NOVA ABORDAGEM PARA A SEPARAÇÃO CEGA DE FONTES EM AMBIENTES REVERBERANTES

Felipe Sander Pereira Clark

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Elétrica.

Orientador: Mariane Rembold Petraglia

Rio de Janeiro
Dezembro de 2013

NOVA ABORDAGEM PARA A SEPARAÇÃO CEGA DE FONTES EM AMBIENTES REVERBERANTES

Felipe Sander Pereira Clark

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Examinada por:

Prof. Mariane Rembold Petraglia, Ph.D.

Prof. Fernando Gil Vianna Resende Junior, Ph.D.

Prof. Jules Ghislain Slama, D.Sc.

Prof. Diego Barreto Haddad, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
DEZEMBRO DE 2013

Clark, Felipe Sander Pereira

Nova Abordagem para a Separação Cega de Fontes em Ambientes Reverberantes/Felipe Sander Pereira Clark. – Rio de Janeiro: UFRJ/COPPE, 2013.

XI, 63 p.: il.; 29, 7cm.

Orientador: Mariane Rembold Petraglia

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2013.

Referências Bibliográficas: p. 53 – 56.

1. Separação cega de fontes. 2. Reverberação. 3. Arranjo direcional de microfones. 4. Direção de chegada. 5. Algoritmos não-supervisionados. 6. Reconstrução senoidal. 7. Solução de Wiener. I. Petraglia, Mariane Rembold. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

Dedico este trabalho a todos os meus familiares e amigos, pois foram estas pessoas que me deram estímulo para realizá-lo.

Agradecimentos

Agradeço às seguintes pessoas pelo papel especial que desempenharam no progresso deste trabalho:

Mariane Rembold Petraglia

Por ser a melhor orientadora que um aluno pode desejar. Sempre compreensiva, permitiu que eu conciliasse a minha vida profissional e acadêmica com harmonia e tranquilidade. O término deste trabalho se deve, sobretudo, ao papel fundamental que ela desempenhou ao longo desses anos.

Ana Maria Sander Pereira Clark e Antonio Castello Branco Clark Filho

Meus pais, por terem me dado todos os subsídios necessários para que eu chegasse ao momento da divulgação desse trabalho, incluindo não só uma boa educação, mas também o carinho familiar e a estabilidade emocional para tal.

Rafael Sander Pereira Clark

Um irmão-aluno exemplar desde a época da escola, sempre me servindo de inspiração. Seu suporte durante os anos de desenvolvimento deste trabalho foi essencial para que eu chegasse até aqui.

Vera Lúcia Sander Pereira

A minha tia que sempre oferece apoio incondicional e consegue me reanimar mesmo nos momentos de maior dificuldade.

Claudia Spector

Minha namorada, melhor amiga e companheira há dez anos. Seu amparo foi precioso, tanto durante o desenvolvimento dessa dissertação, como em todas as fases da vida que passamos juntos.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

NOVA ABORDAGEM PARA A SEPARAÇÃO CEGA DE FONTES EM AMBIENTES REVERBERANTES

Felipe Sander Pereira Clark

Dezembro/2013

Orientador: Mariane Rembold Petraglia

Programa: Engenharia Elétrica

Os algoritmos de separação cega de fontes no domínio da frequência tipicamente são computacionalmente menos custosos que os métodos no domínio do tempo, sendo empregados em contextos de misturas convolutivas com disponibilidade de recursos computacionais limitados, tais como na separação de sinais de voz em tempo real e redução de interferências em arranjos de microfones. Estes algoritmos são, em geral, processos adaptativos cujas funções a serem minimizadas são não quadráticas, necessitando, portanto de um ponto de partida adequado.

Neste sentido, apresenta-se nesta dissertação uma nova técnica de inicialização de processos de separação cega de sinais de voz no domínio da frequência que enfoca na sua aplicação em ambientes reverberantes. O método desenvolvido emprega arranjos diretivos de microfones, ressíntese dos sinais através de osciladores senoidais e filtragem de Wiener. Como método de aprimoramento da solução inicial obtida, adotou-se o algoritmo de exploração de dependências estatísticas de alta ordem entre componentes de frequência dos sinais.

Os resultados obtidos com o método proposto são comparados aos de outros métodos já consolidados na literatura. Os experimentos foram conduzidos em diversos ambientes acústicos, tanto simulados quando reais, nos quais as misturas de sinais de voz foram gravadas por um conjunto de microfones.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

NEW APPROACH TO BLIND SOURCE SEPARATION IN REVERBERANT ENVIRONMENTS

Felipe Sander Pereira Clark

December/2013

Advisor: Mariane Rembold Petraglia

Department: Electrical Engineering

Frequency domain blind source separation algorithms usually have a lower computational cost when compared to time domain methods, being, therefore, employed with convolutive mixtures when computational resources are limited, in applications such as voice signal separation in real time and interference reduction in microphone arrays. These algorithms are, in general, adaptive processes whose cost functions are non-quadratic, requiring a well-chosen starting point.

In this work a novel initialization procedure for frequency domain blind source separation of voice signals in reverberant environments is presented. The proposed method employs beamforming techniques, sinusoidal synthesis and Wiener filtering. The blind source separation algorithm based on exploitation of higher order frequency dependencies is employed in order to derive the final solution.

All results obtained with the new initialization are compared with other well-consolidated approaches. The experiments are conducted considering various acoustic environments, including simulated and real rooms in which the voice signals were recorded by a microphone array.

Sumário

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

Separação cega de fontes é o processo através do qual segrega-se de um conjunto de misturas registradas por um grupo de sensores uma ou mais informações de interesse, reduzindo-se as interferências. Um exemplo clássico desta aplicação é a chamada festa do coquetel (*cocktail party*), na qual deseja-se descobrir o que foi dito por alguma(s) pessoa(s) em um bar ao longo da noite, tendo-se como dados de entrada somente as gravações feitas por um conjunto de microfones dentro do estabelecimento, que conterão misturas das vozes de todos os frequentadores do local.

Os algoritmos que solucionam este tipo de desafio são, normalmente, algoritmos adaptativos cujas funções a serem minimizadas são não quadráticas, necessitando, portanto, de um ponto de partida adequado. Neste sentido, a técnica de inicialização clássica apresentada na literatura é o branqueamento das misturas (vide o Capítulo 2 para maiores detalhes). Em [1, 2] uma nova técnica foi proposta, voltada para algoritmos de separação cega de fontes no domínio da frequência. Nesse contexto, demonstrou-se que a modificação do método de partida dessa classe de algoritmos seria capaz de otimizar, quando comparada à abordagem por branqueamento, o resultado da segregação das fontes, sob a ótica da métrica razão sinal-interferência (do inglês *Signal to Interference Ratio* - SIR, detalhada no Capítulo 7).

Nesta nova proposta de inicialização é sugerida a adoção em cascata de técnicas de estimação da direção de chegada de fontes sonoras, seguida da aplicação de estratégias de mascaramento de frequências e clusterização. O resultado dessa sequência algorítmica é então submetida à linearização por filtragem de Wiener e entregue como insumo aos métodos de separação cega de fontes no domínio da frequência. A representação sistêmica dessa proposta é apresentada na Fig. 1.1.

O sistema supracitado, embora robusto quando aplicado em diversos cenários, demonstrou-se limitado quando utilizado com misturas gravadas em ambientes com muita reverberação, pois, nestes cenários, as imagens das fontes prejudicam severamente o processo de separação [4]. Essa limitação não é simples de ser contornada,

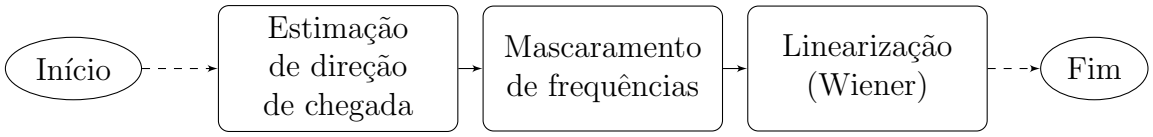


Figura 1.1: Método de inicialização proposto em [1, 2]

permanecendo este cenário como grande desafio para o estado da arte em separação cega de fontes [5, 6]. Nesta esfera, a busca de novas técnicas permanece constante e, reforçando a ideia de que a separação cega de fontes pode se beneficiar de técnicas de mascaramento de frequências, outra linha de pesquisa, conhecida como *Computer Auditory Scene Analysis* - CASA, ganhou notoriedade.

A CASA estabelece como meta o uso de conhecimentos acerca da fisiologia da audição humana para o desenvolvimento de algoritmos de separação cega de fontes [7]. Sob a luz dessas novas técnicas, no presente trabalho busca-se um novo método de inicialização que explore peculiaridades do processo fisiológico auditivo, com o intuito de desenvolver um algoritmo de separação cega de fontes que apresente maior robustez quando alimentado por misturas de fontes em ambientes reverberantes.

Duas dessas características passíveis de serem introduzidas no processo de separação cega de fontes são a capacidade de concentração da nossa audição em apenas uma dentre muitas fontes [8] e o fenômeno conhecido como mascaramento de frequências. Essa peculiaridade da audição humana indica, dentre outros aspectos, que quando duas frequências muito próximas com diferença notável de energia excitam nosso sistema auditivo, apenas aquela de maior energia é percebida [7, 9]. Outra característica importante é o funcionamento da cóclea, órgão que interpreta os sons que ouvimos como uma combinação de frequências que excitam terminais nervosos [9].

Considerando essas características do sistema auditivo e buscando mimetizá-las através de técnicas de processamento de sinais, trabalhar as misturas no domínio da frequência e simular a capacidade de concentração da nossa audição em apenas uma fonte dentre muitas através de técnicas de arranjo diretivo de microfones parecem promissores. Ademais, ressintetizá-las através de componentes senoidais discretos - destacando apenas aqueles que apresentam maior razão sinal reverberação e maior razão sinal ruído - é uma solução potencialmente benéfica, visto que aproxima a capacidade de mascaramento da nossa audição.

O acima exposto originou a nova proposta de inicialização para o problema de separação cega de fontes que será apresentada nessa dissertação. Aplicando o processo de diretividade em arranjo de microfones (em inglês *beamforming* [10]) aliado à decomposição senoidal apresentada em [11, 12] (semelhantemente ao proposto em [13]) e a uma versão otimizada do processo de linearização por filtragem de Wiener empregado em [1, 2], espera-se obter um método de inicialização mais eficaz para a

Figura 1.2: Novo método de inicialização.

separação de misturas reverberantes de voz. A representação sistêmica desse novo método é apresentada na Fig. 1.2, que em suma busca gerar um pré-processamento que minimize o efeito da reverberação nas misturas, fazendo com que o algoritmo de separação seja iniciado com misturas de mais fácil tratamento, sem, contudo, haver o objetivo de completa remoção da reverberação nos sinais estimados ao fim do processo.

A contribuição deste trabalho reside, portanto, no estudo, implementação computacional e análise dos resultados dessa nova proposta de inicialização de algoritmos de separação cega de fontes, de forma a simplificar a solução do problema de segregação de fontes em contextos reverberantes.

1.1 Organização da Dissertação

No Capítulo 2 apresenta-se uma breve introdução teórica ao processo de separação cega de fontes.

No Capítulo 3 serão apresentados métodos de arranjos diretivos de microfones, com foco especial dado à técnica inicialmente introduzida em [14] e que será usada como primeiro passo do algoritmo de inicialização proposto nessa dissertação.

No Capítulo 4 será descrito o método de análise senoidal proposto em [11] e que complementa a inicialização supracitada.

No Capítulo 5 o procedimento *Exploitation of Higher-Order Frequency Dependencies* exposto em [15] será introduzido como meio de aprimorar a solução obtida pelo método de inicialização proposto.

No Capítulo 6 será demonstrado como gerar a solução inicial, a partir dos sinais resultantes das técnicas descritas nos Capítulos 3 e 4, para os métodos de separação cega de fontes no domínio da frequência, em especial para o método apresentado no Capítulo 5.

No Capítulo 7 os ambientes de testes e os resultados obtidos pelo método proposto serão descritos, sendo comparados aos dos métodos convencionais.

No Capítulo 8, apresentar-se-ão as conclusões e as perspectivas de trabalhos futuros.

O Apêndice A contém o formulário adotado para o levantamento dos resultados dos testes subjetivos apresentados no Capítulo 7.

Finalmente, o Apêndice B apresenta a transcrição dos monólogos que constituíram as misturas usadas nos testes de separação cega de fontes apresentados ao longo desta dissertação.

1.1.1 Notação Matemática Adotada

A seguinte notação matemática será empregada ao longo deste trabalho: variáveis escalares e funções serão denotadas por letras não formatadas (exemplo: x, X, $f(x)$, $F(x)$); vetores (exceto no domínio da frequência) serão indicados por letras minúsculas em negrito (exemplo: $\mathbf{x}$) e vetores no domínio da frequência ou matrizes serão representadas por letras maiúsculas em negrito (exemplo: $\mathbf{X}$), sendo diferenciados pelo contexto. Quanto aos vetores, salvo quando expresso o contrário, serão considerados vetores coluna.

Em relação às operações matemáticas que serão empregadas ao longo desta dissertação, $(\cdot)^*$ indica conjugação de valores complexos; $(\cdot)^H$ denota a transposição de vetor ou matriz seguida de conjugação de valores complexos; $(\cdot)^T$ representa a simples transposição vetorial ou matricial; e E[$\cdot$] significa a operação de valor esperado estatístico.

Capítulo 2

Fundamentação Teórica em Separação Cega de Fontes

A separação cega de fontes consiste na recuperação de N fontes individuais a partir de M misturas [3, 16]. Diz-se que esta separação é feita de maneira cega, pois não é assumido qualquer conhecimento prévio dos sinais individuais que compõem as misturas e tampouco do sistema responsável por elas. Formalmente, se denominarmos de $\mathbf{s}(n) = [s_1(n) \quad s_2(n) \cdots s_N(n)]^T$ o vetor composto pelas amostras das fontes individuais no instante n e de $\mathbf{x}(n) = [x_1(n) \quad x_2(n) \cdots x_M(n)]^T$ o vetor composto pelas amostras das misturas observadas no instante n, podemos relacioná-los, supondo misturas lineares, pela equação:

$$\mathbf{x}(n) = \mathbf{H} * \mathbf{s}(n), \tag{2.1}$$

em que $\mathbf{H}$ é a matriz característica do sistema de mistura, denominada matriz de mistura, cuja dimensão é $M \times N$, e o operador $*$ representa a convolução. Adotando o domínio da transformada $\mathcal{Z}$ e assumindo que o sistema de mistura é um sistema causal, podemos representar os elementos de $\mathbf{H}(z)$ - que contém às respostas ao impulso dos diversos caminhos percorridos pelos sinais até os sensores - genericamente por:

$$H_{ij}(z) = \sum_{l=0}^{L-1} h(l) z^{-l}, \tag{2.2}$$

em que se observa que o comprimento destes filtros está diretamente associado ao tempo de reverberação do ambiente de mistura representado por $\mathbf{H}(z)$.

Desta formulação destacam-se dois casos relevantes: quando $L = 1$ e os filtros $H_{ij}(z)$ são apenas valores escalares, as misturas são ditas instantâneas. Caso $H_{ij}(z) = h(l_{ij}) z^{-l_{ij}}$, temos misturas denominadas anecoicas (ou não reverberantes). Destacam-se também as misturas não lineares, que envolvem relações não lineares

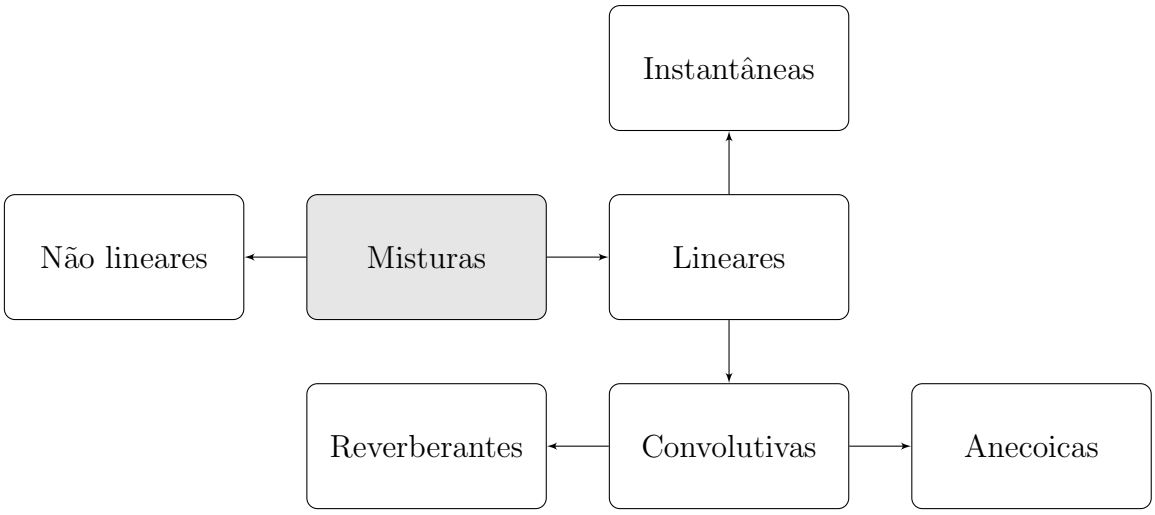


Figura 2.1: Tipos de misturas.

entre os sinais das fontes e os dos sensores, e que não serão abordadas neste trabalho. Sintetizamos estas possibilidades na Fig. 2.1.

Neste ponto, cabe fazermos algumas observações referentes à relação entre M e N. Quando $M < N$ diz-se que temos um problema de separação de fontes indeterminado; quando $M > N$ temos o caso superdeterminado e quando $M = N$ denomina-se o caso de determinado.

Vista esta classificação, nota-se a analogia existente entre separação cega de fontes e a solução de sistemas de equações. De fato, podemos concretizar esta analogia se pensarmos que as variáveis de um sistema de equações são as N fontes individuais que desejamos obter e que as M misturas são as equações de que dispomos. Portanto, assim como na resolução de sistemas de equações, o problema de separação de fontes tem sua dificuldade dependente da relação entre M e N, sendo a solução ótima mais difícil de ser encontrada quando $N > M$ e mais simples quando $N \leq M$.

Ademais, conforme representado pela Eq. (2.2), outra dificuldade que surge quando se desenvolve técnicas de separação de fontes é o desconhecimento *a priori* do número de coeficientes dos filtros de mistura $H_{ij}(z)$. A Fig. 2.2 resume a complexidade de solução da separação de fontes em função dessa análise.

Visando a superar estas dificuldades, as soluções para separação de fontes mais comuns têm como pressuposto o fato de que diferentes sensores capturam diferentes misturas e a conjectura de que as fontes misturadas são estatisticamente independentes. Neste sentido, uma valiosa informação inicial é sabermos que podemos representar a função densidade de probabilidade conjunta do vetor de fontes $\mathbf{s}$ como o produto das densidades marginais, isto é:

$$q(\mathbf{s}) = q_1(s_1)q_2(s_2)\cdots q_N(s_N) = \prod_{i=1}^{N} q_i(s_i) \tag{2.3}$$

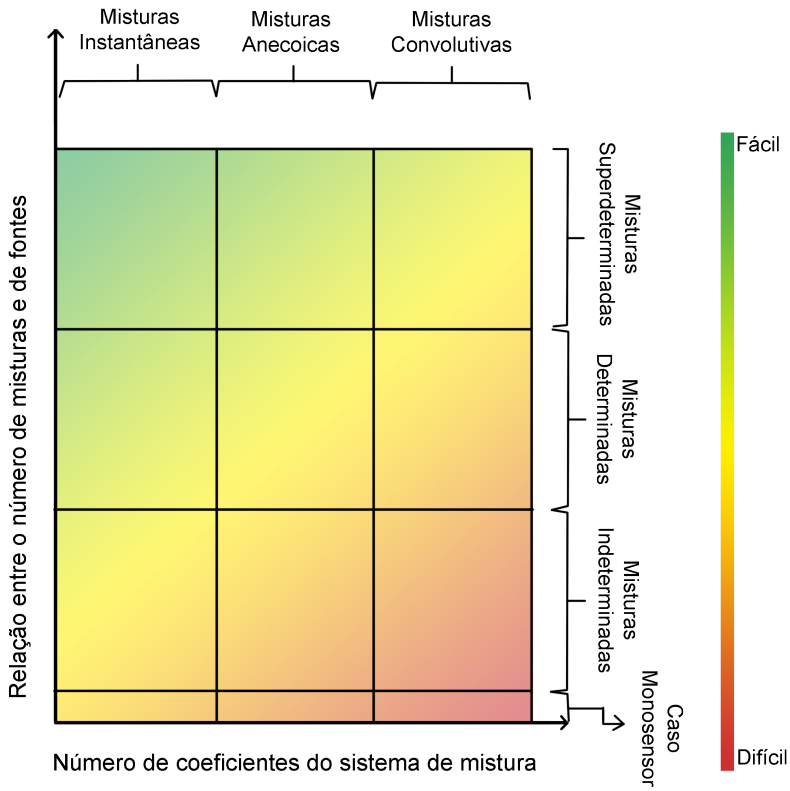


Figura 2.2: Complexidade da separação de fontes (adaptado de [3])

em que $q_i(\cdot)$ representa a função densidade de probabilidade da i-ésima fonte.

Comumente, a obtenção das fontes é iniciada por uma transformada de domínio [3], principalmente quando trabalhamos com misturas convolutivas e/ou subdeterminadas. No caso das misturas convolutivas, costumam-se empregar métodos no domínio da frequência, já que nesta representação as operações de convolução - computacionalmente custosas - tornam-se produtos. No caso de misturas subdeterminadas, é conveniente aplicar transformações esparsificadoras, como a transformada de Fourier em quadros curtos e a *wavelet* [17–20], para permitir a extração das fontes. Entretanto, é importante enfatizar que operações no domínio do tempo também podem prover bons resultados, sobretudo quando as misturas são instantâneas, anecoicas ou mesmo convolutivas determinadas [21].

Em seguida, para o caso das misturas determinadas ou superdeterminadas, faz-se a estimativa da(s) matriz(es) de mistura (ou diretamente de sua(s) inversa(s)) a partir dos coeficientes da transformada e, finalmente, faz-se a transformação inversa para obtenção das fontes individuais no domínio do tempo. No caso das misturas indeterminadas, não podemos recompor as fontes através da estimativa da(s) matriz(es) de mistura (ou sua(s) inversa(s)), sendo necessário estimar os coeficientes das fontes no domínio da transformada.

Quando consideramos apenas os casos determinado ou superdeterminado (os casos de interesse neste trabalho), podemos resumir estes passos afirmando simples-

mente que desejamos encontrar a matriz $\mathbf{G}$ que retorna

$$\mathbf{y}(n) = \mathbf{G} * \mathbf{H} * \mathbf{s}(n) = \mathbf{C} * \mathbf{s}(n) \tag{2.4}$$

como estimativa das N fontes.

Fica claro, portanto, que idealmente, para o caso instantâneo, $\mathbf{G} = \mathbf{H}^{-1} \rightarrow \mathbf{C} = \mathbf{GH} = \mathbf{I}$, em que $\mathbf{I}$ é uma matriz identidade de ordem N. Já para o caso convolutivo, buscamos filtros $\mathbf{G}$ que, combinados com os filtros $\mathbf{H}$, geram filtros $\mathbf{C} = \mathbf{G} * \mathbf{H}$ capazes de reconstituir versões filtradas das fontes que não apresentem interferências. Todavia, visando alcançar estas metas, em geral os algoritmos de separação cega de fontes trabalham por otimização - frequentemente empregando métodos de gradiente - de uma função custo com restrições sobre $\mathbf{G}$, o que origina dois problemas típicos desta classe de algoritmos: o escalamento da matriz $\mathbf{C}$ e a permutação de suas colunas.

O primeiro problema se deve à nossa ignorância sobre o nível dinâmico das fontes, relacionado ao fato de as misturas manterem-se inalteradas caso multipliquemos a i-ésima fonte por um escalar e dividamos a i-ésima coluna da matriz de misturas pelo mesmo escalar.

Uma segunda ambiguidade se origina da falta de conhecimento *a priori* sobre as informações que se deseja separar. Neste cenário, não é possível distinguir qualquer permutação dos dados de entrada - já que a indexação é uma convenção arbitrária - e, portanto, não se pode afirmar categoricamente que houve permutação das saídas. Por este motivo, qualquer ordenação de resultados distinta é válida, sendo comum que as saídas dos mecanismos de separação cega de fontes alternem-se aleatoriamente após cada execução, a partir de arranjos idênticos.

Estes obstáculos tornam-se críticos nas abordagens no domínio da frequência [3], contexto no qual a solução típica é um conjunto de matrizes $\mathbf{H}_k$ que separam individualmente cada uma das k raias das STFTs (vide Seção 2.1) das fontes (assumindo independência) a partir das STFTs das misturas. Nesta configuração, o problema de escalamento implica equalizar as frequências de maneira irregular e o de permutação significa permitir alternância da ordem das componentes de diferentes frequências de cada fonte nas saídas do sistema.

Das diversas técnicas propostas para superar esta dificuldade, este texto abordará, para o primeiro caso, o princípio da distorção mínima para normalizar os coeficientes da matriz de separação, e, para o segundo, o uso de informações estatísticas de alta ordem para decidir quais frequências são pertinentes a cada fonte. Estas abordagens serão desenvolvidas no Capítulo 5.

Neste ponto, visto que o cálculo da matriz $\mathbf{G}$ é comumente feito por métodos de otimização, é interessante averiguarmos como eles são tipicamente inicializados.

É sabido que é possível transformar qualquer mistura de componentes descorrelacionados em um conjunto de componentes independentes através da computação da transformação linear ortogonal correspondente [21].

Portanto, a maneira mais comum de se inicializar $\mathbf{G}$ é através de uma matriz que torne as misturas descorrelacionadas - artifício conhecido como branqueamento ou esferização. Esta técnica implica não só obtermos elementos descorrelacionados, mas, ainda, com variância unitária, facilitando que o método de otimização consiga obter estimativas independentes. Portanto, para um vetor de componentes aleatórias e média zero $\mathbf{z} = [z_1 \cdots z_N]^T$ branqueado, podemos atestar que

$$\begin{aligned} \mathrm{E}[z_i z_j] &= \delta_{ij} \\ \mathrm{E}[\mathbf{z}\mathbf{z}^H] &= \mathbf{I} \end{aligned} \tag{2.5}$$

com δ_{ij} representando elementos unitários para $i = j$ e nulos para $i \neq j$, e $\mathbf{I}$ representando a matriz identidade.

Vemos, portanto, que o objetivo da técnica de branqueamento é a obtenção de uma matriz $\mathbf{V}$ que transformará o vetor de componentes aleatórias $\mathbf{x}$ em outro vetor $\mathbf{z}$ que atenda às Eqs. (2.5).

A solução para este problema é conhecida [21] e parte da decomposição em autovalores e autovetores da matriz de covariância $\mathbf{C_x} = \mathrm{E}[\mathbf{x}\mathbf{x}^H]$. Denotando por $\mathbf{E} = [\mathbf{e}_1, \cdots, \mathbf{e}_N]$ a matriz cujas colunas são os autovetores de norma unitária de $\mathbf{C_x}$ e por $\mathbf{D} = \mathrm{diag}[d_1, \cdots, d_N]$ a matriz diagonal de autovalores de $\mathbf{C_x}$, a matriz $\mathbf{V}$ é dada por:

$$\mathbf{V} = \mathbf{D}^{-1/2}\mathbf{E}^H, \tag{2.6}$$

sendo necessário que os autovalores sejam não-nulos para que $\mathbf{V}$ exista. Na prática, esta restrição não é impeditiva à aplicação do método sobre sinais naturais, como sinais de voz.

Observando que qualquer transformação ortogonal sobre a matriz $\mathbf{V}$ é igualmente válida como matriz de branqueamento, destaca-se a seguinte matriz simétrica comumente empregada:

$$\mathbf{V} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^H \tag{2.7}$$

Para comprovarmos a eficácia deste método, reescrevemos $\mathbf{C_x} = \mathbf{E}\mathbf{D}\mathbf{E}^H$, com $\mathbf{E}$ satisfazendo $\mathbf{E}^H\mathbf{E} = \mathbf{E}\mathbf{E}^H = \mathbf{I}$ (ou seja, uma matriz unitária) e verificamos que:

$$\mathrm{E}[\mathbf{z}\mathbf{z}^H] = \mathrm{E}[\mathbf{V}\mathbf{x}\mathbf{x}^H\mathbf{V}^H] = \mathbf{V}\mathrm{E}[\mathbf{x}\mathbf{x}^H]\mathbf{V}^H = \mathbf{D}^{-1/2}\mathbf{E}^H\mathbf{E}\mathbf{D}\mathbf{E}^H\mathbf{E}\mathbf{D}^{-1/2} = \mathbf{I}, \tag{2.8}$$

ou seja, a matriz de covariância de $\mathbf{z}$ é realmente uma matriz identidade, comprovando o branqueamento.

Conforme apresentado no Capítulo 1, esta dissertação investigará uma alternativa ao método de branqueamento, buscando melhores resultados no contexto da separação cega de fontes em ambientes reverberantes.

2.1 Transformada de Fourier em Janelas Curtas

Todos os algoritmos no domínio da frequência que serão abordados ao longo dos próximos capítulos adotam a transformada de Fourier em janelas curtas (em inglês, *Short Time Fourier Transform* - STFT). Assim, cumpre-nos definir esta operação aplicada ao sinal $x_i(n)$ como

$$X(k,m) = \sum_{l=0}^{K-1} x_i(mJ+l)h(l)\mathrm{e}^{-j2\pi kl/K}, \quad k = 0, \cdots, K-1, \tag{2.9}$$

sendo k o índice da raia espectral, K o total de raias, $h(l)$ uma função de janelamento que converge suavemente para zero em suas extremidades, J o deslocamento (em número de amostras) entre quadros consecutivos e m o índice do quadro.

A transformada inversa (ISTFT) pode ser obtida por:

$$x_i(mJ+l)h(l) = \frac{1}{K}\sum_{k=0}^{K-1} X(k,m)\mathrm{e}^{j2\pi kl/K}, \quad l = 0, \cdots, K-1. \tag{2.10}$$

Capítulo 3

O Arranjo Diretivo de Microfones

A técnica de arranjo diretivo de microfones (em inglês *beamforming*) é o método de processamento de sinais pertencente à classe dos algoritmos de tratamento de dados em grupamentos de transdutores que visa enfatizar ou atenuar sinais oriundos de uma direção específica [22]. Quando aplicado no contexto da separação cega de fontes, não só obtemos uma separação preliminar das fontes quando priorizamos a direção de cada sinal desejado, mas também minimizamos o efeito da reverberação, visto que esta não tem direção específica.

Os processos dessa classe normalmente têm como entrada as informações coletadas por um conjunto de sensores posicionados segundo uma geometria específica e que usualmente possuem as mesmas características construtivas, dentre elas, a isotropia. Este arranjo pode ser linear (unidimensional), planar (bidimensional) ou volumétrico (tridimensional), com espaçamento regular uniforme, regular não uniforme ou até mesmo espaçamento irregular entre componentes. No contexto deste trabalho, consideraremos apenas os arranjos lineares com espaçamento regular uniforme de dois elementos.

Neste âmbito, a característica fundamental que permite o alcance do objetivo supracitado é o conhecimento de que a informação (neste caso, sinais de voz) alcança cada sensor (microfone) em instantes diferentes. A Fig. 3.1 apresenta este conceito, nos permitindo depreender que esta diferença de tempo se dá por uma distância representada por $\delta \approx d\,\mathrm{sen}(\theta)$, sendo d a distância entre microfones consecutivos e θ o ângulo de incidência da frente de onda sobre os microfones, desde que se suponha que a fonte sonora está suficientemente distante dos microfones, de modo que a frente de onda que os excita possa ser considerada plana.

A fim de explorar estas características, diversos métodos de *beamforming* existem na literatura [23]. Entretanto, deste universo plural, destacaremos quatro algoritmos: o método dos atrasos e somas; o método das restrições lineares com mínima variância; o algoritmo de Frost e o algoritmo de Doblinger [14]. Destes, maiores detalhes serão apresentados somente para o último caso, uma vez que, conforme

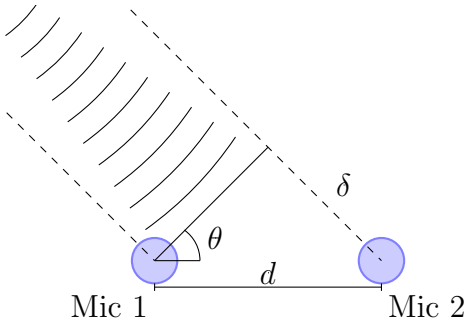


Figura 3.1: Configuração de *beamforming* para dois sensores.

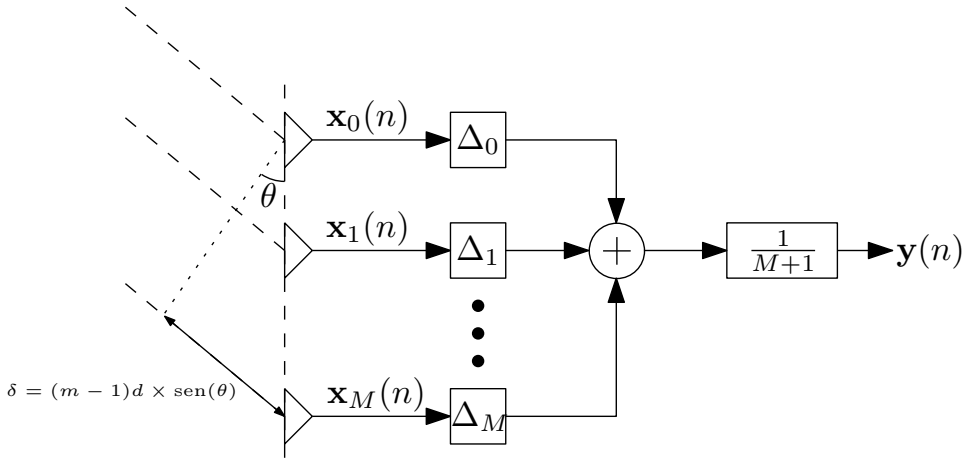


Figura 3.2: A estrutura de atrasos e somas.

comprovado através de experimentos apresentados no Capítulo 7, este método proporciona os melhores resultados para as aplicações estudadas neste trabalho.

É cabível frisar que o emprego dos métodos de *beamforming* supracitados no contexto de separação de fontes torna a aplicação não cega, pois todos eles dependem de informações acerca da direção angular das fontes em relação aos microfones. É possível, contudo, contornar esta limitação através do emprego de mecanismos automáticos de detecção da direção das fontes, como os que são apresentados em [24–27]. Maiores ponderações acerca deste tema serão desenvolvidas no Capítulo 7.

3.1 Método dos Atrasos e Somas

O método dos atrasos e somas constitui-se da compensação das diferenças de tempo de chegada entre os diferentes elementos do arranjo de sensores para um sinal oriundo de uma direção específica. Uma vez que os sensores são alinhados no tempo, estes são coerentemente somados de modo a maximizar a razão sinal ruído na direção almejada, conforme ilustra a Fig. 3.2.

O atraso aplicado ao sinal do m-ésimo sensor é obtido por

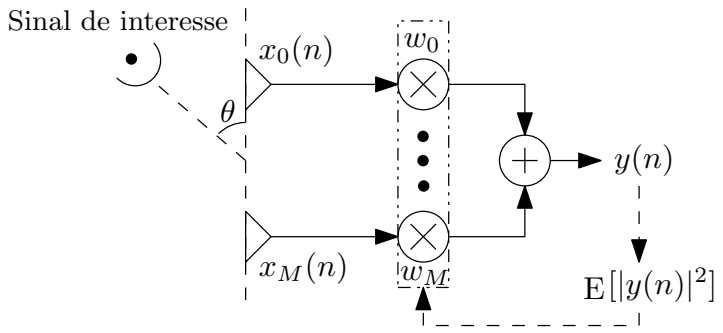


Figura 3.3: Estrutura do LCMV.

$$\Delta_m = \frac{(m-1)\, d\, \mathrm{sen}(\theta)}{c} f_s, \tag{3.1}$$

sendo f_s a frequência de amostragem empregada e d a distância entre cada par de transdutores consecutivos.

3.2 Método das Restrições Lineares e Variância Mínima

O método das restrições lineares e variância mínima (do inglês *Linearly Constrained Minimum Variance beamforming* - LCMV) consiste em um sistema adaptativo de múltiplas entradas ($x_i(n)$) e única saída ($y(n)$), tal que $y(n)$ tenha a menor energia possível, ou seja, este processamento visa ao encontro de $y(n)$ tal que $\mathrm{E}[|y(n)|^2]$ seja minimizado, conforme ilustra a Fig. 3.3. Para tal, aplica-se um conjunto de restrições ao processo de adaptação dos coeficientes de filtragem $\mathbf{w}$, de modo que o critério sobre a energia de saída seja alcançado pela atenuação das interferências que incidem sobre o arranjo de sensores por direções indesejadas, sem distorcer o sinal proveniente da direção particular que se deseja preservar [28].

Visto de modo alternativo, buscamos o filtro ótimo $\mathbf{w}_o$ que minimiza a função custo

$$\begin{array}{c} J = \mathrm{E}[|y(n)|^2] = \mathbf{w}^H \mathbf{R}_{\mathbf{xx}} \mathbf{w} \\ \text{sujeito a } \mathbf{C}^H \mathbf{w} = \mathbf{f} \end{array}, \tag{3.2}$$

em que $\mathbf{R}_{\mathbf{xx}}$ é a matriz de correlação dos dados observados nos sensores, $\mathbf{C}$ é a matriz de restrições do *beamforming* (função das direções de chegada) e $\mathbf{f}$ é o vetor de resposta desejada [22].

Através da aplicação da técnica dos multiplicadores de Lagrange $\boldsymbol{\lambda}$, encontramos a função

$$l(\mathbf{w}, \boldsymbol{\lambda}) = \mathbf{w}^H \mathbf{R}_{\mathbf{xx}} \mathbf{w} + \boldsymbol{\lambda}^H (\mathbf{C}^H \mathbf{w} - \mathbf{f}) + \boldsymbol{\lambda}^T (\mathbf{C}^T \mathbf{w}^* - \mathbf{f}^*) \tag{3.3}$$

a ser minimizada para que encontremos o conjunto ótimo de coeficientes $\mathbf{w}_o$.

É possível demonstrar [10] que esta minimização leva a:

$$\mathbf{w}_o = \mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{C}(\mathbf{C}^H\mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{C})^{-1}\mathbf{f}, \tag{3.4}$$

constituindo, por fim, o cerne do método das restrições lineares e variância mínima.

3.3 O Algoritmo de Frost

É possível depreender da Eq. (3.4) que o processo de determinação de $\mathbf{w}_o$ para o processo LCMV depende das estatísticas de segunda ordem dos dados de entrada do sistema ($\mathbf{R}_{\mathbf{xx}}$) e da inversão desta matriz. Contudo, em muitas aplicações esta estatística pode ser variável no tempo ou desconhecida, ou o custo computacional para o seu cálculo e, sobretudo, de sua inversa, pode ser impraticável.

Visando à mitigação destas dificuldades, Frost [29] propôs uma estrutura adaptativa cuja base é a atualização dos coeficientes $\mathbf{w}$ iterativamente no sentido contrário ao do gradiente da função custo da Eq. (3.3), dado por

$$\nabla_{\mathbf{w}} l(\mathbf{w}, \boldsymbol{\lambda}) = \mathbf{R}_{\mathbf{xx}}\mathbf{w} + \mathbf{C}\boldsymbol{\lambda}, \tag{3.5}$$

e o emprego da aproximação instantânea de $\mathbf{R}_{\mathbf{xx}}$, ou seja, $\tilde{\mathbf{R}}_{\mathbf{xx}}(n) = \mathbf{x}(n)\mathbf{x}^H(n)$ (assumindo $\mathbf{x}(n)$ ergódico). Deste modo, o algoritmo de Frost minimiza a energia instantânea de saída, diferentemente do método LCMV, que minimiza o valor médio quadrático.

Isto posto, o método de Frost pode ser elaborado a partir da definição do processo de minimização através do gradiente apresentado na Eq. (3.5) conforme

$$\mathbf{w}(\mathrm{n}+1) = \mathbf{w}(\mathrm{n}) - \mu(\tilde{\mathbf{R}}_{\mathbf{xx}}(\mathrm{n})\mathbf{w}(\mathrm{n}) + \mathbf{C}\boldsymbol{\lambda}(\mathrm{n})), \tag{3.6}$$

com μ representando um fator de escala que determina o passo do algoritmo.

Uma vez que $\mathbf{w}(\mathrm{n}+1)$ deve atender à restrição imposta na Eq. (3.2), podemos substituir a Eq. (3.6) em $\mathbf{C}^H\mathbf{w} = \mathbf{f}$ e solucioná-la para encontrar os multiplicadores de Lagrange. Estes multiplicadores, então, podem ser aplicados na Eq. (3.6), de onde depreendemos que

$$\mathbf{w}(\mathrm{n}+1) = \mathbf{C}(\mathbf{C}^H\mathbf{C})^{-1}\mathbf{f} + \mathbf{P}(\mathbf{w}(\mathrm{n}) - \mu\mathbf{x}(\mathrm{n})\mathbf{x}^H(\mathrm{n})\mathbf{w}(\mathrm{n})), \tag{3.7}$$

sendo $\mathbf{P} = \mathbf{I} - \mathbf{C}(\mathbf{C}^H\mathbf{C})^{-1}\mathbf{C}^H$.

A Eq. (3.7) constitui a equação de atualização do método de Frost, sendo $\mathbf{w}[0] = \mathbf{C}(\mathbf{C}^H\mathbf{C})^{-1}\mathbf{f}$ uma inicialização comumente empregada para o processo, visto que ela satisfaz a restrição apresentada na Eq. (3.2).

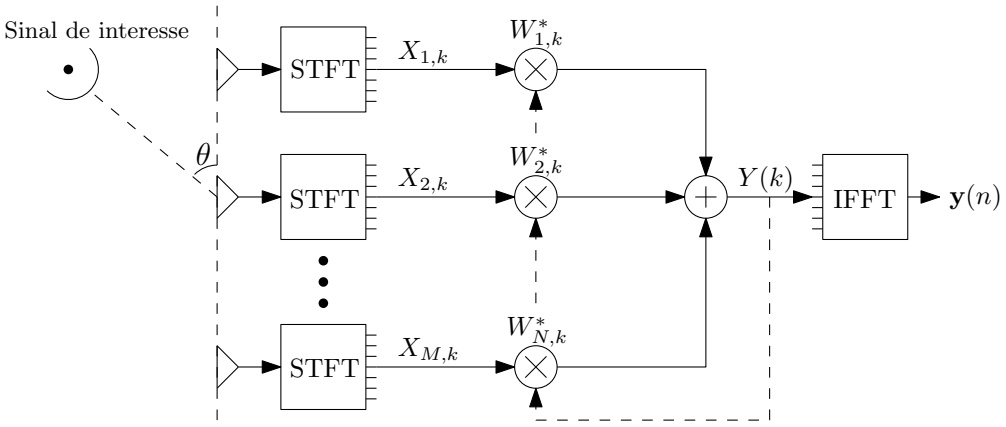


Figura 3.4: Estrutura do algoritmo de Doblinger.

3.4 O Algoritmo de Doblinger

O algoritmo de *beamforming* desenvolvido por Doblinger [14] é uma variação do método de Frost apresentado na Seção 3.3. O principal diferencial da técnica que doravante será apresentada é o domínio do processamento dos sinais, que deixarão de ser trabalhados no tempo e passarão a ser processados na frequência, permitindo o emprego de mais restrições. A estrutura deste método é apresentada na Fig. 3.4, que revela que a informação espectral de cada canal $x_{i,k}$ é modificada quadro a quadro no domínio da transformada da STFT por um conjunto de pesos complexos $w_{i,k}$.

Semelhantemente ao exposto nas seções anteriores, o desenvolvimento do método de Doblinger é formulado como um problema de otimização quadrática sujeito a restrições, ou seja, a minimização da energia de $Y(k)$ através da atenuação das interferências em direções indesejadas, preservando a informação da direção de interesse. Formalmente este processo é descrito pela função custo a ser minimizada

$$\begin{array}{c} J = \mathrm{E}[|Y(k)|^2] = \mathbf{W}_k^H \mathrm{E}[\mathbf{X}_k \mathbf{X}_k^H] \mathbf{W}_k = \mathbf{W}_k^H \mathbf{S}_{\mathbf{X}_k \mathbf{x}_k} \mathbf{W}_k \\ \text{sujeito a } \mathbf{C}_k^H \mathbf{W}_k = \mathbf{f} \end{array}, \tag{3.8}$$

em que $\mathbf{W}_k = [W_{1,k} \quad W_{2,k} \cdots W_{N,k}]^T$, $\mathbf{X}_k = [X_{1,k} \quad X_{2,k} \cdots X_{N,k}]^T$ e a matriz $\mathbf{S}_{\mathbf{X}_k \mathbf{x}_k}$ é a chamada matriz de correlação espaço-espectral de $\mathbf{X}_k$.

As restrições impostas pela matriz $\mathbf{C}_k$, por outro lado, têm que ser tais que o sinal desejado não seja atenuado, ao passo que os sinais provenientes das demais direções sejam suprimidos. Adotando-se um modelo de propagação que assume as frentes de onda que excitam os sensores como planas, a matriz $\mathbf{C}_k$ será composta por vetores

$$\mathbf{d}_k(\theta) = [e^{j\Omega_k \tau_1(\theta)} \quad e^{j\Omega_k \tau_2(\theta)} \cdots e^{j\Omega_k \tau_N(\theta)}]^T, \tag{3.9}$$

sendo o atraso entre microfones τ_i dependente da direção azimutal θ de chegada dos sinais e $\Omega_k = 2\pi f_s \frac{k}{N_f}$, tal que f_s é a frequência de amostragem e N_f o comprimento da FFT.

Adotando a técnica dos multiplicadores de Lagrange para a solução do problema de minimização com restrições desenvolvido até aqui, definimos a função custo

$$l(\mathbf{W}_k, \boldsymbol{\lambda}) = \frac{1}{2}\mathbf{W}_k^H \mathbf{S}_{\mathbf{X}_k\mathbf{X}_k}\mathbf{W}_k + \boldsymbol{\lambda}^H(\mathbf{C}_k^H\mathbf{W}_k - \mathbf{f}), \tag{3.10}$$

cujo gradiente é dado por

$$\nabla_{\mathbf{W}_k} l(\mathbf{W}_k, \boldsymbol{\lambda}) = \mathbf{S}_{\mathbf{X}_k\mathbf{X}_k}\mathbf{W}_k + \mathbf{C}_k\boldsymbol{\lambda}, \tag{3.11}$$

permitindo obter a solução iterativa para o processo de otimização

$$\mathbf{W}_k(\mathrm{n}+1) = \mathbf{W}_k(\mathrm{n}) - \mu\nabla_{\mathbf{W}_k} l(\mathbf{W}_k, \boldsymbol{\lambda}), \tag{3.12}$$

em que μ é o parâmetro de controle do tempo de convergência do algoritmo.

Combinando adequadamente a restrição imposta na Eq. (3.8) com as Eqs. (3.11) e (3.12), obtemos os multiplicadores de Lagrange dados por

$$\boldsymbol{\lambda} = \frac{1}{\mu}(\mathbf{C}_k^H\mathbf{C}_k)^{-1}\mathbf{C}_k^H\mathbf{W}_k(\mathrm{n}) - (\mathbf{C}_k^H\mathbf{C}_k)^{-1}\mathbf{C}_k^H\mathbf{S}_{\mathbf{X}_k\mathbf{X}_k}\mathbf{W}_k(\mathrm{n}) - \frac{1}{\mu}(\mathbf{C}_k^H\mathbf{C}_k)^{-1}\mathbf{f}, \tag{3.13}$$

permitindo definirmos o algoritmo iterativo de *beamforming* segundo

$$\mathbf{W}_k(\mathrm{n}+1) = \mathbf{P}_k[\mathbf{W}_k(\mathrm{n}) - \mu\mathbf{X}_k(\mathrm{n})\mathbf{Y}_k^*(\mathrm{n})] + \mathbf{W}_{ck}, \tag{3.14}$$

desde que adotemos a aproximação $\tilde{\mathbf{S}}_{\mathbf{X}_k\mathbf{X}_k}(n) = \mathbf{X}_k(n)\mathbf{X}_k^H(n)$ (assume-se que X_k é um processo ergódico) e apliquemos a relação $\mathbf{Y}_k(\mathrm{n}) = \mathbf{W}_k^H(\mathrm{n})\mathbf{X}_k(\mathrm{n})$, sendo $\mathbf{P}_k = \mathbf{I} - \mathbf{C}_k(\mathbf{C}_k^H\mathbf{C}_k)^{-1}\mathbf{C}_k^H$ e $\mathbf{W}_{ck} = \mathbf{C}_k(\mathbf{C}_k^H\mathbf{C}_k)^{-1}\mathbf{f}$.

A estrutura algorítmica desenvolvida até então possui como principal falha o potencial de atenuação do sinal na direção de interesse, visto que a restrição $\mathbf{d}_k(\theta)^H\mathbf{W}_k = 1$ é praticamente inviável, já que os sensores normalmente não são perfeitamente isotrópicos, há erros nos seus posicionamentos e as estimativas das direções desejadas comumente são imperfeitas. Neste sentido, Doblinger propõe a modelagem dessas fontes de erro como ruído branco gaussiano e destaca que sua variância é amplificada por $\mathbf{W}_k^H\mathbf{W}_k$, propondo como solução a limitação de $||\mathbf{W}_k||^2$.

De modo a facilitar a modelagem desta restrição sobre $||\mathbf{W}_k||^2$, o seguinte artifício algébrico é empregado: $\mathbf{W}_k(\mathrm{n}) = \mathbf{V}_k(\mathrm{n}) + \mathbf{W}_{ck}$ (observa-se que $\mathbf{P}_k\mathbf{W}_{ck} = \mathbf{0}$). Segue, então, que o limite superior B_k sobre a norma de $\mathbf{W}_k$ pode ser expresso conforme

$$||\mathbf{W}_k(\mathrm{n})||^2 = ||\mathbf{V}_k(\mathrm{n})||^2 + ||\mathbf{W}_{ck}||^2 \leq B_k. \tag{3.15}$$

Deste modo, a parcela variável ($\mathbf{V}_k(\mathrm{n})$) da norma de $\mathbf{W}_k$ deverá ser limitada por

$$||\mathbf{V}_k(\mathrm{n})|| \leq \sqrt{B_k - ||\mathbf{W}_{ck}||^2} = b_k. \tag{3.16}$$

Assim, a fim de acrescentar esta restrição ao algoritmo desenvolvido até então, os seguintes passos devem ser adotados:

$$\tilde{\mathbf{V}}_k(\mathrm{n}+1) = \mathbf{P}_k \left[\mathbf{V}_k(\mathrm{n}) - \frac{\mu}{||\mathbf{X}_k(\mathrm{n})||^2 + \epsilon} \right] \mathbf{X}_k(\mathrm{n}) \mathbf{Y}_k^*(\mathrm{n}), \tag{3.17}$$

$$\mathbf{V}_k(\mathrm{n}+1) = \begin{cases} \tilde{\mathbf{V}}_k(\mathrm{n}+1), & \text{se } ||\tilde{\mathbf{V}}_k(\mathrm{n}+1)|| \leq b_k \\ \frac{b_k \tilde{\mathbf{V}}_k(\mathrm{n}+1)}{||\tilde{\mathbf{V}}_k(\mathrm{n}+1)||}, & \text{se } ||\tilde{\mathbf{V}}_k(\mathrm{it}+1)|| > b_k \end{cases} \quad \text{e} \tag{3.18}$$

$$\mathbf{W}_k(\mathrm{n}+1) = \mathbf{V}_k(\mathrm{n}+1) + \mathbf{W}_{ck}, \tag{3.19}$$

destacando-se que na Eq. (3.17) foi incluída a normalização do passo, de modo a otimizar as condições de convergência do algoritmo. Quanto ao valor dos parâmetros b_k e μ e à inicialização do algoritmo, seguimos as recomendações de [14]. Define-se, assim, por completo, o algoritmo de Doblinger.

Capítulo 4

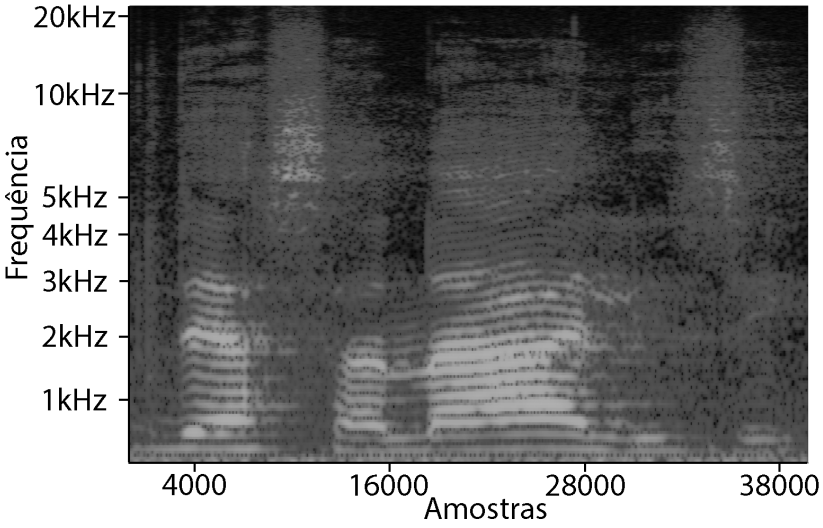
A Síntese Senoidal

Este capítulo descreve o método de síntese de sinais de voz através de osciladores senoidais desenvolvido em [11]. O emprego desta técnica de processamento no contexto de separação cega de fontes parte da observação de que o sinal de voz normalmente tem a sua potência concentrada em um conjunto discreto de frequências, podendo ser modelado por um conjunto de osciladores, ao passo que o ruído de fundo e as componentes correspondentes à reverberação usualmente apresentam maior dispersão e menor potência [13]. Assim, identificando as amplitudes, frequências e fases das trilhas senoidais constituintes do sinal de voz, torna-se possível ressintetizar este sinal, minimizando o ruído e a reverberação.

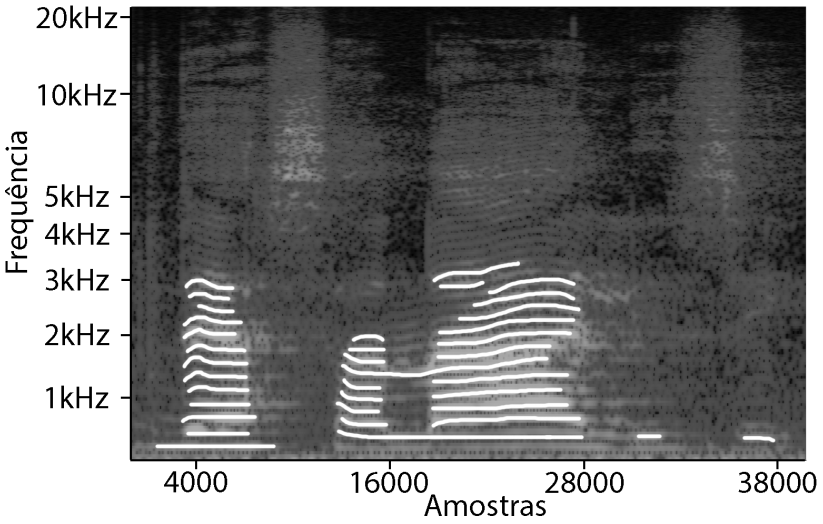
Este processo é realizado no domínio da STFT, cuja análise das magnitudes em função do tempo (espectrograma) revela que a voz é composta por uma série de componentes com picos de energia uniformemente e paralelamente distribuídos, conforme ilustra a Fig. 4.1(a). Neste contexto de síntese, busca-se identificar as trilhas senoidais de maior razão sinal ruído (vide a Fig. 4.1(b)), extrair os seus parâmetros fundamentais (amplitude, frequência e fase) e reconstituir o sinal, conforme ilustra a Fig. 4.1(c).

Aplicando-se este processo (esquematizado na Fig. 4.2) ao sinal de saída de cada *beamformer*, temos disponível um conjunto de sinais menos ruidosos e com menos reverberação. Destarte, entregando estes sinais como entrada para um processo de separação de fontes, facilita-se a execução desta segregação, visto que não será necessário tratar ruídos e reverberações prejudiciais para a obtenção do resultado desejado.

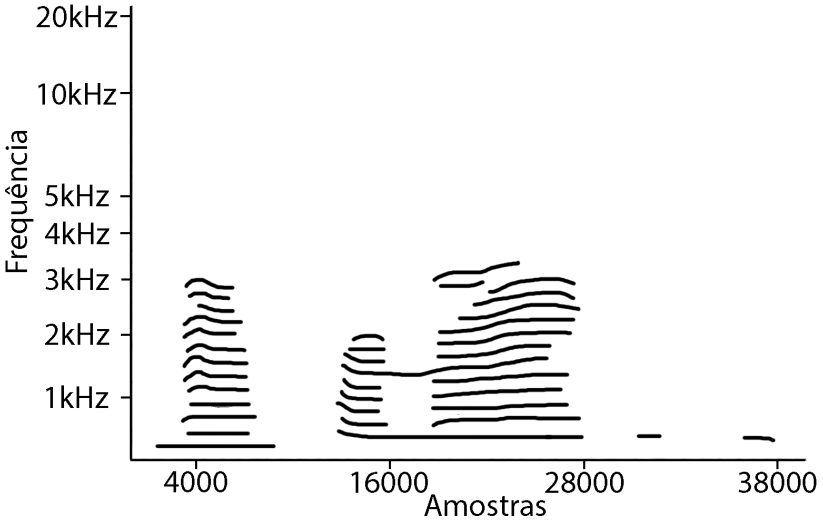
Em seguida detalharemos o sistema de análise e síntese que compõem o método da síntese senoidal. Iniciaremos esta descrição pelas particularidades do sistema de análise, apresentando o modelo senoidal para o sinal de voz. Em seguida, será esclarecido como estimar os parâmetros da fala, como lidar com o casamento de picos espectrais em quadros distintos e, finalmente, como sintetizar o sinal de voz.



(a) Espectrograma original.

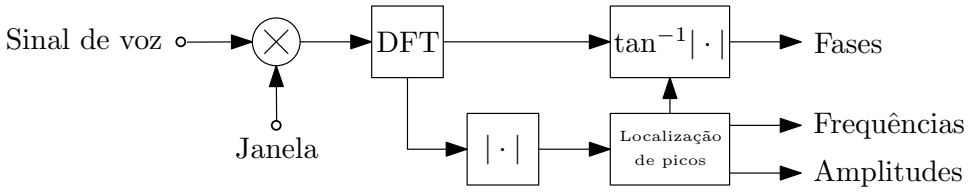


(b) Identificação dos componentes de alta razão sinal ruído.

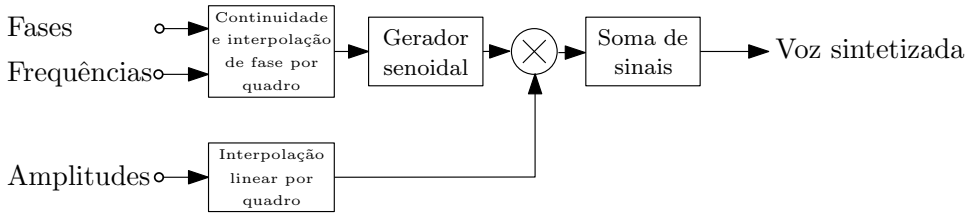


(c) Componentes senoidais utilizados na síntese.

Figura 4.1: Exemplo ilustrativo de reconstrução de sinal de voz amostrado a 44.100 Hz através da síntese senoidal.



(a) Análise



(b) Síntese

Figura 4.2: Fluxograma do processo de análise e síntese senoidal de sinais de voz.

4.1 O Modelo Senoidal

O desenvolvimento do método de síntese senoidal que será apresentado em seguida parte da interpretação do sinal de fala $s(t)$ como a modificação de um sinal de excitação $e(t)$ por um filtro linear e variante no tempo que modela as características do trato vocal. Este filtro é genericamente representado por

$$H(\omega, t) = M(\omega, t)\mathrm{e}^{j\Phi(\omega,t)}, \tag{4.1}$$

sendo M e Φ a magnitude e a fase da função complexa, respectivamente. A excitação em um intervalo de tempo, em contrapartida, é modelada por uma combinação de $L(t)$ senoides de amplitude $a_l(t)$, $w_l(t)$ frequências instantâneas e variáveis no tempo e fases fixas ϕ_l arbitrárias, conforme

$$e(t) = \Re\left\{ \sum_{l=1}^{L(t)} a_l(t)\mathrm{e}^{j\int_0^t \omega_l(\tau)\mathrm{d}\tau + \phi_l} \right\}. \tag{4.2}$$

Assim, se assumirmos que os parâmetros de excitação são invariantes ao longo da duração da resposta ao impulso da Eq. (4.1), depreende-se que

$$s(t) = \sum_{l=1}^{L(t)} A_l(t)\mathrm{e}^{j\Psi_l(t)}, \tag{4.3}$$

sendo

$$A_l(t) = a_l(t)M(\omega_l(t), t) \tag{4.4}$$

a amplitude da l-ésima senoide e

$$\Psi(t) = \int_0^t \omega_l(\tau)\mathrm{d}\tau + \Phi[\omega_l(t), t] + \phi_l \tag{4.5}$$

o seu perfil de frequências.

O processo de extração de frequências, fases e amplitudes de um sinal, a fim de realizar a sua síntese, será detalhado em seguida.

4.2 Estimação dos Parâmetros

Neste seção será demonstrado como extrair os parâmetros senoidais que representam a forma de onda de interesse. Para tal, o sinal será dividido em k trechos quase estacionários (janelas ou quadros de análise) de duração T, cujos centros ocorrem no tempo t_k.

Assumindo que os parâmetros da excitação $e(t)$ e do modelo do trato vocal $H(\omega, t)$ permanecem invariáveis durante um intervalo que inclui a duração da janela de análise e a duração da resposta ao impulso do modelo do trato vocal, então

$$\Psi_l(t) = \omega_l^k(t - t_k) + \theta_l^k, \tag{4.6}$$

já que $\int_0^t \omega_l(\tau)\mathrm{d}\tau = \omega_l^k(t - t_k)$, sendo t_k o tempo correspondente ao centro do intervalo de observação e t um tempo arbitrário dentro do k-ésimo quadro.

Assim, aplicando o resultado da Eq. (4.6) na Eq. (4.3) concluímos que

$$s(t) = \sum_{l=1}^{L(t)} A_l(t)\mathrm{e}^{j\left[\omega_l^k(t-t_k)+\theta_l^k\right]}, \tag{4.7}$$

cuja representação no domínio discretizado é

$$s(n) = \sum_{l=1}^{L^k} A_l^k \mathrm{e}^{j\theta_l^k}\mathrm{e}^{jn\omega_l^k} \Rightarrow s(n) = \sum_{l=1}^{L^k} \gamma_l^k \mathrm{e}^{jn\omega_l^k} \tag{4.8}$$

com $\gamma_l^k = A_l^k \mathrm{e}^{jn\theta_l^k}$ indicando a l-ésima amplitude complexa dentre as L^k senoides e n corresponde a amostras de $t - t_k$ na faixa de $-N/2$ a $N/2$, com $n = 0$ representando o centro da janela de análise e $N + 1$ o seu comprimento em número de amostras.

O objetivo da extração de parâmetros é, portanto, encontrar os valores dos parâmetros que compõem as formas de onda descritas na Eq. (4.8), de modo que a sua combinação represente da melhor forma possível o sinal medido $y(n)$. Assim, a abordagem através da minimização do erro médio quadrático é particularmente interessante, podendo ser formulada segundo

$$\epsilon^k = \sum_n |y(n) - s(n)|^2. \tag{4.9}$$

Aplicando a Eq. (4.8) na Eq. (4.9) é possível demonstrar [11] que a estimativa ótima para as amplitudes e fases é dada por

$$\hat{\gamma}_l^k = Y(l\omega_0^k), \tag{4.10}$$

sendo

$$Y(\omega) = \frac{1}{N+1}\sum_n y(n)\mathrm{e}^{-jn\omega}, \tag{4.11}$$

ou seja, o erro é minimizado através da seleção de todas as trilhas senoidais da informação de fala na banda Ω, sendo $L^k = \Omega/\omega_0^k$.

Este estimador é sempre válido para trechos do sinal que contenham informação vozeada [30]; contudo, para trechos não vozeados, visto que a informação espectral varia muito rapidamente, o tamanho da janela deve ser escolhido criteriosamente de modo que a STFT tenha a resolução em frequência necessária para esta análise. Ademais, a escolha de uma janela adequada, de modo que seus lóbulos laterais no domínio da frequência não degradem a performance do estimador, se faz necessária. Neste contexto, o estudo [11] conclui que, para uma janela retangular, o estimador permanece válido nas regiões não vozeadas se o comprimento N da janela respeitar

$$|\omega_i^k - \omega_l^k| \geq \frac{4\pi}{N+1}, \tag{4.12}$$

o que pode ser visto como uma análise do pior caso, já que outros janelamentos como Hanning, Hamming e Kaiser geram lóbulos secundários menos pronunciados [31].

Assim, o estimador ótimo de parâmetros senoidais encontra as frequências e magnitudes correspondentes aos picos de $|Y(\omega)|$, sendo as amplitudes complexas dadas por

$$\hat{\gamma}_l = \hat{A}_l^k \mathrm{e}^{j\hat{\theta}_l^k} \tag{4.13}$$

4.3 Casamento de Trilhas Espectrais

Na etapa de casamento de parâmetros entre quadros, a meta é associar as amplitudes, frequências e fases computadas em uma janela de dados àquelas da próxima janela. Este processo deve ser suficientemente robusto de modo a obter êxito mesmo na presença de picos espúrios introduzidos pelos lóbulos secundários da janela (no domínio da frequência), quando há alteração na localização dos picos devida à mudança de intonação e em regiões de mudanças espectrais rápidas, como na interface entre janelas de trechos vozeados e não vozeados da fala.

A abordagem proposta por [11] visando a solução do problema supracitado se baseia em um algoritmo heurístico composto pelos três passos descritos a seguir.

Primeiro passo:

Para cada trilha estimada $\hat{\omega}_n^k$ busca-se no quadro vizinho a trilha $\hat{\omega}_m^{k+1}$ tal que

$$|\hat{\omega}_n^k - \hat{\omega}_m^{k+1}| < |\hat{\omega}_n^k - \hat{\omega}_i^{k+1}| < \Delta, \tag{4.14}$$

ou seja, após avaliarmos a distância entre todas as trilhas $\hat{\omega}_i^{k+1}$ no quadro vizinho e $\hat{\omega}_n^k$ no quadro atual, encontramos $\hat{\omega}_m^{k+1}$ como a trilha mais próxima de $\hat{\omega}_n^k$ dentro de uma faixa de tolerância Δ.

Se for encontrado $\hat{\omega}_m^{k+1}$ que atenda ao critério acima, o passo 2 é executado. Caso contrário, assume-se que a trilha $\hat{\omega}_m^{k+1}$ "cessou" na interface entre quadros, e ela é casada com uma réplica de magnitude nula no quadro $k+1$.

Segundo passo:

Cada trilha $\hat{\omega}_m^{k+1}$ encontrada no passo 1 como candidata para ser a continuação de $\hat{\omega}_n^k$ no quadro vizinho ainda não pode ser tida como solução definitiva. Isto ocorre porque pode haver uma trilha na janela k que seja mais próxima de $\hat{\omega}_m^{k+1}$, de modo que estas seriam emparelhadas e $\hat{\omega}_n^k$ permaneceria sem par.

Assim, com o intuito de verificar se as trilhas $\hat{\omega}_m^{k+1}$ apontadas no passo 1 são de fato as candidatas definitivas, o seguinte teste é executado:

$$|\hat{\omega}_m^{k+1} - \hat{\omega}_n^k| < |\hat{\omega}_m^{k+1} - \hat{\omega}_{i+1}^k| \text{ para } i \geq n. \tag{4.15}$$

Caso esta verificação seja positiva, o pareamento de trilhas é confirmado. Caso contrário, como não há candidato $\hat{\omega}_m^{k+1}$ para $\hat{\omega}_n^k$, entende-se que $\hat{\omega}_n^k$ "cessou" na fronteira entre quadros e, portanto, ela é continuada por uma réplica de magnitude nula no quadro $k+1$.

Terceiro passo:

Após todas as trilhas do quadro k terem encontrado outra que garante a sua continuidade no quadro $k+1$ ou terem sido dadas como "cessantes" na interface entre quadros, caso existam trilhas no quadro $k+1$ que não têm correspondente no quadro k, diz-se que elas "surgiram" em $k+1$. Neste caso, esta nova trilha é pareada com uma réplica de magnitude nula no quadro k.

A Fig. 4.3 ilustra todas as situações que foram descritas nos passos acima.

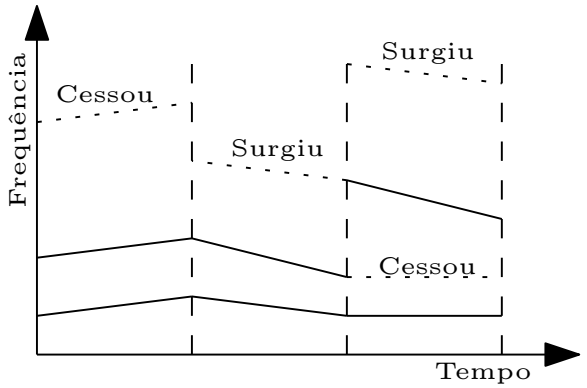


Figura 4.3: Casamento de raias espectrais.

4.4 O Sistema de Síntese

Visto que as fases, frequências e amplitudes das trilhas estimadas a cada quadro são parâmetros variáveis no tempo, é preciso que um mecanismo de interpolação seja desenvolvido de modo que não sejam geradas descontinuidades nas fronteiras entre quadros, o que acarretaria efeitos indesejáveis, tais como ruído musical [32], comprometendo a qualidade do sinal sintetizado. Deste modo, indicando as magnitudes, frequências e fases estimadas para o quadro k por $\hat{A}_l^k$, $\hat{\omega}_l^k$ e $\hat{\theta}_l^k$, respectivamente, e por $\hat{A}_l^{k+1}$, $\hat{\omega}_l^{k+1}$ e $\hat{\theta}_l^{k+1}$ os mesmos parâmetros no quadro $k+1$, uma maneira eficaz de interpolação das magnitudes é dada por:

$$\tilde{A}_l(n) = \hat{A}_l^k + \frac{\hat{A}_l^{k+1} - \hat{A}_l^k}{S} n, \tag{4.16}$$

sendo $n = 0, \cdots, S-1$ o índice das amostras tomadas do k-ésimo quadro.

Por outro lado, esta mesma abordagem mostra-se inadequada para a interpolação das fases, visto que esta informação não é obtida como uma função contínua, mas sim limitada no intervalo $[0, 2\pi)$. Deste modo, os autores do método de síntese sugerem tornar a fase uma função contínua (processo conhecido em inglês como *Unwrapping*) e adotar uma função cúbica de interpolação. Como vantagem desta abordagem, pode-se solucionar concomitantemente a interpolação das frequências (conforme será apresentado em seguida), visto que a frequência instantânea é a derivada da fase contínua.

A referida função cúbica tem a forma

$$\tilde{\theta}_l(t) = \zeta_l + \gamma_l t + \alpha t^2 + \beta t^3 \tag{4.17}$$

e, visto que a frequência instantânea é a derivada desta função, temos

$$\dot{\tilde{\theta}}_l(t) = \gamma_l + 2\alpha t + 3\beta t^2, \tag{4.18}$$

sendo $\zeta_l = \tilde{\theta}(0) = \hat{\theta}_l^k$, $\gamma_l = \dot{\tilde{\theta}}(0) = \hat{\omega}_l^k$, e t a variável tempo contínua tal que $t = 0$ corresponde ao quadro k e $t = T$ corresponde ao quadro $k + 1$.

Quando $t = T$ temos

$$\begin{aligned} \tilde{\theta}_l(t) &= \hat{\theta}_l^k + \hat{\omega}_l^k T + \alpha T^2 + \beta T^3 = \hat{\theta}_l^{k+1} + 2\pi M \\ \dot{\tilde{\theta}}(t) &= \hat{\omega}_l^k + 2\alpha T + 3\beta T^2 = \hat{\omega}_l^{k+1} \end{aligned} \tag{4.19}$$

Uma vez que a fase deve ser contínua (em vez de restrita a $[0, 2\pi)$), um fator inteiro M deve ser determinado de modo que a evolução de fase seja maximamente suave. Em [11] é demonstrado que o valor de M é o inteiro mais próximo de

$$m = \frac{1}{2\pi}\left[(\hat{\theta}_l^k - \hat{\omega}_l^k T - \hat{\theta}_l^{k+1}) + (\hat{\omega}_l^{k+1} - \hat{\omega}_l^k)\frac{T}{2}\right]. \tag{4.20}$$

A função de interpolação finalmente assume a forma

$$\tilde{\theta}_l(t) = \hat{\theta}_l^k + \hat{\omega}_l^k t + \alpha(M)t^2 + \beta(M)t^3, \tag{4.21}$$

sendo

$$\begin{bmatrix} \alpha(M) \\ \beta(M) \end{bmatrix} = \begin{bmatrix} \frac{3}{T^2} & -\frac{1}{T} \\ -\frac{2}{T^3} & \frac{1}{T^2} \end{bmatrix} \begin{bmatrix} \hat{\theta}_l^{k+1} - \hat{\theta}_l^k - \hat{\omega}_l^k T + 2\pi M \\ \hat{\omega}_l^{k+1} - \hat{\omega}_l^k \end{bmatrix} \tag{4.22}$$

derivados da Eq. (4.19). A forma de onda sintética é enfim representada por

$$\tilde{s}(n) = \sum_{l=1}^{L^k} \tilde{A}_l(n)\cos[\tilde{\theta}_l(n)], \tag{4.23}$$

sendo $\tilde{A}_l(n)$ calculado segundo a Eq. (4.16) e $\tilde{\theta}_l(n)$ calculado conforme a Eq. (4.21).

Capítulo 5

Separação Cega de Fontes no Domínio da Frequência

Realizados os processos descritos nos Capítulos 3 e 4, obtém-se uma segregação primária das fontes, que pode ser empregada como inicialização de qualquer método de pós-processamento a fim de maximizar a qualidade da separação. Neste capítulo será apresentada uma dessas técnicas: a exploração de dependências estatísticas de alta ordem entre raias espectrais [15]. No Capítulo 6 será descrito como a integração entre a inicialização proposta e o método de exploração de dependências estatísticas de alta ordem pode ser alcançada através da conversão da separação primária em um conjunto de matrizes de segregação para cada raia espectral por meio da solução de Wiener.

5.1 Exploração de dependências estatísticas de alta ordem entre raias espectrais

O método de separação cega de fontes proposto em [15] é um algoritmo banda larga[1] que explora dependências estatísticas de alta ordem entre raias espectrais de uma mesma estimativa visando a mitigação de permutações de frequências entre fontes distintas. Para tal fim, em cada raia de frequência as misturas são interpretadas como instantâneas, de tal modo que as componentes de frequências da i-ésima fonte são estimadas por

$$\hat{S}_i(k) = \sum_{j=1}^{M} g_{ij}(k) X_j(k),\ i = 1, \cdots, N \tag{5.1}$$

[1] Um método no domínio da frequência é classificado de banda larga se o critério utilizado para separação contempla todas as frequências simultaneamente, isto é, as componentes frequenciais não são tratadas de forma independente [3].

sendo M e N, respectivamente, o número de misturas e de fontes (assume-se que se tem este conhecimento) e $g_{ij}(k)$ o k-ésimo elemento (no domínio da frequência) do filtro de separação que atua sobre a j-ésima mistura $x_j(k)$.

Procuramos, então, pelas matrizes ótimas $\mathbf{G}(k)$, formadas pelos elementos $g_{ij}(k)$, que minimizam, para cada raia espectral, a seguinte função custo derivada da distância de Kullback-Leibler entre a densidade de probabilidade (do inglês *Probability Density Function - PDF*) conjunta dos vetores de estimativas das fontes (contém as K raias espectrais da i-ésima fonte estimada em um dado quadro) e o produto das PDFs individuais estimadas por um modelo estatístico, como métrica de independência:

$$C = d_{KL}\left(p(\hat{\mathbf{S}}_1, \ldots, \hat{\mathbf{S}}_N) \parallel \prod_{i=1}^{N} q(\hat{\mathbf{S}}_i)\right). \tag{5.2}$$

Sendo $d_{KL}(p(x)||q(x)) = \int p(x) \log \frac{p(x)}{q(x)}$, depreende-se que esta função custo avalia o grau de independência entre as fontes, sendo $C = 0$ quando a densidade de probabilidade conjunta das estimativas das fontes iguala-se ao produto das densidades marginais, correspondendo à independência estatística [33]. Prosseguindo com o desenvolvimento de C através de transformações de variáveis [15], obtém-se

$$C = c - \sum_{k=1}^{K} \log \mid \det \mathbf{G}(k) \mid - \sum_{i=1}^{N} \mathrm{E}[\log q(\hat{\mathbf{S}}_i)], \tag{5.3}$$

em que $c = \int p(\mathbf{X}_1, \ldots, \mathbf{X}_M) \log p(\mathbf{X}_1, \ldots, \mathbf{X}_M) d\mathbf{X}_1, \ldots, d\mathbf{X}_M$, sendo $p(\mathbf{X}_1, \ldots, \mathbf{X}_M)$ a PDF conjunta dos vetores de misturas, representa a entropia das misturas. Assim, c é constante, fazendo com que a minimização de C independa de c. Por outro lado, $\sum_{i=1}^{N} \mathrm{E}[\log q(\hat{\mathbf{S}}_i)]$, mede a verossimilhança entre a real distribuição probabilística das fontes estimadas e o modelo tomado como referência. Deste modo, quanto maior é este valor, mais próximo estamos do mínimo de C, já que esta parcela tem peso negativo.

Cabe destacar que, caso estejamos trabalhando com sinais com alta densidade de valores nulos (ou quase nulos), estimativas que convergem para zero maximizariam o termo $\sum_{i=1}^{N} \mathrm{E}[\log q(\hat{\mathbf{S}}_i)]$, visto que $q(\hat{\mathbf{S}}_i)$ pode ser modelada como uma PDF supergaussiana, conforme será apresentado adiante, pois esta é uma propriedade típica da distribuição das amostras de sinais de voz. Estas soluções são devidamente penalizadas pela parcela $\log \det \mathbf{G}(k)$, dado que ela converge para $-\infty$ quando as estimativas tendem para zero, incrementando C. É interessante notar que o mesmo fenômeno também ocorre caso as estimativas convirjam para uma mesma solução, já que, neste caso, $\mathbf{G}$ terá linhas linearmente dependentes.

Deste modo, esta função custo tem a propriedade de minimizar a dependência

estatística entre as estimativas, contornado soluções nulas e estimativas repetidas, adequando-as a um modelo estatístico adotado *a priori*. Adicionalmente, escolhendo-se o processo de atualização via gradiente natural, conforme será apresentado em seguida, consegue-se maximizar a dependência estatística entre raias espectrais de uma mesma estimativa, minimizando o efeito de permutação entre componentes espectrais de estimativas distintas, fenômeno intrínseco à maioria dos métodos de separação cega de fontes no domínio da frequência.

Assim, procedendo à minimização de C, podemos empregar a seguinte equação de atualização dos coeficientes das matrizes de separação:

$$g_{ij}(k, m+1) = g_{ij}(k, m) + \eta \Delta g_{ij}(k), \tag{5.4}$$

em que η é a taxa de aprendizado e $\Delta G_{ij}(k)$ é o gradiente natural dado por [15]:

$$\Delta g_{ij}(k) = \sum_{n=1}^{N} \left(\delta_{in} - \mathrm{E}\left[\varphi^{(k)} \left(\hat{S}_i(1), \ldots, \hat{S}_i(K) \right) \hat{S}_n^*(k) \right] \right) g_{nj}(k), \tag{5.5}$$

com $\delta_{in} = 1$ quando $i = n$ e zero nos demais casos.

A função multivariável $\varphi^{(k)} \left(\hat{S}_i(1), \ldots, \hat{S}_i(K) \right)$ apresentada na Eq. (5.5) é responsável pala maximização da dependência estatística entre raias espectrais de uma mesma estimativa e é dada por [15]:

$$\varphi^{(k)} \left(\hat{S}_i(1), \ldots, \hat{S}_i(K) \right) = -\frac{\partial \log p \left(\hat{S}_i(1), \ldots, \hat{S}_i(K) \right)}{\partial \hat{S}_i(k)}. \tag{5.6}$$

A distribuição *a priori* das fontes proposta pelo método é dada pela função super-gaussiana multivariável (com dependência) [15]

$$q(\hat{\mathbf{S}}_i) = \alpha \mathrm{e}^{-\sqrt{(\hat{\mathbf{S}}_i - \hat{\mu}_i)^H \sum_i^{-1} (\hat{\mathbf{S}}_i - \hat{\mu}_i)}}, \tag{5.7}$$

sendo α um fator de normalização e μ_i e $\sum_i$ respectivamente o vetor média e a matriz covariância do sinal de i-ésima fonte.

Assumindo que as componentes espectrais são descorrelacionadas entre si, a matriz de covariância se torna diagonal e, admitindo média zero dos coeficientes na frequência, podemos reescrever a Eq. (5.7) conforme:

$$q(\hat{\mathbf{S}}_i) = \alpha \mathrm{e}^{-\sqrt{\sum_K |\frac{S_i(k)}{\sigma_i(k)}|^2}}, \tag{5.8}$$

sendo $\sigma_i(k)$ a variância da i-ésima fonte na k-ésima raia. Como este fator é desconhecido, o método o assume como unitário, resolvendo o escalamento pelo método da distorção mínima, apresentado em seguida. Assim, podemos reescrever $\varphi^{(k)} \left(\hat{S}_i(1), \ldots, \hat{S}_i(K) \right)$ como

$$\varphi^{(k)}\left(\hat{S}_i(1),\ldots,\hat{S}_i(K)\right) = \frac{\hat{S}_i(k)}{\sqrt{\sum_{l=1}^{K} \mid \hat{S}_i(l) \mid}}, \tag{5.9}$$

que é a função multivariável de fato empregada pelo método de exploração de dependências estatísticas de alta ordem.

5.2 Princípio da Distorção Mínima

Visando a redução da má equalização das fontes causada pela aplicação de constantes de escalamento muito distintas entre frequências, característica típica de processos de separação cega de fontes no domínio da frequência, emprega-se o princípio da distorção mínima, apresentado em [34]. Este princípio consiste na substituição das matrizes de coeficientes do filtro de separação $\mathbf{G}(k)$ para cada raia por uma versão escalada obtida pela transformação:

$$\mathbf{G}(k) \leftarrow \mathrm{diag}(\mathbf{G}^{-1}(k))\mathbf{G}(k), \tag{5.10}$$

sendo $\mathrm{diag}(\cdot)$ o operador que retorna a matriz dada como argumento com todos os elementos fora da diagonal principal zerados.

Este artifício pode ser justificado da seguinte maneira: supondo que a separação das fontes obtida pelas matrizes $\mathbf{G}(k)$ seja suficientemente acurada e que não há permutações, então

$$\mathbf{G}(k) \approx \mathbf{D}(k)\mathbf{H}^{-1}(k), \tag{5.11}$$

sendo $\mathbf{D}(k)$ a matriz diagonal de coeficientes de escalamento de uma determinada componente frequencial e $\mathbf{H}(k)$ a matriz de mistura desta mesma componente. Assim, $\mathrm{diag}(\mathbf{G}^{-1}(k))\mathbf{G}(k) \approx \mathrm{diag}(\mathbf{H}(k))\mathbf{H}^{-1}(k)$, ou seja, a transformação da Eq. (5.10) permite um escalamento não arbitrário, já que os elementos das diagonais principais das matrizes $\mathbf{H}(k)$ são oriundos de um processo físico coerente.

Assim, temos por completo a formalização do método de separação cega de fontes através de exploração de dependências estatísticas de alta ordem entre raias espectrais. Destaca-se como vantagem deste algoritmo o fato de não só minimizarmos o efeito da permutação, mas também do escalamento arbitrário dos componentes espectrais.

Capítulo 6

Integração Através da Solução de Wiener

Conforme apresentado na Eq. (5.4), o método de exploração de dependências estatísticas de alta ordem requer um conjunto de matrizes de coeficientes de separação no domínio da frequência de modo a proceder à separação de fontes. Neste capítulo será descrito como a integração entre a inicialização proposta e o método de exploração de dependências estatísticas de alta ordem pode ser alcançada através da conversão da separação primária das fontes em um conjunto de matrizes de segregação para cada raia espectral por meio da solução de Wiener.

O método de inicialização proposto é composto pela etapa de *beamforming*, visando atenuar ruídos e caminhos de reverberação em direções não relevantes, além da fonte interferente, seguida da síntese senoidal, visando extrair somente frequências de interesse e minimizar distorções não lineares.

6.1 A Solução de Wiener Clássica

A solução de Wiener [35] busca, conforme ilustra a Fig. 6.1, o filtro linear de resposta ao impulso finita[1], composto por coeficientes w_k, que processe um sinal $x(n)$ de modo a tornar sua saída $y(n)$ tão próxima de um sinal desejado $d(n)$ que o erro $e(n)$ se torne o menor possível sob o ponto de vista de um critério estatístico específico. Assim, tem-se

$$y(n) = \sum_{k=0}^{N-1} w_k^* x(n-k), \qquad n = 0, 1, 2, ... \tag{6.1}$$

sendo N o comprimento da resposta ao impulso do filtro, permitindo definir

[1]No contexto de aplicação desta dissertação a apresentação da filtragem de Wiener será restrita ao caso de resposta ao impulso finita.

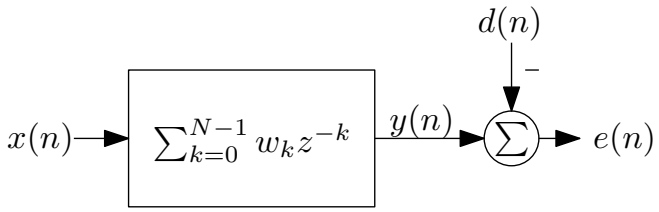


Figura 6.1: Diagrama de bloco da filtragem de Wiener.

$$e(n) = d(n) - y(n). \tag{6.2}$$

Tipicamente adota-se a minimização do erro médio quadrático como critério de otimização do filtro, isto é,

$$J = \mathrm{E}[|e(n)|^2]. \tag{6.3}$$

Neste âmbito, é possível demonstrar [36] que o filtro ótimo é aquele que explora plenamente a correlação estatística entre $x(n)$ e $d(n)$, de modo que o erro remanescente seja ortogonal ao sinal $x(n)$, ou seja,

$$\mathrm{E}[x(n-k)e_o^*(n)] = 0, \qquad k = 0, 1, ..., N-1 \tag{6.4}$$

Combinando as Eqs. (6.1), (6.2) e (6.4), o filtro que atende a este critério é tal que

$$\sum_{i=0}^{N-1} w_{oi} r(i-k) = p(-k), \tag{6.5}$$

sendo w_{oi} o i-ésimo coeficiente da resposta ao impulso do filtro, $r(i-k) = \mathrm{E}[x(n-k)x^*(n-i)]$ a função de autocorrelação do sinal de entrada do filtro para um atraso $i-k$ e $p(-k) = \mathrm{E}[x(n-k)d^*(n)]$ a correlação cruzada entre a entrada do filtro $x(n)$ e o sinal desejado $d(n)$ para um atraso k.

Sob forma matricial, este resultado é expresso segundo

$$\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{p}. \tag{6.6}$$

6.2 Aplicação da Solução de Wiener em Separação Cega de Fontes

Podemos expandir a derivação da seção anterior ao contexto de separação cega de fontes no domínio da frequência, uma vez que dispomos das misturas originais $\mathbf{X}_j$ e das estimativas das fontes separadas $\tilde{\mathbf{S}}_i$, obtidas pela combinação dos métodos apresentados nos Capítulos 3 e 4, no domínio tempo-frequência (STFT). Assim,

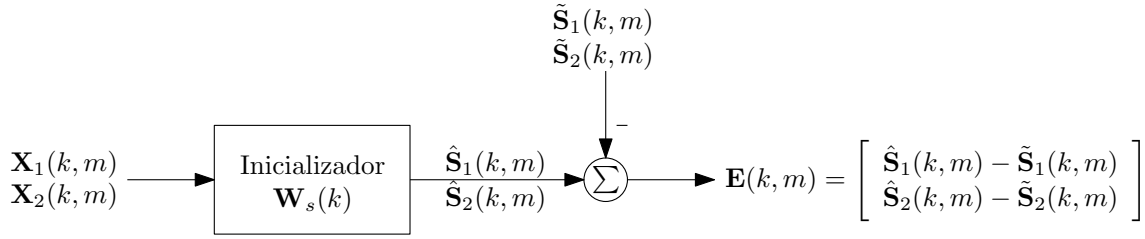


Figura 6.2: Identificação do sistema de inicialização da separação de fontes.

através da técnica do filtro ótimo de Wiener, para cada componente frequencial em cada quadro individual, obtemos as matrizes de separação de cada raia que compõe as fontes, conforme apresentado em [1, 2]. Visto por outra perspectiva, procuramos identificar o sistema linear que atua sobre cada componente espectral das misturas e segrega as raias espectrais das fontes, uma a uma e quadro a quadro.

A Fig. 6.2 apresenta de maneira esquemática o processo de identificação num cenário com duas fontes e duas misturas. Nesse contexto, o bloco "Inicializador" representa o sistema que desejamos identificar e que realiza no domínio da frequência, para cada raia espectral k em cada quadro m, a separação dos componentes frequenciais das fontes. A Eq. (6.7) apresenta o processamento necessário para se estimar o sistema "Inicializador" de modo que a energia $\mathrm{E}[|E_i(k,m)|^2]$ de cada componente $E_i(k,m)$ do vetor $\mathbf{E}(k,m)$ seja mínima.

$$\begin{aligned} \mathbf{P}_i(k) &= \mathrm{E}\left\{ \begin{bmatrix} \mathbf{X}_1(k,m) \\ \mathbf{X}_2(k,m) \end{bmatrix} \tilde{\mathbf{S}}_i^*(k,m) \right\}, \\ \mathbf{R}(k) &= \mathrm{E}\left\{ \begin{bmatrix} \mathbf{X}_1(k,m) \\ \mathbf{X}_2(k,m) \end{bmatrix} \begin{bmatrix} \mathbf{X}_1^*(k,m) & \mathbf{X}_2^*(k,m) \end{bmatrix} \right\}, \\ \mathbf{W}_i(k) &= \mathbf{R}(k)^{-1}\mathbf{P}_i(k), \\ \mathbf{W_s}(k) &= \begin{bmatrix} \mathbf{W}_1(k) & \mathbf{W}_2(k) \end{bmatrix}, \end{aligned} \tag{6.7}$$

Na Eq. (6.7), $\mathrm{E}\{\cdot\}$ representa o valor esperado estatístico[2], $\mathbf{X}_i(k,m)$ e $\tilde{\mathbf{S}}_i(k,m)$ são, respectivamente, as k-ésimas raias das STFTs do sinal do i-ésimo sensor e da i-ésima estimativa no m-ésimo quadro, e $\mathbf{W_s}(k)$ é a matriz de separação 2×2 que combina $\mathbf{W}_1(k)$ e $\mathbf{W}_2(k)$ em uma única matriz, tal que ambas as fontes possam ser estimadas por

$$\begin{bmatrix} \hat{\mathbf{S}}_1(k,m) \\ \hat{\mathbf{S}}_2(k,m) \end{bmatrix} = \mathbf{W}_\mathbf{s}^H(k) \begin{bmatrix} \mathbf{X}_1(k,m) \\ \mathbf{X}_2(k,m) \end{bmatrix}, \tag{6.8}$$

Destaca-se que no presente contexto esta maneira de calcular a solução de Wiener, se aplicada sem restrições, tende a gerar uma solução polarizada, visto que

[2]Na prática, empregam-se médias temporais (em m), assumindo-se que as misturas e as fontes são processos ergódigos.

são levados em consideração todos os dados disponíveis, inclusive aqueles que representam janelas das misturas em que não há informação relevante, como trechos de silêncio ou contendo apenas ruído. Objetivando contornar este problema, sugere-se como artifício para otimizar o método proposto por [1, 2], a eliminação de todos os elementos em um quadro cujas energias sejam menores do que um limiar multiplicativo aplicado à energia média daquela frequência ao longo de todos os quadros usados na computação da solução da Eq. (6.7). Em suma, assume-se que a energia média de uma componente de frequência ao longo dos quadros será fortemente balizada pelo ruído de fundo e momentos de silêncio e que, quando a energia desta componente em um dado quadro suplantar este limiar, haverá informação pertinente a ser tratada (mistura de vozes). A sistematização matemática deste método é dada abaixo.

Seja a energia na k-ésima raia de frequência:

$$\varepsilon_{i,k} = \frac{1}{B} \sum_{m=1}^{B} |\mathbf{X}_i(k,m)|^2, \tag{6.9}$$

em que B é o número total de quadros. Assim, $\mathbf{X}_i(k,m)$ só será levado em conta no processo de estimação da solução de Wiener se $|\mathbf{X}_i(k,m)|^2 \geq \varepsilon_{i,k}$ para um dado k.

Capítulo 7

Avaliação de Desempenho

A avaliação de desempenho dos processos de separação cega de fontes pode se dar por métricas objetivas ou subjetivas. Quando tratamos de quantificações de desempenho por vias objetivas, empregamos métodos numéricos capazes de qualificar um algoritmo através de um parâmetro matemático. Por outro lado, ao tratarmos de avaliações subjetivas, referimo-nos a reproduzir o resultado dos experimentos para um público diversificado de sorte que cada avaliador possa quantificar, dentro de uma escala numérica pré-determinada, o quanto aquele resultado o agradou. Estes métodos de distinção de resultados serão apresentados em seguida.

7.1 Métricas Objetivas

As quatro métricas objetivas de desempenho de separação cega de fontes mais comuns são as propostas em [37]: a razão sinal interferência (do inglês, *Signal to Interference Ratio* - SIR), a razão sinal distorção (do inglês, *Signal to Distortion Ratio* - SDR), a razão sinal artefato (do inglês, *Signal to Artifact Ratio* - SAR) e a razão sinal ruído (do inglês, *Signal to Noise Ratio* - SNR). O propósito destes avaliadores é quantificar, respectivamente, a interferência entre fontes nos sinais estimados (o grau de sucesso da separação), a distorção imposta pelo mecanismo de separação de fontes aos sinais obtidos, como distorções não lineares, a introdução de artefatos pelo algoritmo, como ruído musical, e o quão relevante é o ruído total introduzido ao longo de todo o processo de separação das fontes. Nesta dissertação consideramos somente a SIR, visto que trata-se do método mais empregado na literatura. Contudo, como as suas deduções são semelhantes, as quatro métricas serão introduzidas.

O conceito fundamental para o cálculo destes avaliadores de desempenho é a interpretação das j estimativas de fontes como a combinação do sinal desejado, de interferências, de ruídos e de artefatos:

$$\hat{\mathbf{s}}_j = \mathbf{s}_{\text{esp}} + \mathbf{s}_{\text{int}} + \mathbf{s}_{\text{rud}} + \mathbf{s}_{\text{art}}, \tag{7.1}$$

em que $\mathbf{s}_{\text{esp}}$ é o sinal esperado (separação ideal), $\mathbf{s}_{\text{int}}$ é a parcela de interferência, $\mathbf{s}_{\text{rud}}$ é a parcela de ruído, e $\mathbf{s}_{\text{art}}$ é a parcela de artefatos.

Assim, define-se:

$$\text{SIR} = 10\log_{10}\frac{||\mathbf{s}_{\text{esp}}||^2}{||\mathbf{s}_{\text{int}}||^2}; \tag{7.2}$$

$$\text{SAR} = 10\log_{10}\frac{||\mathbf{s}_{\text{esp}} + \mathbf{s}_{\text{int}} + \mathbf{s}_{\text{rud}}||^2}{||\mathbf{s}_{\text{art}}||^2}; \tag{7.3}$$

$$\text{SDR} = 10\log_{10}\frac{||\mathbf{s}_{\text{esp}}||^2}{||\mathbf{s}_{\text{int}} + \mathbf{s}_{\text{rud}} + \mathbf{s}_{\text{art}}||^2} \quad \text{e} \tag{7.4}$$

$$\text{SNR} = 10\log_{10}\frac{||\mathbf{s}_{\text{esp}} + \mathbf{s}_{\text{int}}||^2}{||\mathbf{s}_{\text{rud}}||^2}. \tag{7.5}$$

7.2 Métrica Subjetiva

A avaliação subjetiva de processos de separação de fontes consiste em reproduzir para um público diversificado os resultados gerados pelo método proposto e solicitar que cada avaliador submeta seu parecer acerca da qualidade da separação através de uma nota em uma escala determinada. No escopo desta dissertação solicitou-se ao público voluntário que fosse avaliado em uma escala de 1 a 5 o quanto as interferências remanescentes após o processo de separação o perturbava em comparação às separações ideais (neste caso, 1 indica "interferência extremamente prejudicial" e 5 indica "interferência imperceptível"). O Apêndice A inclui uma folha de apresentação do teste subjetivo entregue ao público avaliador e explica detalhadamente a condução do procedimento.

7.3 Delimitações dos Testes

Os testes realizados neste estudo estão limitados à separação cega de fontes no contexto determinado com duas fontes e dois sensores. Os ambientes de testes e as características dos sinais e *hardware* empregados são apresentados em seguida.

7.3.1 Os Ambientes de Testes

Foram adotados três ambientes de testes, descritos abaixo.

Laboratório de Processamento Analógico e Digital de Sinais

A Fig. 7.1 apresenta a configuração de testes que foi montada no Laboratório de Processamento Analógico e Digital de Sinais (PADS) da Universidade Federal do Rio de Janeiro. Neste ambiente, cujo tempo de reverberação estimado é de $T_{60} = 700\,\text{ms}$, foram instalados dois microfones idênticos, distanciados de 5 cm, e posicionados na base de uma mesa circular de $1,2\,\text{m}$ de raio. Na outra extremidade foram posicionados dois alto-falantes reproduzindo monólogos individuais de 19 s de duração gravados em ambiente controlado (estúdio com pouca reverberação), gerando as misturas necessárias aos estudos apresentados em seguida. A posição do primeiro alto-falante foi definida em 60º ou 90º (em relação ao centro dos arranjo de microfones) e a posição do segundo alto-falante foi escolhida em 105º ou 120º. As vozes reproduzidas foram sempre de pessoas de sexos opostos, e as gravações foram realizadas tanto quando o laboratório estava vazio (doravante referido como "PADS silencioso") quanto quando o laboratório estava ocupado por alunos e professores trabalhando, gerando ruído de fundo (doravante referido como "PADS ruidoso").

Ambiente acústico simulado

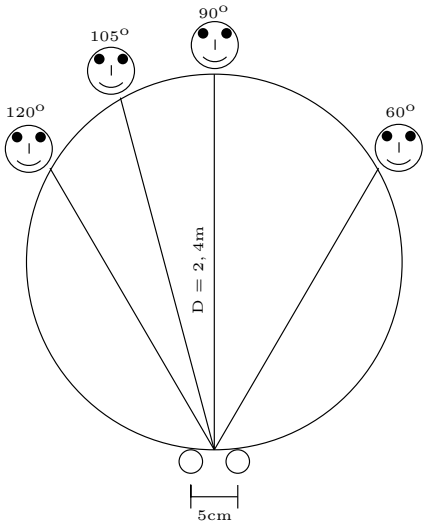
Foram empregados dois ambientes acústicos simulados pelo método Image-Source Model [38], cujos tempos de reverberação são de $T_{60} = 700\,\text{ms}$ e $T_{60} = 1\,\text{s}$. Foram posicionadas duas fontes de vozes masculina e feminina em 45º e 135º e a $1,2\,\text{m}$ em relação ao centro do arranjo de microfones, composto por dois elementos distanciados de 5 cm, gerando misturas de 10 s de duração.

Sala D-105

A Fig. 7.2 apresenta a configuração de testes que foi montada na sala D-105 do Centro de Tecnologia da Universidade Federal do Rio de Janeiro (CT UFRJ) durante o desenvolvimento do trabalho [39]. Neste ambiente, cujo tempo de reverberação estimado é de $T_{60} = 932\text{ms}$, foram posicionados alto-falantes reproduzindo monólogos individuais de 8 s de duração gravados em ambiente controlado, gerando as misturas necessárias aos estudos apresentados em seguida. O primeiro alto-falante foi posicionado em 84º, 85º, 109º ou 115º (em relação ao centro do arranjo de microfones) e o segundo em 109º, 115º, 128º ou 136º. As vozes reproduzidas foram sempre de pessoas de sexos opostos.

A Tabela 7.1 resume todos os casos de teste cujos resultados serão apresentados em seguida.

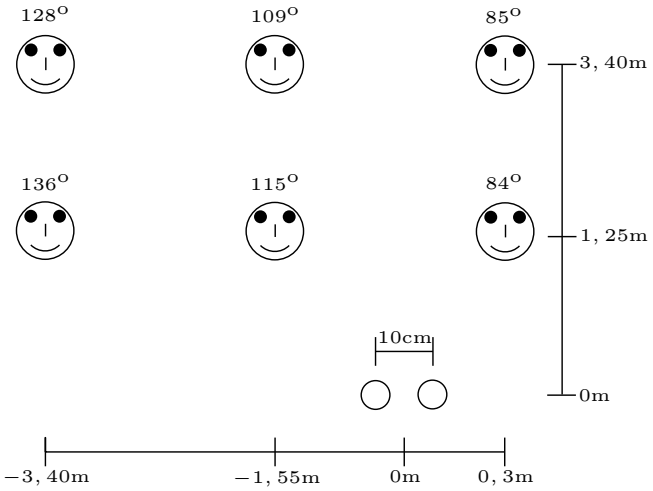
(a) Vista do PADS.



(b) Configuração de testes no PADS.

Figura 7.1: Vista e configuração de testes no PADS.

(a) Vista da sala D-105.



(b) Configuração de testes na sala D-105.

Figura 7.2: Vista e configuração de testes na sala D-105.

Tabela 7.1: Cenários de teste.

Cenário	Ambiente	Ângulo das fontes	Distância entre microfones
Cenário 1	PADS ruidoso	105° e 90°	5 cm
Cenário 2	PADS ruidoso	120° e 60°	5 cm
Cenário 3	PADS silencioso	105° e 60°	5 cm
Cenário 4	Simulação $T_{60} = 700\,\text{ms}$	135° e 45°	5 cm
Cenário 5	Simulação $T_{60} = 1\,\text{s}$	135° e 45°	5 cm
Cenário 6	Sala D-105	109° e 128°	10 cm
Cenário 7	Sala D-105	84° e 136°	10 cm
Cenário 8	Sala D-105	84° e 115°	10 cm
Cenário 9	Sala D-105	115° e 136°	10 cm

7.3.2 Características dos Sinais, *Hardware* e Processamento

Os sinais gravados no PADS e na sala D-105 foram captados por microfones omnidirecionais Behrigner ECM8000 alimentados via *phanto power* por uma mesa de som Behringer Eurorack MX 3242X, cujas saídas *direct-out* foram conectadas às entradas analógicas de uma placa de captura de áudio M-AUDIO Firewire 1814. Esta captura foi conduzida por um computador Intel DG31PR, Intel Core2Duo 2.8GHz e 2GB DDR2 667 na resolução de 16 bits e taxa de amostragem de 44.100 Hz a fim de preservar o máximo de informação do sinal de voz. Entretanto, para o escopo deste trabalho, todos os sinais foram reamostrados a 8.000 Hz. Os sinais que foram convoluídos com as respostas acústicas dos ambientes simulados foram gravados na resolução de 16 bits e taxa de amostragem de 44.100 Hz em ambiente anecoico. A execução do método proposto foi conduzida em um laptop MacBook Pro (Intel Core2Duo 3.06 GHz e 8GB 1067 MHz DDR). Quanto às avaliações subjetivas, os voluntários usaram fones de ouvido Roland RH-5 para escutar os resultados e julgar o processo de separação cega de fontes.

Seguindo as recomendações de [40], no método de *beamforming* foram empregadas janelas de dados de 1024 amostras, com sobreposição de 768 amostras. O método de síntese senoidal, por outro lado, foi implementado conforme as recomendações de [12], com janelas de 256 amostras e sobreposição de 128 amostras. Quanto ao método de exploração de dependências estatísticas de alta ordem, foram adotadas janelas de 2048 amostras, sobreposição de 1536 amostras, 5000 iterações e taxa de aprendizado $\eta = 0,05$. Este tamanho de janela foi determinado através de testes variando este parâmetro até que se encontrasse o valor ótimo, enquanto o valor de η foi escolhido em função dos resultados apresentados em [1, 2].

7.4 Resultados dos Métodos de Separação Cega

Primeiramente investigou-se qual dentre os métodos de *beamforming* descritos no Capítulo 3 alcançaria o melhor resultado - medido através da SIR - quando aplicado no algoritmo proposto na Fig. 1.2. A Tabela 7.2 apresenta este resultado para cada cenário descrito na Tabela 7.1, sendo que cada célula da tabela apresenta o formato "$\text{SIR}_{\text{fonte estimada 1}}$ / $\text{SIR}_{\text{fonte estimada 2}}$".

Pode-se depreender que o método de Doblinger e o algoritmo de Frost apresentam resultados semelhantes na maioria dos cenários. No entanto, a modelagem dos erros associados ao processo de *beamforming* como um ruído branco gaussiano conferiu ao primeiro método melhores resultados nos Cenários 1, 5 e 9. As técnicas de atrasos e somas e LCMV apresentaram SIRs inferiores aos demais métodos em todos os cenários. Portanto, o algoritmo de Doblinger foi escolhido como passo integrante do

Tabela 7.2: Comparação da SIR resultante entre os algoritmos de *beamforming*.

Método / cenário	Cenário 1	Cenário 2	Cenário 3	Cenário 4	Cenário 5
Doblinger	10,6 / 7,7	18,3 / 9,9	14,7 / 17,5	12,4 / 6,4	7,2 / 5,6
Frost	0,7 / 0,7	18,3 / 9,9	14,7 / 17,5	12,2 / 6,4	3,3 / 3,2
LCMV	3,7 / 7,3	3,2 / -1,9	14,6 / 17,5	11,1 / -9,7	-2,1 / 2,1
Atrasos e somas	5,6 / 12,2	15,1 / 6,1	14,8 / 16,5	12,8 / 6,3	3,1 / 3,0

Método / cenário	Cenário 6	Cenário 7	Cenário 8	Cenário 9
Doblinger	2,8 / 6,7	-2,4 / 13,1	-1,9 / 14,1	1,7 / 8,2
Frost	2,5 / 7,0	-2,4 / 13,1	-1,8 / 14,6	1,5 / 8,1
LCMV	3,1 / 7,0	-2,6 / 13,0	-2,1 / 13,9	2,1 / 8,3
Atrasos e somas	-0,9 / 3,8	-2,7 / 13,0	-1,7 / 14,6	-0,6 / 6,5

Tabela 7.3: Comparação da SIR resultante entre os métodos de inicialização.

Método / cenário	Cenário 1	Cenário 2	Cenário 3	Cenário 4	Cenário 5
Proposto	10,6 / 7,7	18,3 / 9,9	14,7 / 17,5	12,4 / 6,4	7,2 / 5,6
[1, 2]	9,3 / 7,2	15,9 / 9,7	11,6 / 14,7	12,0 / 6,0	6,6 / 5,4
Branqueamento	13,2 / 4,4	15,2 / 6,21	14,6 / 17,5	12,6 / 6,2	7,3 / 5,9

Método / cenário	Cenário 6	Cenário 7	Cenário 8	Cenário 9
Proposto	2,8 / 6,7	-2,4 / 13,1	-1,9 / 14,1	1,7 / 8,2
[1, 2]	0,6 / 4,8	-2,9 / 12,4	-2,5 / 13,4	1,7 / 7,6
Branqueamento	-1,0 / 5,4	-2,4 / 13,1	-1,8 / 14,6	2,0 / 8,3

processo de inicialização proposto.

Como segundo passo, definido que o algoritmo de *beamforming* empregado no método de inicialização proposto nesta dissertação seria o algoritmo de Doblinger, comparações do seu desempenho com o do método proposto em [1, 2] e com o do método de branqueamento foram conduzidas. A Tabela 7.3 apresenta o resultado, medido pela SIR, destas comparações. A análise dos resultados (Tabela 7.3) obtidos nos Cenários 1, 2 e 3 permite identificar que o método proposto apresentou melhor resultado para o mesmo ambiente quando este apresenta ruído de fundo moderado, típico de um ambiente acústico real sem fontes ruidosas adicionais (Cenário 3). Todavia, quando consideramos um caso ideal (ambiente acústico simulado do Cenário 4), observa-se um prejuízo em relação à SIR das separações. Este fenômeno ocorre porque o passo de síntese senoidal beneficia-se do descarte de componentes frequenciais do ruído de fundo em seu processamento. Entretanto, quando este ruído é nulo ou sua energia equipara-se a dos sinais de interesse, componentes que outrora seriam favoráveis ao processo de exploração de dependências estatísticas de alta ordem acabam sendo equivocadamente desprezadas. Nota-se também que quanto maior é a distância entre fontes, melhores são os resultados obtidos, o que é razoável de se esperar dos métodos de separação cega de fontes, conforme corroboram os resultados

Tabela 7.4: Comparação da SIR resultante entre os métodos de inicialização com fontes normalizadas.

Método / cenário	Cenário 6	Cenário 7	Cenário 8	Cenário 9
Proposto	4,8 / 5,8	2,9 / 7,5	3,8 / 10,6	2,8 / 5,9
[1, 2]	1,3 / 2,6	2,8 / 7,1	3,7 / 9,8	3,0 / 5,8
Branqueamento	5,1 / 6,0	2,9 / 7,6	3,8 / 10,7	3,2 / 6,1

obtidos pela inicialização [1, 2] e pelo branqueamento.

Observa-se, adicionalmente, para os cenários supracitados, que a SIR obtida pelo método proposto foi, de maneira geral, mais elevada do que as dos métodos concorrentes quando se considera a SIR média entre as fontes separadas.

Em relação ao Cenário 5, percebe-se que o método apresentado neste trabalho apresentou resultados superiores aos obtidos pela aplicação da inicialização proposta em [1, 2], porém semelhantes aos obtidos pelo branqueamento.

Quanto aos demais cenários, destaca-se que o método proposto mostrou-se vantajoso no Cenário 6 e superior à inicialização de [1, 2] nos demais cenários, embora não tenha obtido vantagem em relação ao método de branqueamento nestes casos. É interessante notar que em todos os testes realizados a SIR correspondente da voz masculina sempre se sobressaiu, o que pode ser explicado pela maior energia desta, associada à sua maior proximidade aos microfones, facilitando a sua extração. A fim de investigar a importância da equalização energética entre as fontes, uma terceira série de testes, cujos resultados são apresentados na Tabela 7.4, foi conduzida considerando ambas as fontes normalizadas nos microfones.

A Tabela 7.4 revela que quando as fontes são normalizadas nos microfones os resultados de todos os métodos apresentam melhorias. Este artifício, entretanto, não pode ser utilizado na prática, visto que em situações cotidianas de emprego de um sistema de separação cega de fontes não se tem acesso às excitações individuais de cada fonte nos microfones, somente às misturas. Contudo, este recurso permite o desenvolvimento de uma análise mais criteriosa do método proposto, revelando que o passo de síntese senoidal das misturas possivelmente é prejudicado pela desigualdade de energia das fontes. Quando uma fonte é mais proeminente do que a outra, o processo de síntese não consegue encontrar com facilidade os picos espectrais da fonte de menor energia, prejudicando a composição das matrizes iniciais de separação para esta voz.

Posto que o método de inicialização proposto busca inicializar o algoritmo apresentado no Capítulo 5 em um ponto da sua função custo mais próximo possível do mínimo global, investigou-se, através de novos experimentos, o quão distante do mínimo global alcançável pelo procedimento de exploração de dependências estatísticas de alta ordem estão os resultados apresentados na Tabela 7.3.

Tabela 7.5: Comparação do parâmetro ξ da Eq. (7.6) entre os métodos de inicialização.

Método / cenário	Cenário 1	Cenário 2	Cenário 5	Cenário 6	Cenário 8
Proposto	0,2	0,0	0,0	0,1	0,4
[1, 2]	1,0	1,3	0,4	2,2	0,3
Branqueamento	0,4	3,4	0,2	2,7	0,7

Para tal, empregou-se a estratégia apresentada na Fig. 6.2, adotando-se como sinais de referência ($\tilde{\mathbf{S}}_1$ e $\tilde{\mathbf{S}}_2$) para o método de Wiener as fontes individuais que foram aquisitadas pelos microfones a fim de gerar as misturas. Assim, o algoritmo de separação cega de fontes foi inicializado pelo melhor conjunto de matrizes lineares, correspondente ao mínimo erro quadrático médio entre as fontes ideais e suas estimativas.

O resultado deste estudo foi mensurado pelo seguinte parâmetro:

$$\xi = \frac{\mathrm{SIR}_{\text{ideal 1}} + \mathrm{SIR}_{\text{ideal 2}}}{2} - \frac{\mathrm{SIR}_{\text{estimativa 1}} + \mathrm{SIR}_{\text{estimativa 2}}}{2}, \tag{7.6}$$

sendo $\mathrm{SIR}_{\text{estimativa n}}$ a SIR obtida para a n-ésima estimativa ao se inicializar o processamento de separação de fontes pelo método proposto, pelo método [1, 2] ou pelo branqueamento, e $\mathrm{SIR}_{\text{ideal n}}$ a SIR obtida para a estimativa da n-ésima fonte ao se inicializar o processamento de separação de fontes com a inicialização ideal conforme descrito acima.

A Tabela 7.5 apresenta os valores de ξ, obtidos pela Eq. (7.6), para os Cenários 1, 2, 5, 6 e 8, e cada método confrontado. Salienta-se que, para os cenários selecionados, o método proposto foi aquele que mais se aproximou da inicialização ideal. Portanto, podemos concluir que as baixas SIRs se devem às limitações intrínsecas do método descrito no Capítulo 5, e não ao método de inicialização.

Uma maneira intuitiva de interpretar estes resultados é entendermos que quando o método de separação cega de fontes é inicializado com as fontes idealmente separadas, temos como ponto de partida uma solução que não pertence aos resultados alcançáveis pela função custo a ser processada. Assim, conforme as iterações do processo de separação são executadas, esta solução inicial é modificada até alcançar aquela mais próxima do ponto de partida que pertence à função custo do método de separação de fontes empregado.

No que concerne aos testes subjetivos, os mesmos cinco cenários (1, 2, 5, 6 e 8) foram selecionados para avaliação. Os resultados médios das apreciações de dez voluntários são apresentados na Tabela 7.6.

Nota-se que estes resultados apresentam algum grau de correlação com os resultados da Tabela 7.3. De fato, resultados com SIRs maiores tendem a apresentar avaliações subjetivas mais próximas do resultado máximo, conforme evidencia a

Tabela 7.6: Resultado das avaliações dos testes subjetivos - média e desvio padrão das notas (de 1 a 5) atribuídas às duas estimativas.

Método / cenário	Cenário 1	Cenário 2	Cenário 5
Proposto	3,8 ± 0,5 / 3,8 ± 1,0	4,5 ± 0,4 / 4,6 ± 0,4	3,3 ± 0,5 / 2,6 ± 0,7
[1, 2]	3,8 ± 0,5 / 3,8 ± 0,9	4,6 ± 0,5 / 4,7 ± 0,4	3,3 ± 0,6 / 3,0 ± 0,8
Branqueamento	4,1 ± 0,7 / 3,9 ± 0,5	4,5 ± 0,5 / 4,2 ± 0,6	3,5 ± 0,8 / 2,8 ± 0,8

Método / cenário	Cenário 6	Cenário 8
Proposto	1,9 ± 0,6 / 3,7 ± 0,4	2,5 ± 0,7 / 4,4 ± 0,5
[1, 2]	1,7 ± 0,7 / 3,7 ± 0,6	2,8 ± 0,8 / 4,4 ± 0,4
Branqueamento	1,6 ± 0,7 / 2,6 ± 0,6	2,7 ± 0,8 / 4,3 ± 0,5

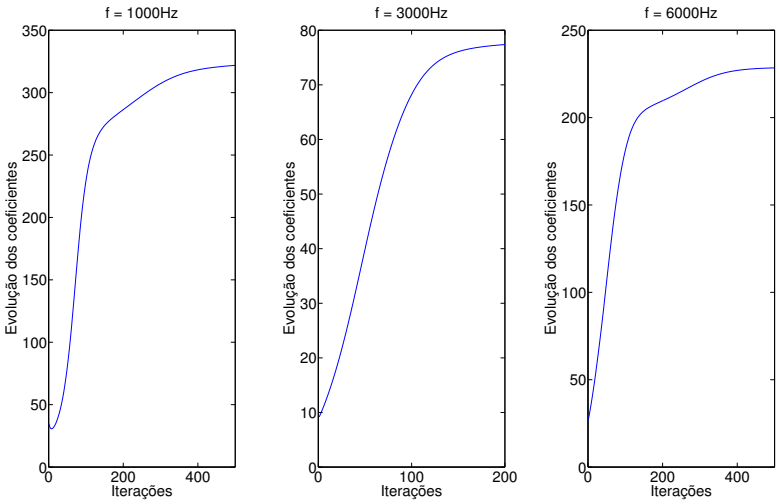
comparação dos Cenários 1 e 2 com os Cenários 6 e 8. Coerentemente, SIRs menores geraram médias subjetivas menores, conforme comprovam os Cenários 6 e 8. Todavia, a sensibilidade do público voluntário mostrou-se menor do que a do cálculo da SIR, o que se torna evidente quando verificamos que mesmo em cenários nos quais a SIR tornou-se negativa a média do público avaliador não foi inferior a 1,6.

Cabe frisar que os resultados subjetivos médios mostram-se semelhantes quando comparados no mesmo cenário, mesmo com a adoção de inicializações diferentes. Este resultado parece reforçar a suposição de que se saturou a capacidade do método de exploração de dependências estatísticas de alta ordem. Destaca-se, contudo, que no Cenário 6 o método de branqueamento apresentou resultado 1,1 unidades (em relação à segunda estimativa) inferior quando comparado aos métodos rivais, ou seja, o branqueamento foi a única técnica a ser depreciada pelos avaliadores.

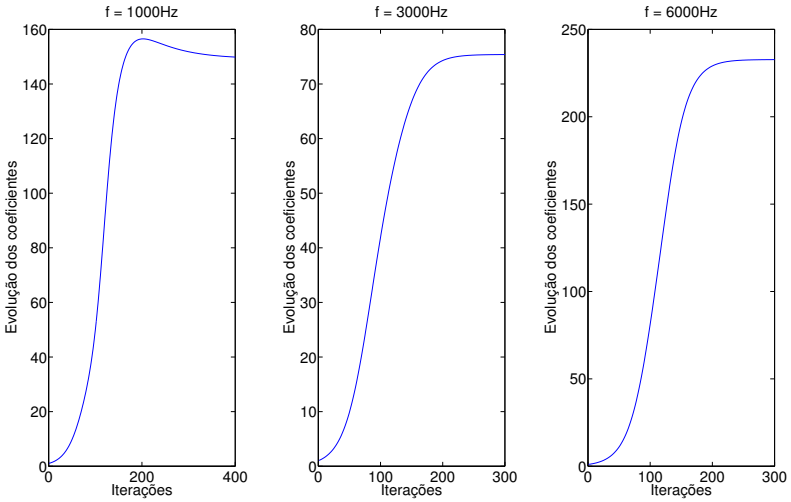
Além dos resultados médios, é pertinente observarmos o desvio padrão dos resultados apresentados. Sob esta perspectiva, nota-se que, excetuando-se o Cenário 1, o método proposto nesta dissertação foi aquele que apresentou avaliações mais consistentes, ao passo que a inicialização por branqueamento apresentou maior dispersão de resultados.

O tempo de convergência do método principal também é um parâmetro importante a ser avaliado, pois ele é diretamente afetado pelo método de inicialização aplicado. As Figs. 7.3 a 7.5 apresentam os gráfico da evolução da soma dos módulos dos coeficientes das matrizes de separação para as frequências $f = 1000\,\mathrm{Hz}$, $f = 3000\,\mathrm{Hz}$ e $f = 6000\,\mathrm{Hz}$ ao longo da execução do método de exploração de dependências estatísticas de alta ordem durante o processamento dos Cenários 2, 6 e 9. Para cada cenário investigou-se a convergência quando o método foi iniciado via branqueamento, através da proposta de [1, 2] e pela proposta desta dissertação. Para a composição de cada figura limitou-se o número de iterações do método núcleo em 8000.

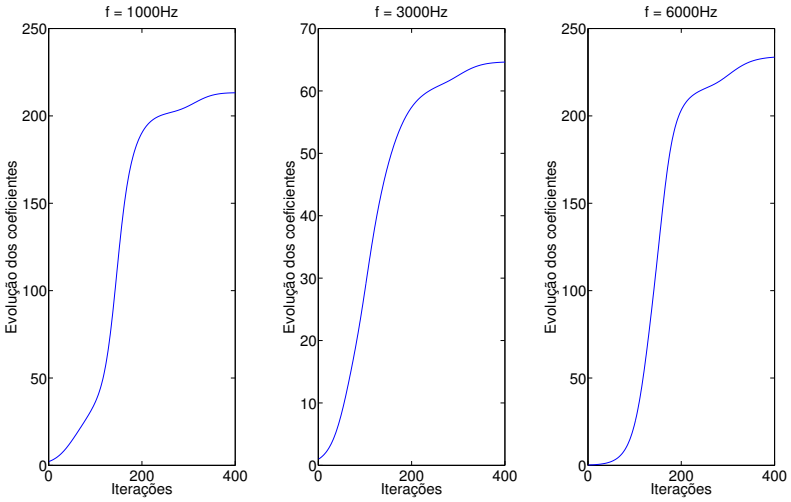
Verificamos que o tempo de convergência quando adota-se o branqueamento é tão menor quanto menor é o ruído de fundo e a reverberação nos sinais a serem



(a) Inicialização por branqueamento.

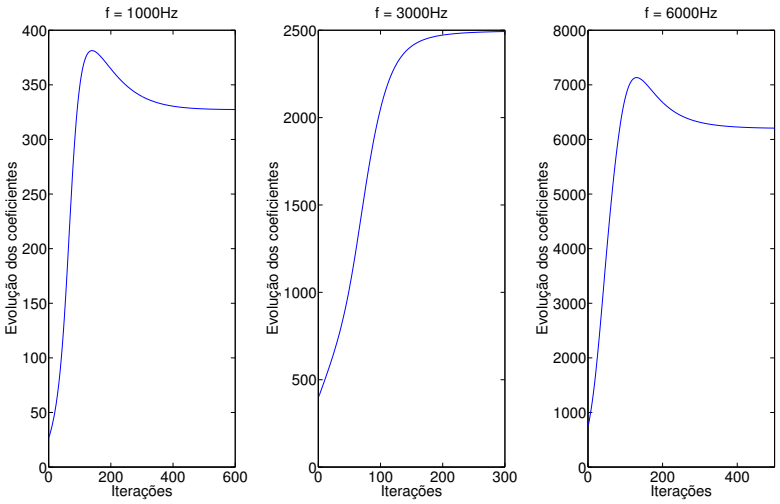


(b) Inicialização conforme [1, 2].

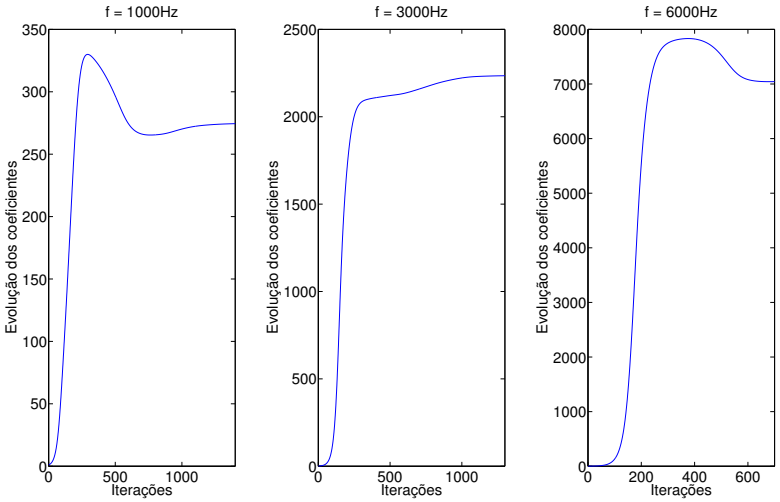


(c) Inicialização proposta.

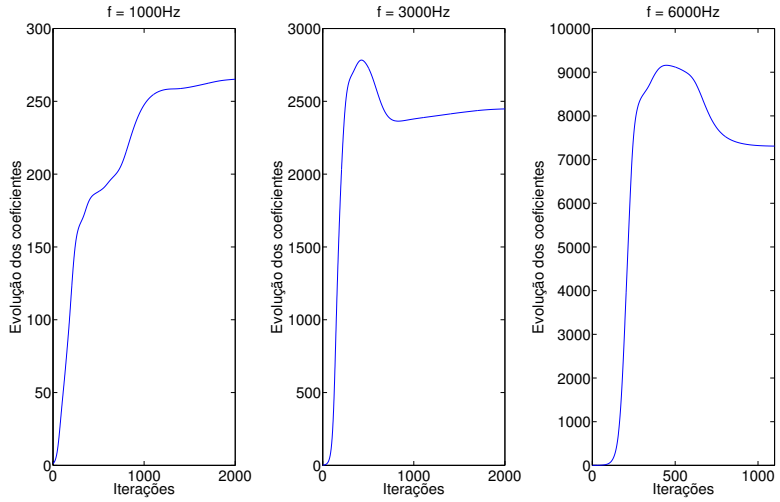
Figura 7.3: Convergências dos algoritmos no Cenário 2.



(a) Inicialização por branqueamento.

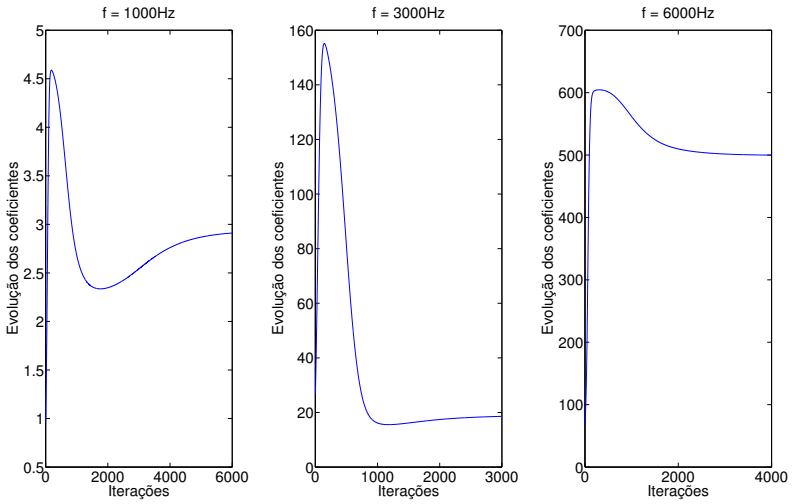


(b) Inicialização conforme [1, 2].

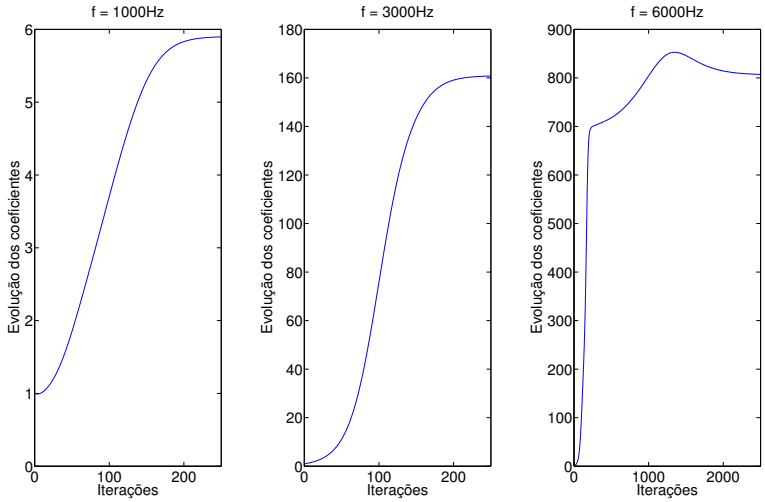


(c) Inicialização proposta.

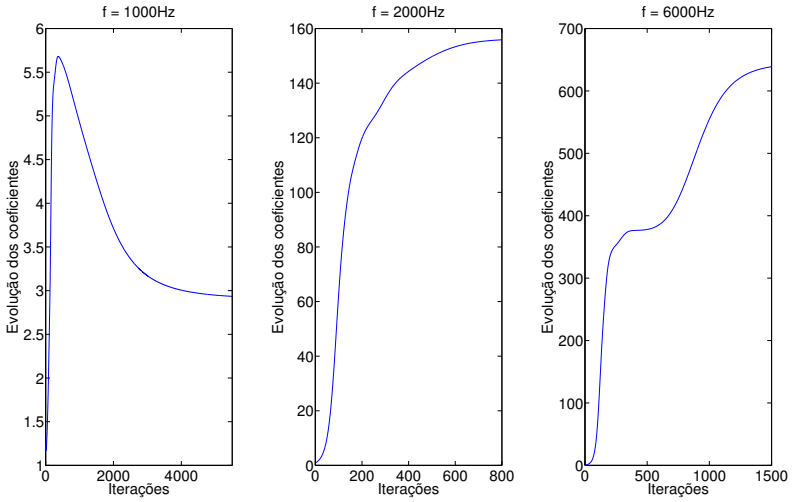
Figura 7.4: Convergências dos algoritmos no Cenário 4.



(a) Inicialização por branqueamento.



(b) Inicialização conforme [1, 2].



(c) Inicialização proposta.

Figura 7.5: Convergências dos algoritmos no Cenário 9.

Tabela 7.7: Variação da SIR em função do erro de determinação do ângulo das fontes.

Método / cenário	Cenário 1	Cenário 2
0% de erro	10,1 / 8,6	18,3 / 9,9
5% de erro	10,4 / 7,9	18,3 / 9,9
10% de erro	10,8 / 5,7	18,3 / 9,9
15% de erro	10,5 / 4,5	18,3 / 9,9
20% de erro	0,3 / 3,1	18,3 / 9,9
25% de erro	5,4 / 1,6	18,3 / 9,9
30% de erro	1,0 / 6,3	18,3 / 9,9

separados. Deste modo, no Cenário 4 o branqueamento mostrou-se o método de inicialização mais vantajoso. Por outro lado, nos Cenários 2 e 9, nos quais o ruído de fundo e a reverberação são maiores, o método proposto e o de [1, 2] mostraram-se vantajosos. Ademais, observa-se que a convergência dos coeficientes de separação de baixas frequências é, de modo geral, mais demorada, excetuando-se a da Fig. 7.5(b), em que a convergência das altas frequências foi cerca de dez vezes mais lenta que a convergência das baixas e médias frequências. Neste contexto, a inicialização proposta mostrou-se notavelmente vantajosa nas altas frequências quando comparada aos demais métodos.

É pertinente destacar que o método de inicialização, conforme apresentado nesta dissertação, não pode ser entendido *stricto sensu* como um método de separação cega de fontes, visto que o emprego das técnicas de *beamforming* pressupõe conhecimento do ângulo das fontes em relação ao centro do arranjo de microfones. Frisa-se, contudo, que métodos como o GCC-PHAT (do inglês *Generalized Cross Correlation with Phase Transform*) e similares [24–27] permitiriam a determinação cega destes ângulos. Este tipo de técnica, contudo, não foi utilizada, pois julgou-se mais interessante investigar a performance máxima que seria obtida pela inicialização proposta quando ela fosse alimentada pela informação acurada dos ângulos. Como vantagem adicional, a abordagem aplicada permite analisar a magnitude da degradação da SIR advinda de erros nas estimativas dos ângulos. Os resultados deste último estudo são apresentados na Tabela 7.7, na qual foram selecionados dois cenários: Cenário 1, em que as fontes foram alocadas angularmente próximas uma da outra (diferença de 15^o), e o Cenário 2 em que as fontes foram posicionadas com maior distância (diferença de 60^o), sendo que desvios percentuais iguais foram aplicados às direções angulares das fontes.

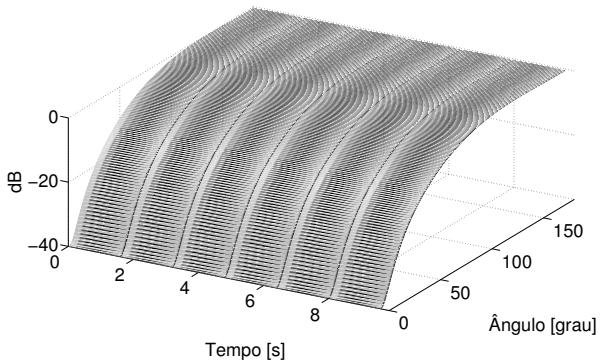
Percebe-se que quando as fontes estão próximas, a SIR resultante permanece razoavelmente inalterada com até 10% de erro na determinação dos ângulos. Por outro lado, quando as fontes estão distanciadas, a SIR não se altera até mesmo para um erro de 30% na determinação dos ângulos. Este resultado certamente está condi-

cionado à seletividade do método de *beamforming* - quanto mais seletivo é o método, maior é a sensibilidade aos erros na determinação dos ângulos de chegada. Todavia, para um arranjo com dois microfones, conforme revela a Eq. (3.2), o número de restrições impostas pela matriz $\mathbf{C}_k$ não deve exceder dois, de modo a não gerar um sistema superdeterminado. Assim, não é possível alcançar uma configuração que seja muito seletiva em toda a faixa de frequências, o que é convenientemente benéfico para a redução da sensibilidade do método proposto à acurácia da determinação da direção angular das fontes. A Fig. 7.6 mostra o padrão de seletividade típico obtido pelo *beamforming* de Doblinger nas frequências de 2 kHz e 3 kHz, exemplificado pelo processamento do Cenário 5 em dois microfones, revelando a seletividade desigual ao longo das frequências.

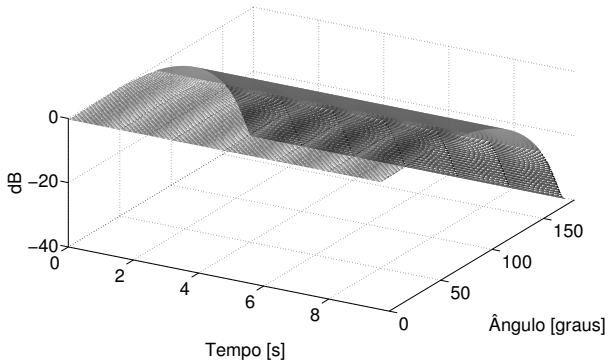
Além dos estudos apresentados até aqui, é igualmente importante que se avaliem as modificações espectrais impostas por cada passo do processo de separação cega de fontes proposto. A Fig. 7.7 mostra os espectrogramas das fontes misturadas em cada canal, dos sinais obtidos após a separação primária pelo *beamforming*, dos sinais após a síntese senoidal e das fontes separadas após a exploração das dependências estatísticas de alta ordem para o Cenário 5. Como referência, apresentam-se também os espectrogramas obtidos pelo processo ideal de separação de fontes.

A análise da Fig. 7.7 permite verificar que, coerentemente ao que se mostra na Fig. 7.6, o processo de *beamforming* é mais seletivo nas altas frequências. Em seguida, do espectro modificado por este processamento, a síntese senoidal reduz a participação de componentes com baixa razão sinal ruído, sobretudo na faixa de altas frequências, região na qual os sinais de voz têm pouca energia. Após a exploração das dependências estatísticas de alta ordem, obtemos o espectrograma do sinal final do processo, que pode ser comparado ao da separação ideal. Percebe-se, principalmente em relação à Separação 1, a semelhança em relação ao espectro ideal. Quanto à Separação 2, após a exploração das dependências estatísticas de alta ordem, alguns componentes de alta frequência de baixa relevância foram resgatados. O resultado apresentado na Fig. 7.7 ilustra o que tipicamente ocorre após o processamento pelos métodos de *beamforming* e síntese senoidal, isto é, a minimização das componentes de pouca relevância no espectro de voz, sobretudo na porção superior do espectro. Esta atenuação, conforme indicado na Eq. (6.9), é trazida para o processo de exploração de dependências estatísticas de alta ordem através da solução de Wiener. Notamos, contudo, que mesmo com esta restrição sobre a solução de Wiener, parte do espectro de altas frequências é resgatado ao fim do processo, embora com menos energia do que na composição original das misturas.

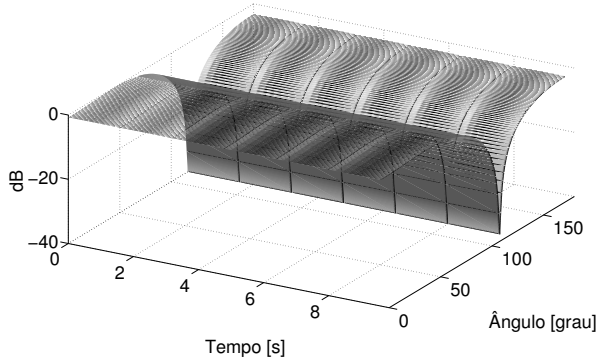
Finalmente, a Fig. 7.8 apresenta as formas de onda dos sinais obtidos ao longo das etapas do processamento proposto e os sinais originais (separação ideal) para o Cenário 5.



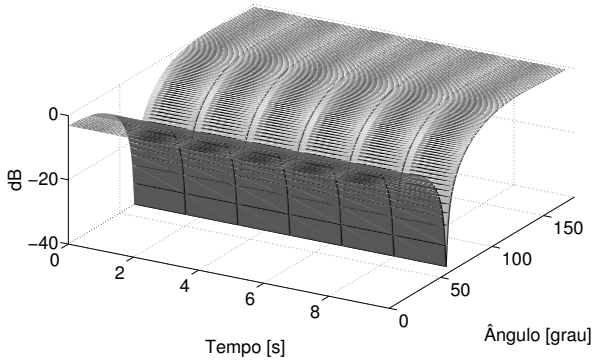
(a) Seletividade em $f = 2000\,\text{Hz}$, fonte 1



(b) Seletividade em $f = 2000\,\text{Hz}$, fonte 2



(c) Seletividade em $f = 3000\,\text{Hz}$, fonte 1



(d) Seletividade em $f = 3000\,\text{Hz}$, fonte 2

Figura 7.6: Seletividade do método de *beamforming*.

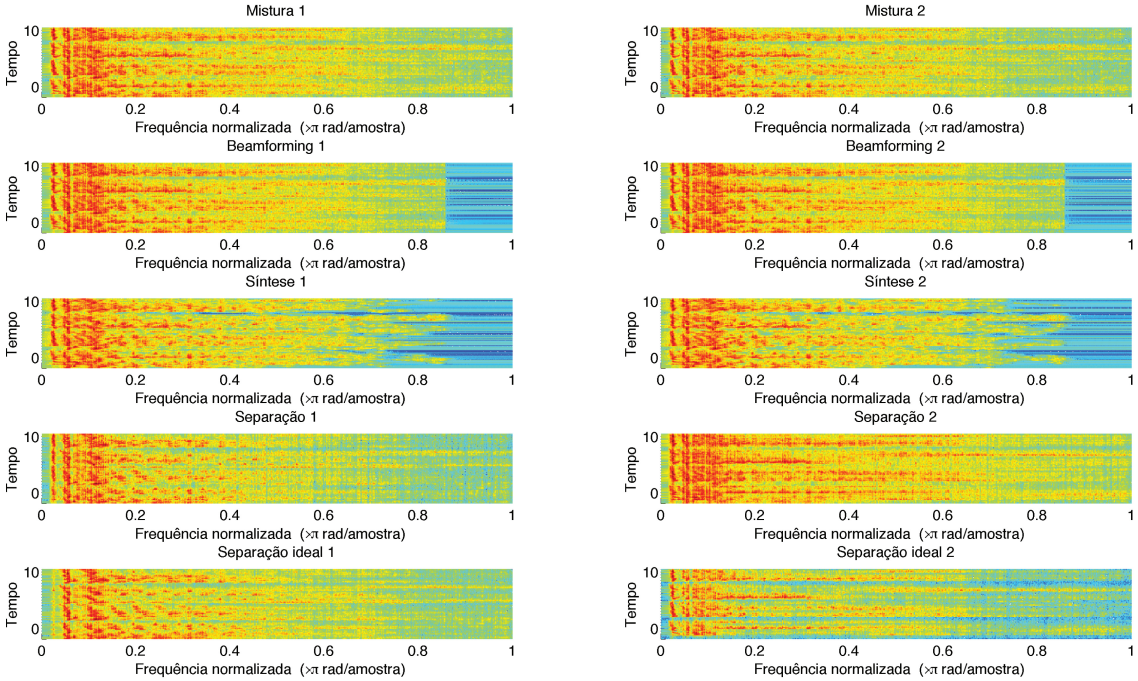


Figura 7.7: Espectrogramas ao longo do processo.

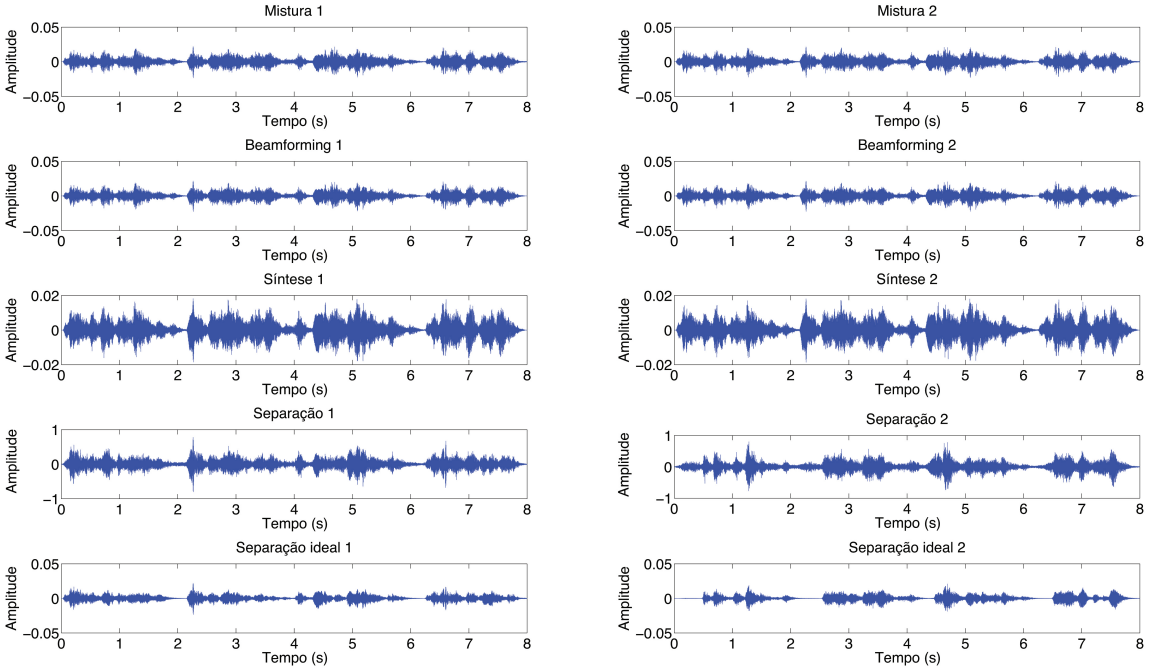


Figura 7.8: Formas de onda ao longo do processo.

Capítulo 8

Conclusão e Trabalhos Futuros

Esta dissertação apresentou uma nova abordagem para a separação cega de fontes em contexto reverberante. Através do emprego da técnica de *beamforming* proposta por Doblinger [14] e do método de síntese senoidal introduzido por [11], aplicados ao paradigma desenvolvido em [1, 2], procurou-se alcançar resultados superiores aos obtidos pela técnica de branqueamento e aos alcançados em [1, 2] na inicialização dos coeficientes do sistema de separação.

As análises realizadas no Capítulo 7, principalmente dos resultados da Tabela 7.3, permitem identificar o método proposto como superior à proposta de [1, 2] sob o ponto de vista da SIR, embora a vantagem sobre o método de branqueamento não tenha sido tão pronunciada, havendo até mesmo cenários em que o branqueamento se sobressaiu. Por outro lado, conforme revela a Tabela 7.5, o método de inicialização proposto, quando se considera a SIR média das estimativas, é o que mais se aproxima da inicialização ideal, de onde se infere que resultados superiores não foram alcançados não por consequência da inicialização, mas pelas limitações do método núcleo de separação, baseado na exploração de dependências estatísticas de alta ordem.

Ao considerarmos o valor médio das análises subjetivas, revela-se que todos os métodos de inicialização apresentaram resultados semelhantes. Assim, ressalva-se que todos os métodos mostraram-se, na média, adequados para a aplicação em cenários reais com usuários humanos. Por outro lado, o emprego das técnicas apresentadas em contextos de inteligência computacional, como transcrição automática de diálogos, identificação de comandos de voz e reconhecimento de assinatura vocal, de acordo com a métrica SIR, ainda exigirá aperfeiçoamentos.

É importante enfatizar que, conforme apontado no Capítulo 7, a sensibilidade do método subjetivo de avaliação mostrou-se bem menor do que a do método objetivo. Deste modo, fica claro que, embora exista correlação entre as duas métricas, a SIR não pode ser vista como o método definitivo para qualificar sistemas cujos usuários serão humanos. Por outro lado, para sistemas automatizados, a SIR pode

representar um quantitativo de facilidade de tratamento dos dados via inteligência computacional.

Tão importante quanto os resultados objetivos e subjetivos, é o fato do método proposto, conforme indicam as Figs. 7.3 a 7.5, permitir a convergência rápida do método central de separação de fontes. Ainda que a convergência do método apresentado nesta dissertação não tenha sido superior em todos os contextos, ficou claro que em situações ruidosas e muito reverberantes a vantagem sobre o branqueamento é significativa.

Assim, embora esta dissertação tenha progredido sobre o tema de separação cega de fontes reverberantes, uma solução definitiva ainda não foi encontrada, cabendo aos trabalhos futuros análises complementares até que se alcance o completo entendimento desta temática. Logo, visto que esta é uma área de pesquisa de grande relevância, propõem-se em seguida algumas abordagens para pesquisas futuras derivadas daquilo que se apresentou nesta dissertação e que podem trazer melhorias aos resultados alcançados:

1. Visto que a Tabela 7.5 indica que a limitação fundamental para o alcance de melhores SIRs encontra-se no método núcleo de separação de fontes, sugere-se que a mesma estratégia de inicialização ideal empregada para o levantamento desta tabela seja executada visando a comparação de diversos métodos de separação cega de fontes no domínio da frequência. Após a identificação do método mais robusto, este poderia ser inicializado conforme a proposta desta dissertação;

2. Além da sugestão supramencionada, em função dos resultados da Tabela 7.4, indica-se como linha de pesquisa o aperfeiçoamento de métodos de separação cega de fontes no domínio da frequência de modo que estes se tornem menos sensíveis às diferenças de energia entre as fontes misturadas;

3. Não só a diferença de energia entre fontes é prejudicial aos métodos de separação cega de fontes, mas também a diferença de absorção acústica de ambientes reais faz com que a covariância entre as frequências de uma mesma fonte não seja facilmente reconhecida. Assim, aponta-se a abordagem em subbandas proposta em [41] como possibilidade de melhoria da técnica adotada nesta dissertação;

4. Conforme apontado no Capítulo 7, o método de *beamforming* proposto tem sua seletividade limitada em função da estrutura da Eq. (3.2), já que, a fim de evitar um sistema superdeterminado, as restrições impostas em $\mathbf{C}_k$ não devem ultrapassar o número de microfones. Isto posto, recomenda-se o estudo dos

resultados alcançáveis por arranjos com mais de dois microfones e também por arranjos não uniformes e irregulares;

5. Apontou-se que o método de inicialização proposto poderia ser complementado por uma etapa de pré-processamento que determinasse automaticamente e de maneira cega os ângulos das fontes em relação ao centro do arranjo de microfones. Aconselha-se, pois, o estudo comparativo de algoritmos de detecção de ângulos de chegada e o emprego do método mais promissor;

6. A análise das Figs. 7.3 a 7.5 permite identificar que diferentes métodos de inicialização alcançam convergências mais rápidas em determinadas faixas de frequência. Deste modo, o estudo de técnicas híbridas de inicialização, de modo a acelerar a convergência ao longo de todo o espectro, é um viés de estudo que pode ser profícuo principalmente para aplicações em tempo real.

Referências Bibliográficas

[1] CLARK, F., PETRAGLIA, M., HADDAD, D. "A New Initialization Method for Frequency-Domain Blind Source Separation Algorithms", *Signal Processing Letters, IEEE*, v. 18, n. 6, pp. 343–346, Jun. 2011.

[2] CLARK, F., PETRAGLIA, M., HADDAD, D. *Cancelamento de Eco Acústico e Separação Cega de Fontes Aplicados à Telefonia Viva-Voz*. Projeto de graduação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Dez. 2010.

[3] HADDAD, D. B., PETRAGLIA, M. R., BATALHEIRO, P. B. "Métodos de Separação Cega de Fontes". In: *Tutoriais do XVII Congresso Brasileiro de Automática*, v. 1, Book Editora, pp. 133–157, Juiz de Fora, Set. 2008.

[4] ALLEN, J. B., BERKLEY, D. A. "Image Method for Efficiently Simulating Small-Room Acoustics", *Acoustical Society of America*, v. 65, n. 4, pp. 943–950, Abr. 1979.

[5] KIM, L.-H., HASEGAWA-JOHNSON, M. "Toward Overcoming Fundamental Limitation in Frequency-Domain Blind Source Separation for Reverberant Speech Mixtures". In: *2010 Conference Record of the Forty Fourth Asilomar Conference*, pp. 542–545, Nov. 2010.

[6] ARAKI, S., MUKAI, R., MAKINO, S., et al. "The Fundamental Limitation of Frequency Domain Blind Source Separation for Convolutive Mixtures of Speech", *Speech and Audio Processing, IEEE Transactions on*, v. 11, n. 2, pp. 109–116, Abr. 2003.

[7] WANG, D. "On Ideal Binary Mask As the Computational Goal of Auditory Scene Analysis". In: Divenyi, P. (Ed.), *Speech Separation by Humans and Machines*, pp. 181–197. Springer US, Mar. 2005.

[8] MESGARANI, N., CHANG, E. F. "Selective Cortical Representation of Attended Speaker in Multi-Talker Speech Perception", *Nature*, v. 485, n. 7397, pp. 233 – 236, Mai. 2012.

[9] GELFAND, S. *Hearing: An Introduction to Psychological and Physiological Acoustics*. 5 ed. Sheepen Place, Informa Healthcare, Dez. 2009.

[10] BOURGEOIS, J., MINKER, W. *Time-Domain Beamforming and Blind Source Separation: Speech Input in the Car Environment*, v. 3, *Lecture Notes in Electrical Engineering*. 1 ed. New York, Springer US, Abr. 2009.

[11] MCAULAY, R., QUATIERI, T. "Speech Analysis/Synthesis Based on a Sinusoidal Representation", *Acoustics, Speech and Signal Processing, IEEE Transactions on*, v. 34, n. 4, pp. 744–754, Ago. 1986.

[12] ELLIS, D. P. W. "Sinewave and Sinusoid+Noise Analysis/Synthesis in Matlab". Acesso em Out. 2013. Disponível em: <`http://www.ee.columbia.edu/~dpwe/resources/matlab/sinemodel/`>.

[13] ZHANG, W., RAO, B. D. "A Two Microphone-Based Approach for Source Localization of Multiple Speech Sources", *Audio, Speech, and Language Processing, IEEE Transactions on*, v. 18, n. 8, pp. 1913–1928, Nov. 2010.

[14] DOBLINGER, G. "Localization and Tracking of Acoustical Sources". In: Hänsler, E., Schmidt, G. (Eds.), *Topics in Acoustic Echo and Noise Control*, Signals and Communication Technology, Springer Berlin Heidelberg, cap. 4, pp. 91–122, Mai. 2006.

[15] KIM, T., ATTIAS, H. T., LEE, S.-Y., et al. "Blind Source Separation Exploiting Higher-Order Frequency Dependencies", *Audio, Speech, and Language Processing, IEEE Transactions on*, v. 15, n. 1, pp. 70–79, Jan. 2007.

[16] CARDOSO, J.-F. "Blind Signal Separation: Statistical Principles", *Proceedings of the IEEE*, v. 86, n. 10, pp. 2009–2025, Out. 1998.

[17] ARAKI, S., SAWADA, H., MUKAI, R., et al. "Underdetermined Sparse Source Separation of Convolutive Mixtures with Observation Vector Clustering". In: *2006 IEEE International Symposium on Circuits and Systems*, p. 4, Singapore, Dez. 2006.

[18] YASHITA, M., HAMADA, N. "Time-Frequency Masking Method Using Wavelet Transform for BSS Problem". In: *TENCON 2006. 2006 IEEE Region 10 Conference*, pp. 1–4, Seville, Nov. 2006.

[19] SAWADA, H., ARAKI, S., MUKAI, R., et al. "Blind Extraction of a Dominant Source From Many Mixtures of Many Sources Using ICA and

Time-Frequency Masking". In: *2005 IEEE International Symposium on Circuits and Systems*, v. 6, pp. 5882–5885, Kobe, Mai. 2005.

[20] ARAKI, S., SAWADA, H., MUKAI, R., et al. "Underdetermined Blind Separation for Speech in Real Environments with Sparseness and ICA". In: *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, v. 3, pp. iii–811–814, Montréal, Mai. 2004.

[21] HYVÄRINEN, A., KARHUNEN, J., E., O. *Independent Component Analysis*. New York, John Wiley e Sons, Mai. 2001.

[22] LIU, W., WEISS, S. *Wideband Beamforming: Concepts and Techniques*. Wireless Communications and Mobile Computing. West Sussex, Wiley, Mai. 2010.

[23] JOHNSON, D., DUDGEON, D. *Array Signal Processing: Concepts and Techniques*. Prentice-Hall Signal Processing Series. New York, P T R Prentice Hall, Fev. 1993.

[24] NOGUEIRA, L. C. F., PETRAGLIA, M. R. "Sistema de Localização de Fontes Sonoras Baseado em Algoritmo de Separação Cega". In: *XXXI Simpósio Brasileiro de Telecomunicações (SBrT), 2013*, Set. 2013.

[25] NESTA, F., OMOLOGO, M. "Generalized State Coherence Transform for Multidimensional TDOA Estimation of Multiple Sources", *Audio, Speech, and Language Processing, IEEE Transactions on*, v. 20, n. 1, pp. 246–260, Jan. 2012.

[26] NGUYEN, U., PHAM, T. "Performance Assessment of Generalized Cross-Correlation Based Algorithms for Multisource Point-Based Localization and Detection", *Advanced Technologies for Communications (ATC), 2011 International Conference on*, pp. 303–306, Ago. 2011.

[27] KNAPP, C., CARTER, G. C. "The Generalized Correlation Method for Estimation of Time Delay", *Acoustics, Speech and Signal Processing, IEEE Transactions on*, v. 24, n. 4, pp. 320–327, Ago. 1976.

[28] DINIZ, P. S. R. *Adaptive Filtering: Algorithms and Practical Implementation*. 3 ed. New York, Springer, Out. 2010.

[29] FROST-III, O. L. "An Algorithm for Linearly Constrained Adaptive Array Processing". In: *Proceedings of the IEEE*, v. 60, Ago. 1972.

[30] JALIL, M., BUTT, F., MALIK, A. "Short-Time Energy, Magnitude, Zero Crossing Rate and Autocorrelation Measurement for Discriminating Voiced and Unvoiced Segments of Speech Signals". In: *Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE), 2013 International Conference on*, pp. 208–212, Mai. 2013.

[31] DINIZ, P., DA SILVA, E., NETTO, S. *Digital Signal Processing: System Analysis and Design*. Cambridge Textbooks. 2 ed. Cambridge, Cambridge University Press, Set. 2010.

[32] THIEMANN, J. *Acoustic Noise Suppression for Speech Signals using Auditory Masking Effects*. M.s. thesis, McGill University, Montréal, Jul. 2001.

[33] PEEBLES, P. *Probability, Random Variables, and Random Signal Principles*. McGraw-Hill Series in Electrical and Computer Engineering Series. 4 ed. New York, McGraw-Hill Companies,Incorporated, Jul. 2000.

[34] MATSUOKA, K. "Minimal Distortion Principle for Blind Source Separation". In: *Proc. of the 41st SICE Annual Conf.*, v. 4, pp. 2138–2143, Ago. 2002.

[35] HAYKING, S. *Adaptive Filter Theory*. 4 ed. New Jersey, Prentice Hall, Set. 2001.

[36] FARHANG-BOROUJENY, B. *Adaptive Filters Theory and Applications*. 2 ed. Chichester, John Wiley e Sons, Jun. 2013.

[37] VINCENT, E., GRIBOVAL, R., FEVOTTE, C. "Performance Measurement in Blind Audio Source Separation", *Audio, Speech and Language Processing, IEEE Transactions on*, v. 14, n. 4, pp. 1462–1469, Jul. 2006.

[38] LEHMANN, E. A., JOHANSSON, A. M. "Prediction of Energy Decay in Room Impulse Responses Simulated With an Image-Source Model", *The Journal of the Acoustical Society of America*, v. 124(1), pp. 269–277, Jul. 2008.

[39] DE CARVALHO ABI ABIB, G. *Separação Cega de Fontes Acústicas em Ambientes com Reverberação: Testes e Análises*. Projeto de graduação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Ago. 2013.

[40] DOBLINGER, G. "Adaptive Microphone Array Demos". Acesso em Out. 2013. Disponível em: <`http://www.nt.tuwien.ac.at/about-us/staff/gerhard-doblinger/demo-page/`>.

[41] LEE, I. "Permutation Correction in Blind Source Separation Using Sliding Subband Likelihood Function". In: *Independent Component Analysis and*

Signal Separation, v. 5441, *Lecture Notes in Computer Science*, pp. 767–774, Paraty, Mar. 2009. Springer Berlin Heidelberg.

Apêndice A

Teste de Qualidade de Áudio

Prezado voluntário, obrigado pela sua participação neste teste. O objetivo desta avaliação é determinar a eficácia de um sistema cuja função é recuperar vozes individuais a partir de gravações em que duas pessoas discursam simultaneamente. As gravações foram realizadas em ambientes distintos (salas que geram mais ou menos eco). A Fig.A.1 ilustra este processo.

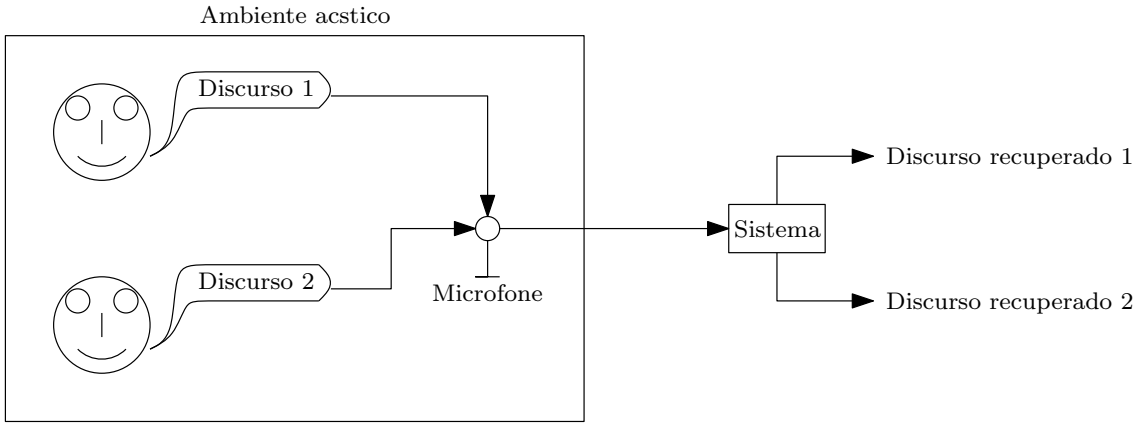


Figura A.1: Arranjo de teste.

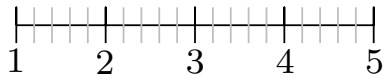
Metodologia: para cada ambiente acústico serão primeiramente apresentados os discursos 1 e 2 separadamente, de modo que o avaliador possa conhecer o resultado ideal que se espera do sistema de separação de vozes. Em seguida, serão apresentados os resultados (discursos recuperados) obtidos por diversos sistemas diferentes.

Objetivo: caberá ao avaliador pontuar, na escala de 1 a 5, o quanto a interferência remanescente do outro discurso prejudica a compreensão do discurso recuperado. Nesta referida escala, 1 significa "interferência extremamente prejudicial" e 5 significa "interferência imperceptível", sendo que o avaliador poderá fazer uso tanto da parte inteira quanto da fracionária da escala.

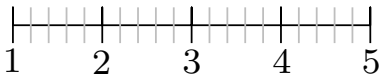
Reprodução 1

Reprodução 2

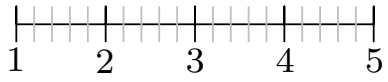
Reprodução 3



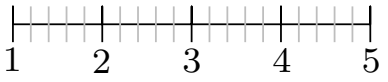
Reprodução 4



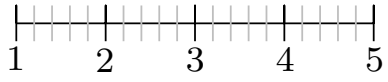
Reprodução 5



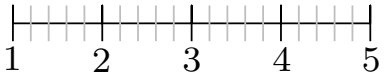
Reprodução 6



Reprodução 7

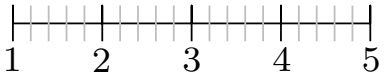


Reprodução 8



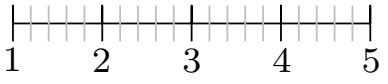
Reprodução 9

Reprodução 10



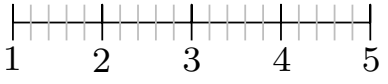
Reprodução 11

Reprodução 12



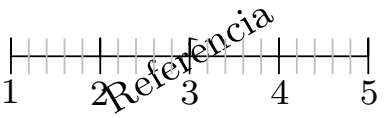
Reprodução 13

Reprodução 14

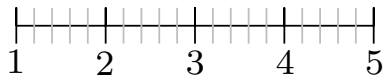


Reprodução 15

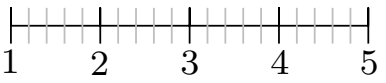
Reprodução 16



Reprodução 17

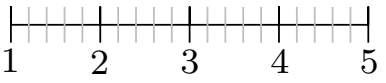


Reprodução 18

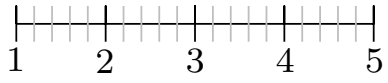


Reprodução 19

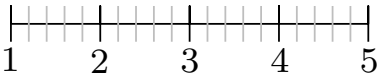
Reprodução 20



Reprodução 21

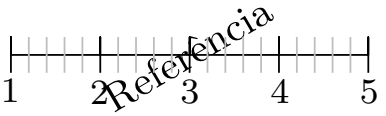


Reprodução 22

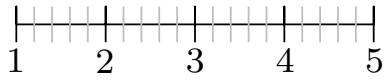


Reprodução 23

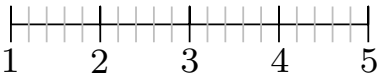
Reprodução 24



Reprodução 25

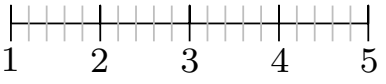


Reprodução 26

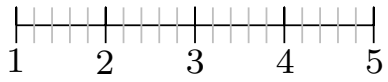


Reprodução 27

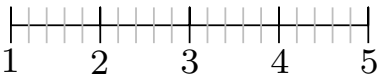
Reprodução 28



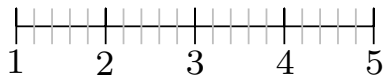
Reprodução 29



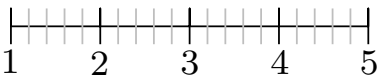
Reprodução 30



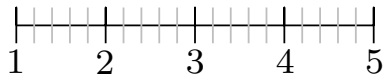
Reprodução 31



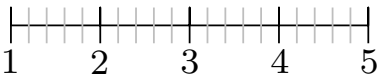
Reprodução 32



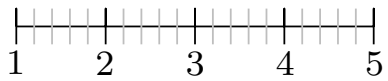
Reprodução 33



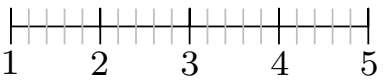
Reprodução 34



Reprodução 35



Reprodução 36



Apêndice B

Monólogos Empregados nos Testes

Em seguida exibe-se a transcrição dos monólogos que constituíram as misturas usadas nos testes de separação cega de fontes apresentados ao longo desta dissertação.

Monólogos gravados no PADS UFRJ

> **Monólogo 1**: A consciência de si nasce de um desejo vivo, ou seja, desejar o desejo do outro enquando desejo vivo de outra consciência de si, originando o homem através de significado pela utilização da linguagem. Dentro desta perspectiva, a relação entre o homem e as coisas é desejar conhecimento. A relação entre homens é desejar [...]
>
> **Monólogo 2**: Um homem estava sentado num banco sem pernas, à luz de um candeeiro apagado, quando viu um peixe afogado ser desenterrado do lago. A sua sorte foi estar de olhos fechados. A múmia que estava ao seu lado gritou baixinho, que essa sorte só acontece a quem não tem olho, grito esse que assustou o elefante sem tromba, fazendo com que desatasse a voar dali.

Monólogos gravados para geração das misturas em ambiente simulado

> **Monólogo 1**: Among them are canvasses by a young artist. Building from the ground-up is very costly. Next year we will see several more exhibitions. The number of works on view will increase.
>
> **Monólogo 2**: "This food is too spicy", he complained. Young men can be very arrogant and rude. So Marus owns the big shipping company. Their eyes met across the table.

Monólogos gravados na sala D-105 do CT UFRJ

Monólogo 1: [...] Pouco sucesso apontam duas realidades na opinião dos pesquisadores: ou as crianças estão entrando tarde na escola, ou estão tendo um ensino [...]

Monólogo 2: Estes são os bons momentos de Michele com o pessoal do interior, gente querida. Boas lembranças, mas é só. A publicitária não quer ninguém.