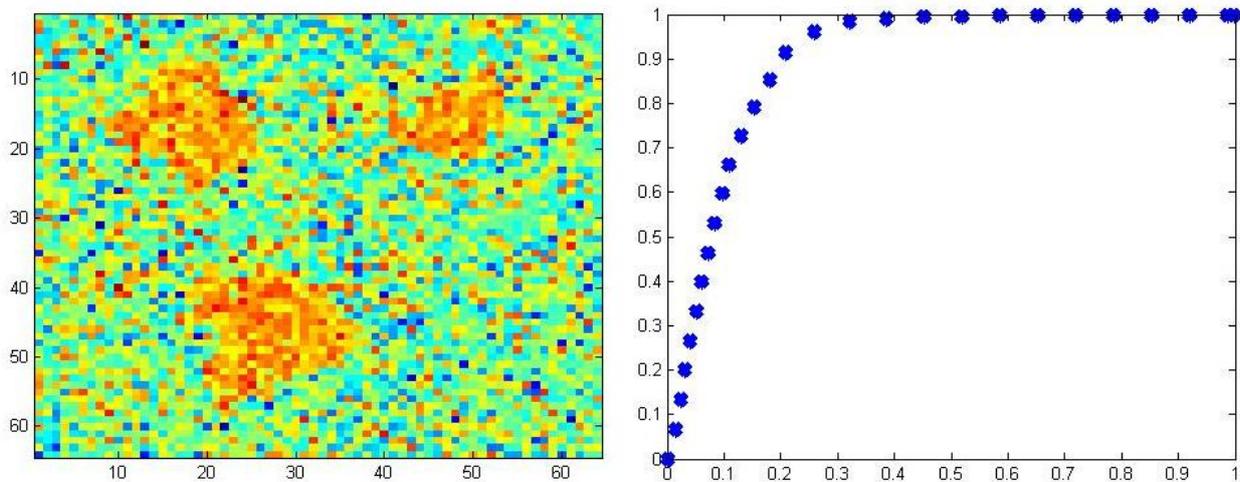


Figura 4.6: Imagem da componente 26 da função G_1 e sua curva ROC

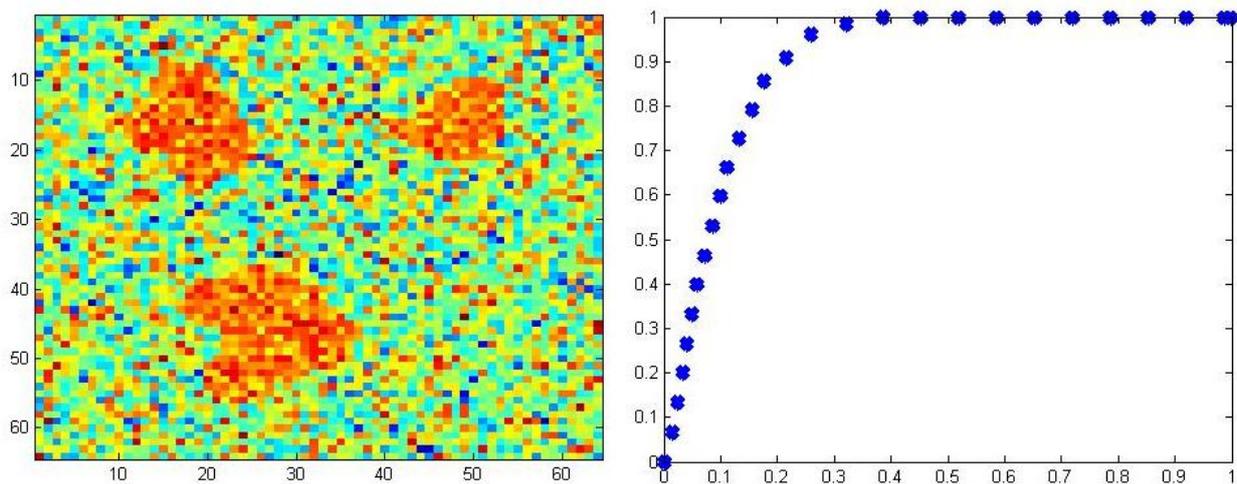


Figura 4.7: Imagem da componente 12 da função G_2 e sua curva ROC

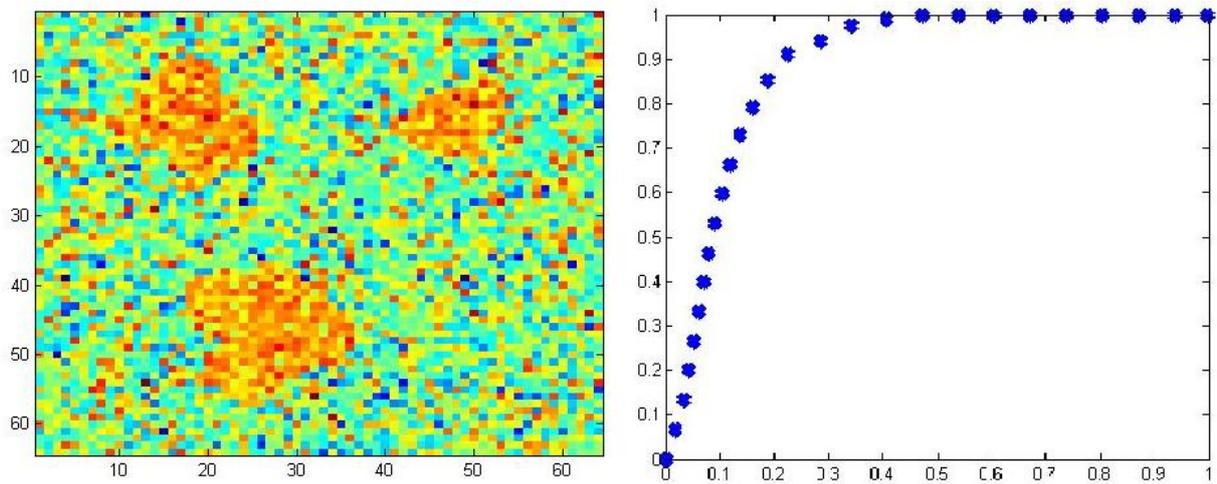


Figura 4.8: Imagem da componente 15 da função G_3 e sua curva ROC

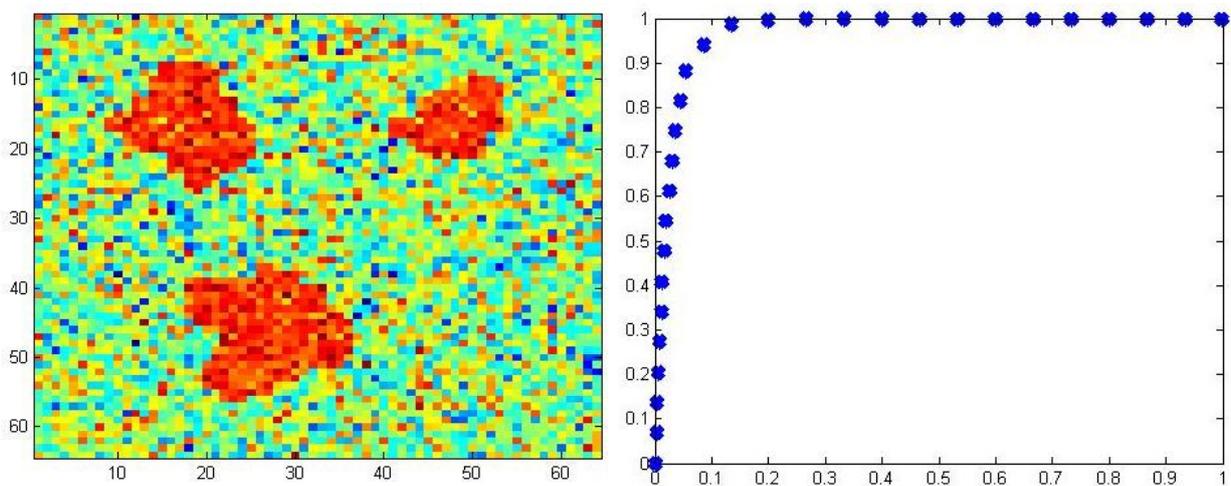


Figura 4.9: Imagem da componente 7 da função G_4 e sua curva ROC

Após as análises de robustez da técnica utilizada é feita a primeira etapa do processo de estimativa das componentes independentes, onde foi aplicado o algoritmo FastICA para obter as imagens referentes a cada componente. Nas figuras (4.10)-(4.13) são mostradas as duas melhores componentes de cada função $G(\cdot)$ utilizada.

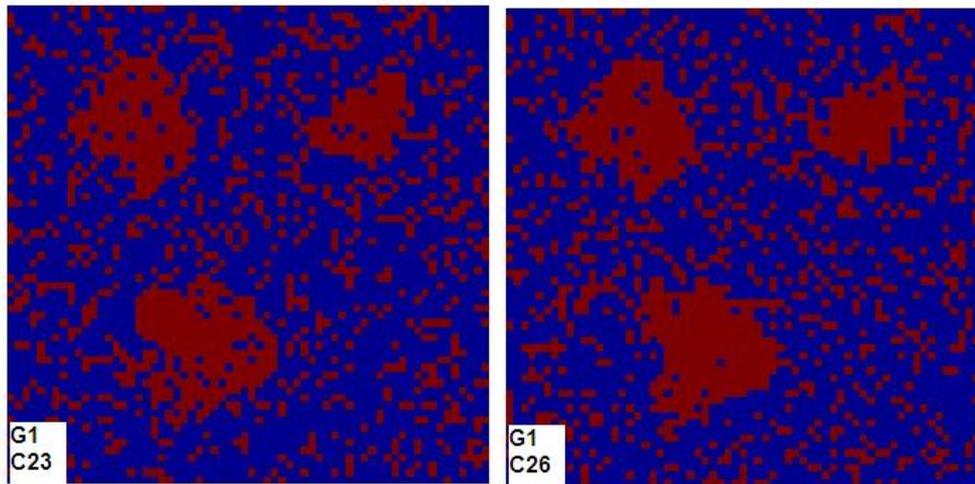


Figura 4.10: As regiões vermelhas na imagem correspondem às regiões ativas marcadas ao usar o limiar no ponto ótimo de operação da curva ROC encontrado para a função G_1

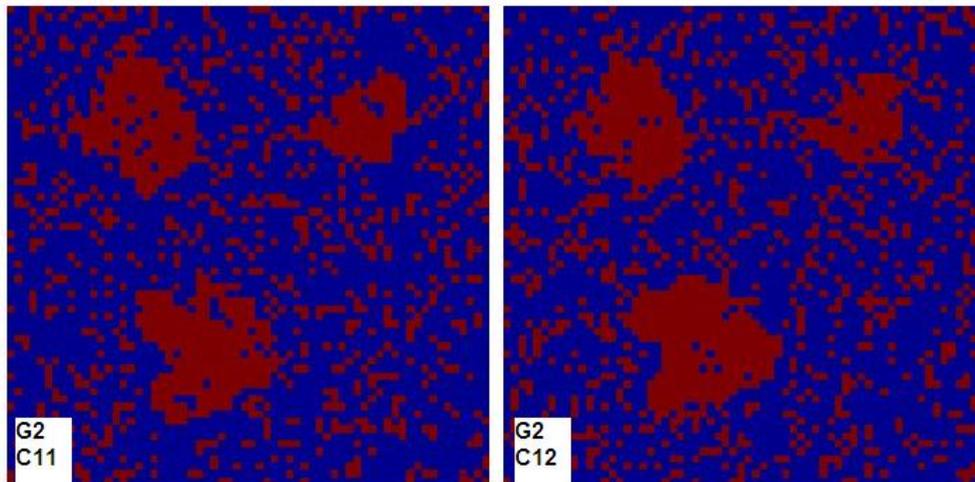


Figura 4.11: As regiões vermelhas na imagem correspondem às regiões ativas marcadas ao usar o limiar no ponto ótimo de operação da curva ROC encontrado para a função G_2

Na tabela abaixo contém: a quantidade de pixels marcados como ativos (PMCA), frações de verdadeiros positivos(FVP), de falsos negativos (FFN), de verdadeiros negativos (FVN) e de falsos positivos(FFP), obtidos nos mapas limitados das figuras(4.10)-(4.13) em seus respectivos ítems. Além disso também é exibido na última coluna as áreas ROC.

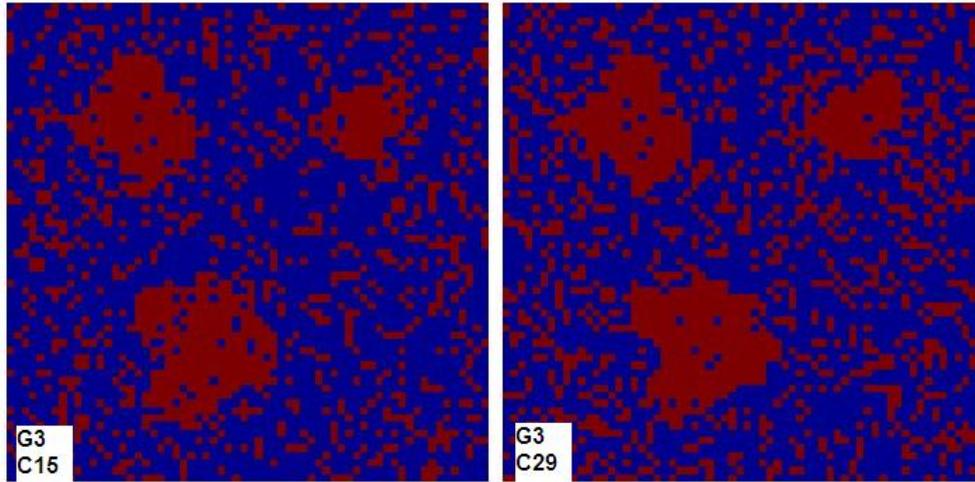


Figura 4.12: As regiões vermelhas na imagem correspondem às regiões ativas marcadas ao usar o limiar no ponto ótimo de operação da curva ROC encontrado para a função G_3

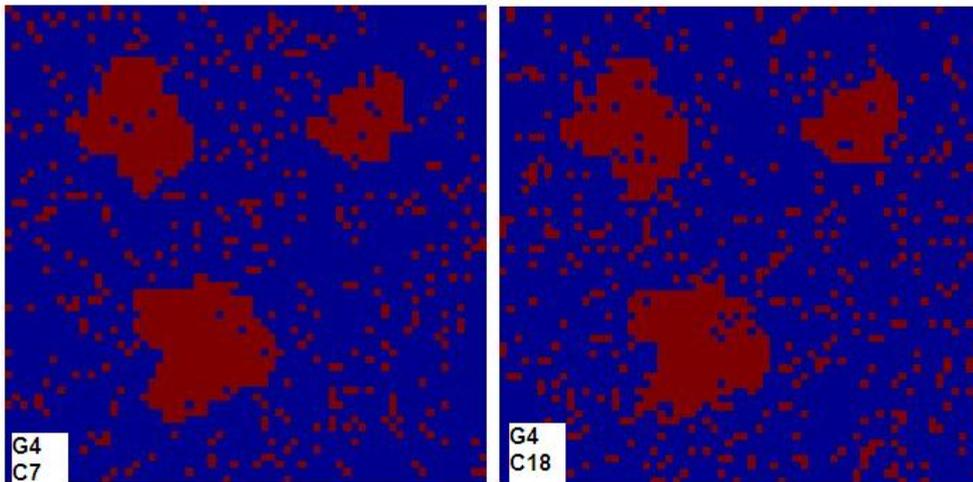


Figura 4.13: As regiões vermelhas na imagem correspondem às regiões ativas marcadas ao usar o limiar no ponto ótimo de operação da curva ROC encontrado para a função G_4

G(.)-Componente	PMCA	FVP	FFN	FVN	FFP	Área ROC
G_1C_{23}	1296	0,9024	0,0976	0,7807	0,2193	0,8946
G_1C_{26}	1290	0,9374	0,0626	0,7802	0,2198	0,9011
G_2C_{11}	1244	0,9153	0,0847	0,7898	0,2102	0,8976
G_2C_{12}	1304	0,9337	0,0663	0,7757	0,2243	0,9018
G_3C_{15}	1299	0,9153	0,0847	0,7743	0,2257	0,8912
G_3C_{29}	1490	0,9540	0,0460	0,7264	0,2736	0,8910
G_4C_7	891	0,9669	0,0331	0,8970	0,1030	0,9723
G_4C_{18}	934	0,9411	0,0589	0,8809	0,1191	0,9551

Os valores das frações de erros e acertos da tabela foram obtidos no ponto ótimo de operação, mais a área ROC. Pode-se observar facilmente que o melhor resultado é obtido pela função G_4 na componente C_7 . De fato a função G_4 , apontada na literatura, é a que mostrou a maior quantidade de componentes significativas. Por essa razão a função utilizada para os dados reais foi a função G_4 .

4.2 Dados Reais

Os dados utilizados neste trabalho corresponde a um estímulo auditivo. O paradigma tem um delineamento em blocos alternando com 6 períodos de 27.5s ON (atividade) e 6 períodos de 27.5 OFF (repouso). Durante a tarefa o indivíduo ouviu passivamente uma estória com uma estrutura narrativa padrão. Imediatamente após o exame, o sujeito teve que informa o conteúdo da narrativa.

Os experimentos foram realizados em um scanner Siemens de 1.5 T, usando-se sequências echo-plana imaging (EPI) com os seguintes parâmetros: TE=60 ms, TR= 4.6s, FA=90°, FOV= 220 mm, espessura das fatias de 6.25 mm, e dimensão do voxel= 172×1.72×6.25 mm. Foram adquiridos 64 volumes cerebrais com 16 fatias cada, sendo o tamanho da matriz 128×128.

Os dados foram pré-processadas para correção de movimento, correção do tempo entre fatia, suavização espacial e também a filtragem temporal usando o software statistical Parametric Mapping (SPM8). As análises dos dados simulados e pós-processamento das imagens, através dos métodos descritos anteriormente, foram utilizando algoritmos implementados em MATLAB 7.2.

Utilizando o ICA para analisar os dados reais correspondentes a um estímulo auditivo, já descrito na seção (4.2), a figura (4.2) mostra o mapa de correlação utilizando a função G_4 e em destaque sua componente de maior representatividade.

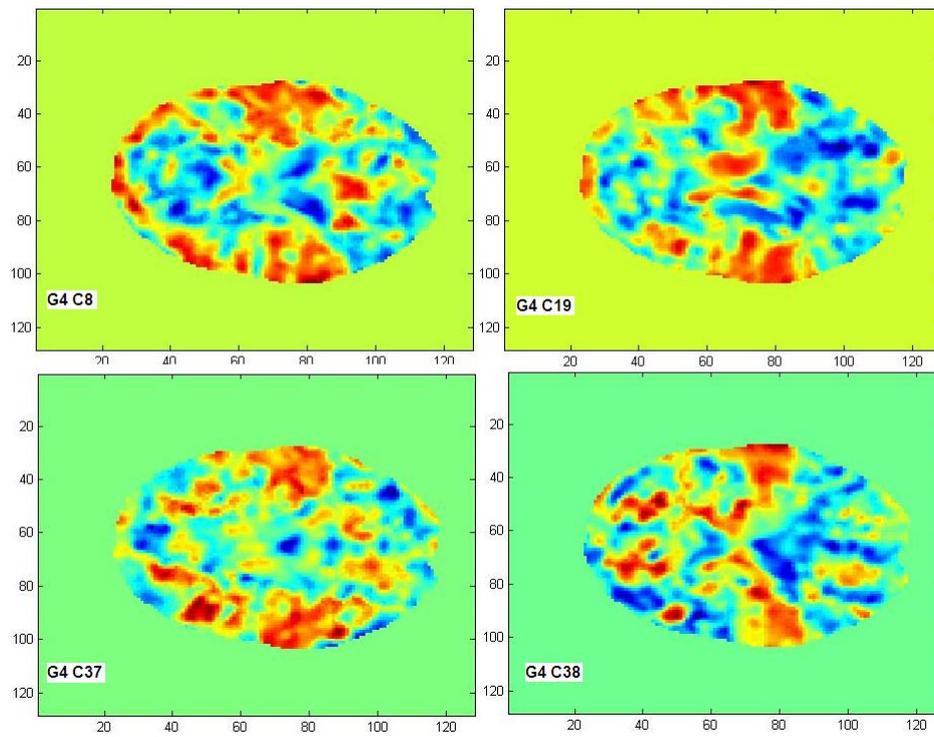


Figura 4.14: Imagens resultante da estimulação auditiva no córtex cerebral segundo análise do FastICA.

Verifica-se que mesmo utilizando a melhor função, G_4 , os resultados não mostraram resultados satisfatórios. A marcação dos pixels ativos não foram bem delimitadas pelo método, de forma que métodos já utilizados em dados de fMRI, [7,1], mostraram um resultado mais satisfatório do que os obtidos pelo ICA.

Capítulo 5

Considerações Finais

Foi utilizado para análise dos dados o método de ICA onde primeiramente foi feito um pré-processamento dos dados de fMRI utilizando o software SPM8. Também foi aplicado um pré-processamento característico do ICA, onde os dados foram centralizados e posteriormente branqueados.

No processo de estimativa da matriz de mistura e desmistura foi utilizada a aproximação da negentropia como função custo, sendo aplicado o algoritmo FastICA. Antes de aplicar o algoritmo FastICA, foi realizado a Decomposição em Valores Singulares (SVD) dos dados pré-processados.

Para validar o método foi feito a análise das curvas ROC utilizando os dados artificiais.

Os resultados obtidos nos dados sintéticos demonstraram um desempenho satisfatório quando utilizada a função G_4 , no entanto, ao aplicar o método nos dados reais os resultados se mostraram pouco expressivo, podendo ser observado visualmente nas imagens das componentes C_8 , C_{19} , C_{37} e C_{38} .

No entanto, a utilização do SVD mostrou-se muito eficiente na otimização do processamento computacional, sem o SVD o tempo computacional é imcomparavelmente maior.

Propõe-se para trabalhos futuros a aplicação da Decomposição em Valores Singulares Supervisionados (SSVD) de maneira a minimizar os efeitos dos picos encontrados em dados reais, por terem grande interferência nos resultados.

Referências Bibliográficas

- [1] Alencar, E. B. de, 2010. “Distribuições de probabilidade multiescala em fMRI”. Dissertação de Mestrado, Universidade Federal do Pará.
- [2] Araujo, D. B. de, 2002. “Sobre neuroimagens funcionais por magnetoencefalografia e ressonância magnética: novos métodos e aplicações”, Tese (Doutorado). Universidade de São Paulo, FFCLRP.
- [3] Bamber, D., 1975. “The Area above the Ordinal Dominance Graph and the Area below the Receiver Operating Characteristic Graph”. *J. Math Psychol* 12, 387-415.
- [4] Beck, J. R. and Shultz, E. K., 1986. “The Use of Relative Operating Characteristic (ROC) Curves in Test Performance Evaluation”. *Arch Pathol Lab. Med.* 110, 13-20.
- [5] Bell, A., Sejnowski, T. J., 1995. “An Information Maximization Approach to Blind Separation and Deconvolution”. *Neural Computing*, vol. 7, no. 6, pp. 1129-1159.
- [6] Buxton, R. B., 2009. “Introduction to Functional Magnetic Resonance Imaging. Principles and Techniques”. Cambridge University Press, New York, USA.
- [7] Campelo, A. D. S., 2010. “Análise de dados de Imagem por Ressonância Magnética Funcional utilizando Mapas auto-organizáveis”. Dissertação de Mestrado, Universidade Federal do Pará.
- [8] Cardoso, J.F., 1998. “Blind Separation: Statistical Principles”. *Proceedings of IEEE*, 86(10), p. 2009-2025.
- [9] Cao, X.R., Liu, R.W., 1996. “General Approach to Blind Source Separation”. *IEEE Trans. on Signal Processing*, 44(3), pp. 562-571.
- [10] Comon, P., 1994. “Independent Component Analysis, A New Conception Signal Processing”. *Special Issue on High Order Statistics*, n. 36, p. 287-314.

- [11] Fortuna, D. R. da , 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.
- [12] Friston, K. J., Jezzard, P. and Turner, R., 1994. “Analysis of Functional MRI Time-Series”. *Human Brain Mapping* 1,153-171.
- [13] 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”. *Human Brain Mapping* 2, 189-210.
- [14] 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.
- [15] Friston K. J., 2003. “Experimental Design and Statistical Parametric Mapping”. The Wellcome Dept. Of Cognitive Neurology, University College London.
- [16] Geus, K. e Troiano, M., 2004. “Visualização de regiões de ativação cerebral por fMRI sobre volumes multimodais”. I Workcomp Sul.
- [17] Glover, G. H., 1999. “Deconvolution of impulse response in event-related BOLD fMRI”. *NeuroImage*, 9:416-429.
- [18] Golub, G. H. e Van Loan, C. F., 1989. “Matrix Computations”. John Hopkins University Press.
- [19] Gonzalez, R.C. e Woods, R.E., 2000. “Processamento de Imagens Digitai”, Edgard Blucher Ltda.
- [20] Hanley, J. A.,1996. “The use of the binormal model for parametric roc analysis of quantitative diagnostic tests”. *Statistics in Medicine*, 15, 1575-1585.
- [21] Hyvarinem, A., Oja, E., Karhunen, J., 2001. “Independent Component Analysis”. John Wiley e Sons.
- [22] Hyvarinem, A., Oja, E., 2000. “Independent Component Analysis: Algorithms and Applications”. *Neural Networks*, 13(4-5), 411-430.
- [23] Hyvarinem, A., 1999. “Survey on Independent Component Analysis”. *Neural Computing Surveys*, 2, 94-128.

- [24] Hyvarinem, A., 1999(b). "Gaussian Moments for Noisy Independent component analysis". IEEE Signal Processing Letters, v. 6, n. 6, p. 145-147.
- [25] Hyvarinem, A., 1999(c). "Fast and Robust Fixed-Point Algorithms for Independent Component Analysis". IEEE Trans. on Neural Networks, 10(3), p. 626-634.
- [26] Hyvarinem, A., Oja, E., 1997. "A Fast Fixed Point Algorithm for Independent Component Analysis". Neural Computation, 9(7), p.1483-1492.
- [27] Hyvarinem, A., 1997 (b). "One Unit Contrast Function for Independent Component Analysis: A Statistical Analysis". Proc. IEEE Workshop on Neural Network for Signal Processing, p.388-397, Florida.
- [28] Jutten, C., Héroult, J., 1991. "Blind Separation of Sources: An Adaptive Algorithm based on Neuromimetic Architecture". Signal Processing, v.24, p.1-10.
- [29] Kodidis, E., 2001. "Blind Source Separation: Fundamentals and Recent Advances". XIX Simpósio Brasileiro de Telecomunicações (SBrT2001). Mini-curso. Fortaleza, CE, Set.
- [30] Lange, N. A. E Zeger, S. L., 1997. "Non-linear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging". Applied Statistics, 46, 1-29.
- [31] 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.
- [32] Ogawa S, Lee T. M., Nayak A. S., Glynn P., 1990. "Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields". Magn Reson Med. Apr;14(1):68-78.
- [33] Papoulis, A., 1991. "Probability, Random Variables and Stochastic Process". Mc Graw Hill, 3ed.
- [34] 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.
- [35] SPM from the Wellcome Department of Cognitive Neurology <http://www.fil.ion.ucl.ac.uk/SPM>.
- [36] 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.

- [37] Thulborn, K. R., Waterton, J. C., Matthews, P. M., e Radda, G. K., 1982. "Oxygenation dependence of the tranverse relaxation time of water protons in whole blood at highfield". *Biochimica et Biophysica Acta*, 714, 265-270.