



USO DE BANCO DE FILTROS EM SEPARAÇÃO CEGA DE FONTES

Daniel Francisco Pinto

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
Outubro de 2012

USO DE BANCO DE FILTROS EM SEPARAÇÃO CEGA DE FONTES

Daniel Francisco Pinto

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. Amaro Azevedo de Lima, Ph.D.

Prof. Julio Cesar Boscher Torres, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
OUTUBRO DE 2012

Pinto, Daniel Francisco

Uso de Banco de Filtros em Separação Cega de Fontes/Daniel Francisco Pinto. – Rio de Janeiro: UFRJ/COPPE, 2012.

XIV, 67 p.: il.; 29, 7cm.

Orientador: Mariane Rembold Petraglia

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2012.

Referências Bibliográficas: p. 66 – 67.

1. Separação Cega de Fontes. 2. Processamento Multi-taxas. 3. Banco de Filtros. 4. Processamento de Sinais de Áudio. I. Petraglia, Mariane Rembold. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

Aos meus pais e irmão

Agradecimentos

Agradeço, principalmente, a Deus, que iluminou esta trajetória, por mais complicada que tenha sido.

Aos meus pais, Manuel e Maria Aldina, ao meu irmão Rodrigo, aos meus familiares e amigos que sempre me dispensaram uma palavra de incentivo, deram força e me apoiaram em todos os momentos durante o mestrado.

Aos professores do Programa de Engenharia Elétrica (PEE) de COPPE/UFRJ, que colaboraram com o meu aprendizado no mestrado.

À professora Mariane Rembold Petraglia, que sempre com paciência, compreensão e sabedoria me auxiliou, estimulou e orientou durante esse período. Por mais palavras que eu escreva, estas não serão suficientes para demonstrar todo o incentivo que me foi dado por ela.

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

USO DE BANCO DE FILTROS EM SEPARAÇÃO CEGA DE FONTES

Daniel Francisco Pinto

Outubro/2012

Orientador: Mariane Rembold Petraglia

Programa: Engenharia Elétrica

A Separação Cega de Fontes tem sido utilizada em vários setores da engenharia como telecomunicações, biomedicina e separação de áudio. Uma das abordagens para a resolução deste problema no caso de misturas convolutivas é realizar a separação no domínio da frequência (FDBSS, do inglês *Frequency Domain Blind Source Separation*). Nesta dissertação propomos realizar um pré-processamento das misturas, decompondo-as em sub-bandas, antes de aplicar o algoritmo FDBSS. Desta forma, podemos explorar as características do sinal de voz e da sua propagação acústica em ambientes reais para melhorar o desempenho do método de separação. Verificamos, através de simulações e de sinais gravados em uma sala, a influência dos bancos de filtros e do tamanho do bloco de dados usado em cada subbanda no resultado final da separação.

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

USING FILTER BANKS IN BLIND SOURCE SEPARATION

Daniel Francisco Pinto

October/2012

Advisor: Mariane Rembold Petraglia

Department: Electrical Engineering

Blind Source Separation has been used in various engineering sectors such as telecommunications, biomedicine and audio separation. One approach to solve this problem in the case of convolutive mixtures is to perform the separation in the frequency domain (FDBSS, *Frequency Domain Blind Source Separation*). In this thesis, we propose to perform a preprocessing of the mixtures, decomposing them into sub-bands, before applying the FDBSS algorithm. In this way, we can exploit the characteristics of the voice signal and its acoustic propagation in real environments to improve the performance of the separation method. We determined, through simulations and using signals recorded in a room, the influence of the filter banks and the size of data block used in each subband in the quality of the final separation.

Sumário

Lista de Figuras	x
Lista de Tabelas	xii
1 Introdução	1
1.1 Organização da Dissertação	2
2 Separação Cega de Fontes	4
2.1 Tipos de Misturas de Sinais	4
2.1.1 Misturas Instantâneas	5
2.1.2 Misturas Convolutivas	6
2.2 Problemas com Ambiguidades em Separação Cega	7
2.3 Análise de Componentes Independentes	9
2.4 Abordagens de ICA	9
2.4.1 Maximização da não-gaussianidade	9
2.4.2 Maximização da verossimilhança	10
2.4.3 Minimização de informação mútua	10
2.5 O Algoritmo <i>Natural</i> ICA	11
2.6 Separação Cega de Fontes no Domínio da Frequência	12
2.6.1 Passando para o Domínio da Frequência	13
2.6.2 Branqueamento	14
2.6.3 Permutação	14
2.6.4 Escalonamento	18
2.6.5 Suavização	19
2.6.6 Voltando para o Domínio do Tempo	20
3 Bancos de Filtros	21
3.1 Sistemas Multitaxas	21
3.1.1 Decimador	21
3.1.2 Interpolador	22
3.2 Banco de Filtros	23
3.2.1 Banco de Filtros Modulados Por Cosseno	24

3.2.2	Banco de Filtros Não-Uniforme com Decomposição Diádica . . .	27
4	Separação Cega de Fontes em Subbandas	31
4.1	Características da Reverberação em Salas	32
4.2	Separação de Fontes por Subbanda	33
5	Resultados das Simulações	37
5.1	Ambiente de Simulação	37
5.2	Arquivos de Áudio	38
5.3	Avaliação de Desempenho	38
5.4	FDBSS	39
5.5	FDBSS em Subbandas com Banco Modulado por Cosseno	40
5.6	FDBSS em Subbandas com Banco de Filtros Não-Uniforme	54
6	Conclusões	63
	Referências Bibliográficas	66

Lista de Figuras

1.1	<i>Cocktail Party</i> - Dejesa-se obter as fontes presentes num meio com interferências	2
2.1	Problemas de ambiguidades em separação cega de fontes	8
2.2	Estágios do FDBSS	13
2.3	O Modelo do Campo Próximo	15
2.4	Exemplo de <i>clusters</i> para localização de fontes	17
3.1	Decimação do sinal $x(n)$ com $M = 3$	22
3.2	Efeito do decimador na resposta em frequência de $x(n)$	22
3.3	Interpolação do sinal $x(n)$ com $L = 4$	23
3.4	Efeito do interpolador no espectro de $x(n)$	23
3.5	Espectro de frequência dividido em subbandas	24
3.6	Bancos de filtros de (a) análise e (b) síntese	24
3.7	Filtros formadores do banco de filtros modulado por cosseno	25
3.8	Banco de Filtros Modulados por Cossenos	27
3.9	Identities Nobres do (a) Decimador e (b) Interpolador	28
3.10	Estruturas de banco de filtros não uniforme de (a) análise e (b) síntese	28
3.11	Estruturas equivalentes dos bancos de (a) análise e (b) síntese	29
3.12	Respostas em frequência de bancos de filtros de análise e síntese não-uniformes	30
4.1	Características de tempo de reverberação e frequência de um ambiente fechado	33
4.2	Etapas do método FDBSS	34
4.3	FDBSS em subbandas com bancos de filtros modulados por cosseno	35
4.4	FDBSS em subbandas com bancos de filtros não-uniformes	36
5.1	Arranjo da sala utilizada nas simulações	38
5.2	Respostas em frequência dos filtros modulados por cossenos com 2 bandas	41

5.3	Respostas em frequência dos filtros modulados por cossenos com 4 bandas	41
5.4	Respostas em frequência dos filtros modulados por cossenos com 8 bandas	42
5.5	Respostas em frequência dos filtros modulados por cossenos com 16 bandas	42
5.6	Banco de Filtros Não-Uniformes com Decimação Não-Crítica	58

Lista de Tabelas

5.1	FDBSS para Reverberação de 130 ms	40
5.2	FDBSS para Reverberação de 300 ms	40
5.3	FDBSS com banco de filtros modulado por cosseno com 2 bandas . .	43
5.4	FDBSS com banco de filtros modulado por cosseno com 4 bandas . .	43
5.5	FDBSS com banco de filtros modulado por cosseno com 8 bandas . .	44
5.6	FDBSS com banco de filtros modulado por cosseno com 16 bandas . .	44
5.7	FDBSS com banco de filtros modulado por cosseno com 2 bandas . .	45
5.8	FDBSS com banco de filtros modulado por cosseno com 4 bandas . .	45
5.9	FDBSS com banco de filtros modulado por cosseno com 8 bandas . .	46
5.10	FDBSS com banco de filtros modulado por cosseno com 16 bandas . .	46
5.11	FDBSS com banco de filtros modulado por cosseno com 4 bandas . .	47
5.12	FDBSS com banco de filtros modulado por cosseno com 8 bandas . .	47
5.13	FDBSS com banco de filtros modulado por cosseno com 16 bandas . .	48
5.14	FDBSS com banco de filtros modulado por cosseno com 4 bandas . .	48
5.15	FDBSS com banco de filtros modulado por cosseno com 8 bandas . .	49
5.16	FDBSS com banco de filtros modulado por cosseno com 16 bandas . .	49
5.17	FDBSS com banco de filtros modulado por cosseno com janela de 256 pontos	50
5.18	FDBSS com banco de filtros modulado por cosseno com janela de 512 pontos	50
5.19	FDBSS com banco de filtros modulado por cosseno com janela de 1024 pontos	51
5.20	FDBSS em bandas de banco de filtros de cossenos FDBSS com banco de filtros modulado por cosseno com janela de 2048 pontos	51
5.21	FDBSS com banco de filtros modulado por cosseno com janelas distintas	52
5.22	FDBSS com banco de filtros modulado por cosseno com janela de 256 pontos	52
5.23	FDBSS com banco de filtros modulado por cosseno com janela de 512 pontos	53

5.24 FDBSS com banco de filtros modulado por cosseno com janela de 1024 pontos	53
5.25 FDBSS com banco de filtros modulado por cosseno com janela de 2048 pontos	53
5.26 FDBSS com banco de filtros modulado por cosseno com janelas diferentes	54
5.27 FDBSS em Subbandas com Banco Não-Uniforme com janela de 256 pontos	55
5.28 FDBSS em Subbandas com Banco Não-Uniforme com janela de 512 pontos	55
5.29 FDBSS em Subbandas com Banco Não-Uniforme com janela de 1024 pontos	55
5.30 FDBSS em Subbandas com Banco Não-Uniforme com janela de 2048 pontos	56
5.31 FDBSS em Subbandas com Banco Não-Uniforme com janela de 256 pontos	56
5.32 FDBSS em Subbandas com Banco Não-Uniforme com janela de 512 pontos	57
5.33 FDBSS em Subbandas com Banco Não-Uniforme com janela de 1024 pontos	57
5.34 FDBSS em Subbandas com Banco Não-Uniforme com janela de 2048 pontos	57
5.35 FDBSS em Subbandas com Banco Não-Uniforme com Decimação Não-Crítica com janela de 256 pontos	59
5.36 FDBSS em Subbandas com Banco Não-Uniforme com Decimação Não-Crítica com janela de 512 pontos	59
5.37 FDBSS em Subbandas com Banco Não-Uniforme com Decimação Não-Crítica com janela de 1024 pontos	59
5.38 FDBSS em Subbandas com Banco Não-Uniforme com Decimação Não-Crítica com janela de 2048 pontos	60
5.39 FDBSS em Subbandas com Banco Não-Uniforme com Decimação Não-Crítica com tamanhos de janela distintos	60
5.40 FDBSS em Subbandas com Banco Não-Uniforme com Decimação Não-Crítica com janela de 256 pontos	61
5.41 FDBSS em Subbandas com Banco Não-Uniforme com Decimação Não-Crítica com janela de 512 pontos	61
5.42 FDBSS em Subbandas com Banco Não-Uniforme com Decimação Não-Crítica com janela de 1024 pontos	61

5.43 FDBSS em Subbandas com Banco Não-Uniforme com Decimação	
Não-Crítica com janela de 2048 pontos	62
5.44 FDBSS em Subbandas com Banco Não-Uniforme com Decimação	
Não-Crítica com comprimentos de janela diferentes	62

Capítulo 1

Introdução

A separação cega de fontes é um assunto bastante pesquisado atualmente na área de processamento de sinais. Isso se deve pois há uma gama de aplicações nas áreas de telecomunicações, especialmente telefonia celular e sistemas de conferência, de separação de áudio e de engenharia biomédica.

O objetivo da separação cega de fontes é recuperar fontes independentes a partir de sinais observados em múltiplos sensores, sem se ter conhecimento de como estas fontes independentes se misturaram ou mesmos das próprias características das fontes. Um exemplo, muito usado na literatura, é o problema do *Cocktail Party*. Nesse problema temos, dentro de um mesmo ambiente, diversas pessoas conversando sobre diversos assuntos. Um ser humano consegue filtrar a conversa de interesse, desprezando as demais fontes de informação presentes no ambiente. A Fig. 1.1 exemplifica esse problema.

Existem diversas abordagens para resolver esse problema. A Análise de Componentes Principais (PCA), a Decomposição de Valores Singulares (SVD) e a Análise de Componentes Independentes (ICA) são três das técnicas mais utilizadas para isso e, cada uma delas está associada a uma família de algoritmos.

Para misturas instantâneas de sinais, as quais ocorrem em condições ideais, os algoritmos atuais apresentam bons desempenhos. Na prática, os sinais podem sofrer reverberações, devido às reflexões e refrações nas paredes e objetos do ambiente em que são observados, além da presença de ruídos inerentes ao meio. Nesses casos, os algoritmos propostos recentemente na literatura ainda apresentam dificuldades para realizar uma clara separação. Para melhorar os seus desempenhos, propomos a utilização de técnicas conhecidas de outras áreas de processamento de sinais, os bancos de filtros e processamento em multitaxas.

Nesta dissertação, verificaremos através de simulações, como o pré-processamento dos sinais observados, empregando bancos de filtros e processamento em multitaxas, e a utilização de parâmetros específicos para o algoritmo de separação nas diferentes faixas de frequência podem resultar na melhoria de desempenho dos

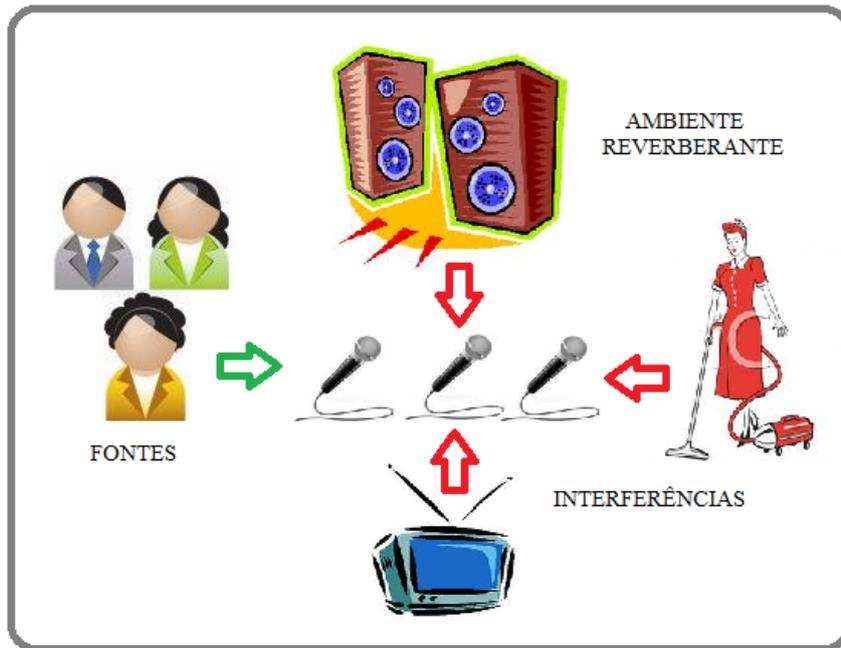


Figura 1.1: *Cocktail Party* - Dejesa-se obter as fontes presentes num meio com interferências

algoritmos baseados no princípio de ICA, que admite que todas as fontes a serem estimadas são independentes entre si. Nas simulações realizadas nesta dissertação, utilizaremos sinais gravados em estúdio e simuladores acústicos de salas reverberantes. Desta forma, temos um controle maior das observações utilizadas, podendo associar os parâmetros ótimos encontrados ao tempo de reverberação do ambiente.

1.1 Organização da Dissertação

No Capítulo 2 será apresentado o problema da separação cega de fontes. Neste capítulo, mostraremos os tipos de mistura de sinais existentes e os problemas encontrados após a separação. Citaremos as técnicas mais conhecidas para a resolução deste problema, dando maior foco ao algoritmo *Natural* ICA. Em seguida, será descrito o passo-a-passo do processamento da separação cega das fontes no domínio da frequência.

No Capítulo 3 serão apresentados os conceitos de processamento em multitaxas utilizados neste trabalho. Apresentaremos os componentes básicos que permitem o processamento em multitaxas, o decimador e o interpolador. Também apresentaremos estruturas de bancos de filtros conhecidas como Banco de Filtro Modulados por Cosseno e Banco de Filtros Não-Uniforme na Forma de Árvore Diádica.

No Capítulo 4, introduziremos a motivação para o algoritmo de separação cega de fontes proposto nesta dissertação. Discutiremos os fatores a serem considerados na seleção dos parâmetros do banco de filtros, utilizado no pré-processamento, e do

algoritmo de separação, empregado em cada subbanda.

No Capítulo 5, são apresentados o ambiente de simulações e as figuras de mérito utilizadas para a análise de resultados. Em seguida, são mostrados os resultados das simulações realizadas com as diferentes estruturas de bancos de filtros, comparando-se os resultados obtidos com as várias configurações.

Finalizando, o Capítulo 6 apresenta as conclusões desta dissertação.

Capítulo 2

Separação Cega de Fontes

Neste capítulo faremos uma apresentação sucinta do problema de separação cega de fontes. Inicialmente, serão mostrados os tipos de mistura de sinais existentes e o problema de ambiguidade, de escalonamento e permutação da solução da separação cega.

Em seguida, será descrita a análise de componentes independentes. Serão apresentadas as abordagens mais utilizadas para obter a solução desse problema, citando seus algoritmos mais representativos. Daremos destaque ao algoritmo *Natural ICA*, pois este será utilizado em nossas simulações.

A separação cega de fontes no domínio da frequência será então descrita, desde a passagem dos sinais para o domínio da frequência, passando pelos algoritmos de branqueamento, separação, solução dos problemas de permutação e escalonamento, além da suavização das estimativas das fontes e do retorno dos sinais ao domínio do tempo.

2.1 Tipos de Misturas de Sinais

Em separação cega de fontes, podemos definir o tipo de mistura de acordo com a forma na qual as N fontes de informação $s_i(n)$, $i = 1, \dots, N$, se combinam, gerando os sinais observados em múltiplos sensores. As misturas podem ser classificadas como lineares e não-lineares, sendo apenas as lineares investigadas neste trabalho.

Basicamente, existem duas formas de misturas lineares: instantâneas e convolutivas. Para expressá-las matematicamente, inicialmente definimos o vetor de fontes de informação no instante n como:

$$\mathbf{s}(n) = \begin{bmatrix} s_1(n) \\ \dots \\ s_N(n) \end{bmatrix} \quad (2.1)$$

2.1.1 Misturas Instantâneas

Denomina-se a mistura instantânea quando seu sinal é composto conforme a equação:

$$x_l(n) = \sum_{j=1}^N h_{l,j} s_j(n) \quad (2.2)$$

para $x_l(n)$, $l = i, \dots, M$, onde M é o número de sinais observados. Este tipo de mistura caracteriza-se pelos coeficientes $h_{l,j}$ numéricos, formando uma combinação linear das fontes [1]. Matricialmente podemos reescrever a equação acima como:

$$\mathbf{x}(n) = \mathbf{H}\mathbf{s}(n) \quad (2.3)$$

onde

$$\mathbf{H} = \begin{bmatrix} h_{1,1} & h_{1,2} & \dots & h_{1,N} \\ h_{2,1} & h_{2,2} & \dots & h_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ h_{M,1} & h_{M,2} & \dots & h_{M,N} \end{bmatrix} \quad (2.4)$$

e

$$\mathbf{x}(n) = \begin{bmatrix} x_1(n) \\ x_2(n) \\ \vdots \\ x_M(n) \end{bmatrix} \quad (2.5)$$

A matriz \mathbf{H} é denominada matriz de misturas e $\mathbf{x}(n)$ é o vetor de observações. Vale lembrar que no caso acima, além de ser instantânea, a mistura é não-ruídosa e linear.

Desejamos, portanto, obter estimativas dos sinais das fontes de informação, $s_j(n)$, $j = 1, \dots, N$, a partir do vetor de observações, $\mathbf{x}(n)$. Portanto, seja \mathbf{W} a matriz separadora e $y_i(n)$, $i = 1, \dots, N$ os sinais recuperados a partir de $\mathbf{x}(n)$. Matricialmente, podemos escrever:

$$\begin{bmatrix} y_1(n) \\ \vdots \\ y_N(n) \end{bmatrix} = \begin{bmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,M} \\ w_{2,1} & w_{2,2} & \dots & w_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N,1} & w_{N,2} & \dots & w_{N,M} \end{bmatrix} \begin{bmatrix} x_1(n) \\ x_2(n) \\ \vdots \\ x_M \end{bmatrix} \rightarrow \mathbf{y}(n) = \mathbf{W}\mathbf{x}(n) \quad (2.6)$$

Portanto, das equações matriciais acima, se conhecemos \mathbf{H} , podemos obter \mathbf{W} facilmente por um processo de inversão de matrizes, no caso do número de observações ser o mesmo que o número de fontes ($M = N$), ou pelo cálculo da matriz pseudo-

inversa, quando $M > N$. Normalmente somente possuímos $\mathbf{x}(n)$ e a análise de componentes independentes é a ferramenta utilizada para estimar os sinais originais [1].

2.1.2 Misturas Convolutivas

A mistura convolutiva é a que se aproxima de uma situação real na maioria das aplicações que envolvem sinais de áudio. Nos receptores das fontes no instante n , além da componente $s_j(n)$, aparecem suas versões atrasadas devido ao multipercurso que o sinal sofre num ambiente fechado devido à reverberação ou a objetos presentes no seu entorno. Matematicamente, representamos uma mistura convolutiva como:

$$x_l(n) = \sum_{j=1}^N \left(\sum_{p=0}^{P-1} h_{lj}(p) s_j(n-p) \right) \quad (2.7)$$

onde admitimos que o ambiente possa ser modelado por filtros FIR, com resposta ao impulso $h_{l,j}(n)$ de comprimento P . Podemos reescrever a equação acima na forma matricial:

$$\mathbf{x}(n) = \mathbf{H}\mathbf{s}(n) \quad (2.8)$$

onde

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_{1,1} & \mathbf{h}_{1,2} & \dots & \mathbf{h}_{1,N} \\ \mathbf{h}_{2,1} & \mathbf{h}_{2,2} & \dots & \mathbf{h}_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{h}_{M,1} & \mathbf{h}_{M,2} & \dots & \mathbf{h}_{M,N} \end{bmatrix} \quad (2.9)$$

sendo cada elemento de \mathbf{H} um vetor linha contendo os P coeficientes $h_{l,j}(n)$ do filtro FIR que relaciona a j -ésima fonte com a l -ésima mistura. O produto da matriz \mathbf{H} e do $\mathbf{s}(n)$ agora representa o produto entre cada vetor linha $\mathbf{h}_{l,j}$ e o vetor coluna $\mathbf{s}_j(n)$ que contém P amostras da j -ésima fonte $s_j(n)$. Da mesma forma devemos reescrever a equação (2.6) no caso convolutivo como:

$$\mathbf{y}(n) = \mathbf{W}\mathbf{x}(n) \quad (2.10)$$

sendo

$$\mathbf{W} = \begin{bmatrix} \mathbf{w}_{1,1} & \mathbf{w}_{1,2} & \dots & \mathbf{w}_{N,1} \\ \mathbf{w}_{2,1} & \mathbf{w}_{2,2} & \dots & \mathbf{w}_{N,2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{w}_{N,1} & \mathbf{w}_{N,2} & \dots & \mathbf{w}_{N,M} \end{bmatrix} \quad (2.11)$$

em que cada elemento de \mathbf{W} contém os coeficientes $w_{q,l}(n)$ dos filtros de separação usados para estimar a fonte $y_q(n)$ a partir dos sinais dos M sensores para $l = 1, \dots, M$, da seguinte forma:

$$y_q(n) = \sum_{l=1}^M \left(\sum_{p=0}^{Q-1} w_{ql}(p) x_l(n-p) \right) \quad (2.12)$$

Para que tenhamos sucesso ao recuperar as fontes, os algoritmos de análise de componentes independentes devem ser modificados, adaptando-se para o caso convolutivo [1].

2.2 Problemas com Ambiguidades em Separação Cega

Após a recuperação dos sinais originais utilizando a análise de componentes independentes, podemos nos deparar com duas possíveis ambiguidades referentes ao processamento[2], descritas a seguir.

Escalonamento

Uma constante c pode multiplicar a fonte $s_i(n)$ ou a i -ésima coluna de \mathbf{H} , o que resultará no mesmo efeito no sinal de saída $\mathbf{x}(n)$, sem sabermos qual dos dois produtos ocorreu. O mesmo fato também poderá levar a uma ambiguidade do sinal.

Permutação

Uma vez que não temos nenhuma informação sobre a ordem com que os componentes de $\mathbf{s}(n)$ se organizam e de \mathbf{H} , não podemos afirmar que o i -ésimo elemento de \mathbf{y} seja a estimativa do i -ésimo sinal de $\mathbf{s}(n)$.

A Fig. 2.1 ilustra os efeitos de permutação e escalonamento [2].

Podemos modelar matematicamente as ambiguidades. Sejam $\mathbf{\Lambda}$ a matriz de escalonamento e \mathbf{P} a matriz de permutação. A matriz $\mathbf{\Lambda}$ é uma matriz diagonal, enquanto \mathbf{P} é uma matriz com todos os elementos de uma linha iguais à zero exceto o da posição onde o sinal aparece, representado pelo valor 1. Como exemplos temos as matrizes dadas em (2.13) e (2.14) para $N = 3$.

$$\mathbf{\Lambda} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & 3 \end{bmatrix} \quad (2.13)$$

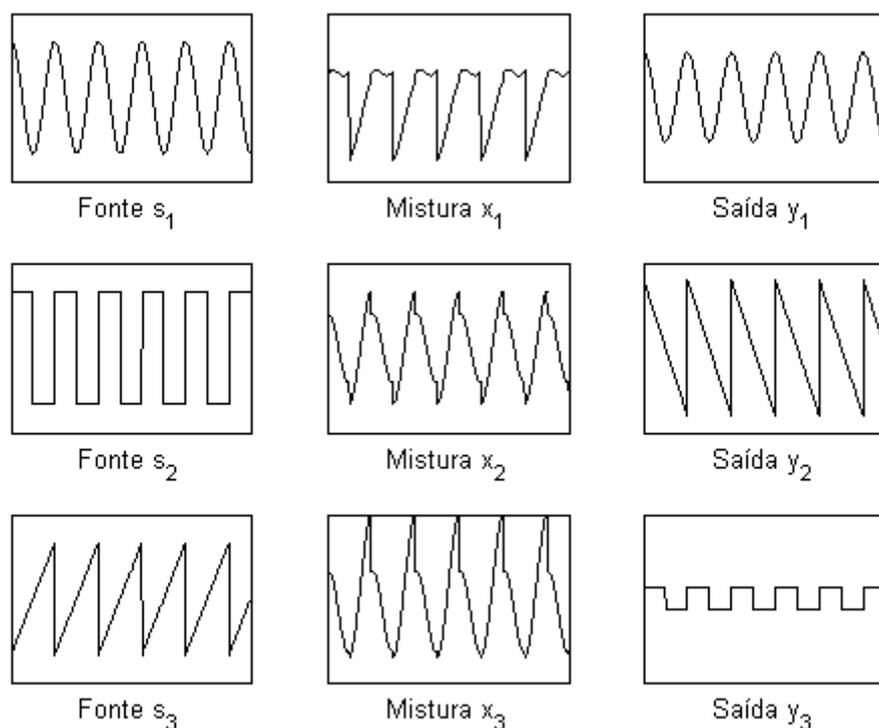


Figura 2.1: Problemas de ambiguidades em separação cega de fontes

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad (2.14)$$

Com isso, se o processamento de separação funcionar, a resposta será um vetor $\mathbf{y}(n)$ tal que:

$$\mathbf{y}(n) = \mathbf{P}\mathbf{\Lambda}\mathbf{s}(n) \quad (2.15)$$

No domínio do tempo essas ambiguidades não interferem na qualidade da separação, pois tanto a ordem dos sinais quanto suas amplitude não são importantes no resultado final. Em compensação, nos algoritmos no domínio da frequência, esses problemas são críticos e devem ser solucionados [1] para obtermos boas estimativas das fontes.

2.3 Análise de Componentes Independentes

A análise de componentes independentes (ICA, do inglês *Independent Component Analysis*) é uma metodologia utilizada para encontrar, a partir de observações de misturas, as fontes de informação originiais. Nesse método, não se tem nenhum conhecimento prévio de como as fontes se comportam. Para que esse processamento seja possível, três fatores são importantes [3]:

- As componentes (fontes de informação) são estatisticamente independentes entre si.
- As componentes devem ser não gaussianas.
- A matriz misturadora é quadrada, ou seja, o número de observações é o mesmo que o número de fontes.

2.4 Abordagens de ICA

Para a resolução do problema de análise de componentes independentes, algumas abordagens foram propostas na literatura [3]. As principais serão descritas a seguir.

2.4.1 Maximização da não-gaussianidade

A não-gaussianidade é uma das características desejadas para os sinais a serem recuperados. Baseia-se na idéia do teorema do limite central que diz que a soma de variáveis aleatórias independentes tende a uma variável gaussiana [3]. Portanto, maximizando-se a não-gaussianidade dos sinais, tende-se a obter sinais os quais são os mais próximos possíveis das fontes originais, assumidamente não-gaussianas. Sejam $s_m(n)$ sinais não-gaussianos independentes e $y_k(n)$ uma combinação linear destas fontes, ou seja:

$$y_k(n) = \sum_m q_m s_m(n) \quad (2.16)$$

Se houvesse uma reconstituição perfeita de $s_k(n)$, o coeficiente q_k seria igual a 1 e os demais iguais a 0. Como a soma das componentes independentes vai ser mais gaussiana que cada uma delas separadamente, basta encontrar os valores de $q_k, k = 1, \dots, N$ que deixem o somatório menos gaussiano possível. Os algoritmos que usam essa abordagem baseiam-se nas seguintes técnicas:

- Curtose ou estatísticas de ordem superior
- Negentropia

- *Projection Pursuit*

A curtose e a negentropia são medidas de gaussianidade [3].

2.4.2 Maximização da verossimilhança

Esse princípio baseia-se nas estimações estatísticas das componentes. Seja $\mathbf{x} = \mathbf{B}\mathbf{s}$, onde \mathbf{B} é uma matriz de misturas. Podemos descrever a densidade de probabilidade do vetor \mathbf{x} como [4]:

$$p_x(\mathbf{x}) = |\det(\mathbf{B})| p_s(\mathbf{s}) = |\det(\mathbf{B})| \prod_{i=1}^N s_i \quad (2.17)$$

$$p_x(\mathbf{x}) = |\det(\mathbf{B})| \prod_{i=1}^N p_i(\mathbf{b}_i^T \mathbf{x}) \quad (2.18)$$

Admitindo que temos T amostras de $\mathbf{x} \rightarrow \{\mathbf{x}(1), \dots, \mathbf{x}(T)\}$, podemos calcular a verossimilhança como:

$$L(\mathbf{B}) = \prod_{t=1}^T \prod_{i=1}^N p_i(\mathbf{b}_i^T \mathbf{x}) |\det(\mathbf{B})| \quad (2.19)$$

Para facilitar os algebrismos, utilizamos o logaritmo da verossimilhança [4]. Também podemos dividir ambos os membros da equação por T , obtendo o operador esperança. Portanto a função que deve ser maximizada é:

$$\frac{1}{T} \log(L(\mathbf{B})) = E \left[\sum_{i=1}^N p_i(\mathbf{b}_i^T \mathbf{x}) \right] + \log(|\det(\mathbf{B})|) \quad (2.20)$$

Os algoritmos que usam essa abordagem são:

- Bell-Sejnowski
- Gradiente Natural
- Infomax (baseado na entropia da saída)
- *Projection Pursuit*

2.4.3 Minimização de informação mútua

Essa abordagem baseia-se na teoria da informação. Deseja-se uma figura de mérito que meça a dependência dos componentes de um vetor aleatório. Define-se o processamento das componentes como aquela que minimiza a informação mútua das misturas.

A definição da entropia diferencial de um vetor de variáveis aleatórias \mathbf{y} , com função densidade de probabilidade $p(\mathbf{y})$, é dada por:

$$H(\mathbf{y}) = - \int p(\mathbf{y}) \log p(\mathbf{y}) d\mathbf{y} \quad (2.21)$$

Um resultado fundamental da teoria da informação é que uma variável gaussiana tem a maior entropia entre todas as variáveis aleatórias de mesma variância. Portanto, a entropia pode ser usada como medida de não-gaussianidade. Define-se uma versão normalizada da entropia diferencial, a negentropia J , como:

$$J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}) \quad (2.22)$$

onde \mathbf{y}_{gauss} é um vetor aleatório gaussiano com mesma matriz correlação que \mathbf{y} . A negentropia tem valor zero para y sendo uma variável gaussiana e maior que zeros para qualquer outra variável. Define-se a informação mútua I entre as componentes y_i , $i = 1, \dots, m$, de \mathbf{y} como:

$$I(y_1, y_2, \dots, y_m) = \sum_{i=1}^m H(y_i) - H(\mathbf{y}) \quad (2.23)$$

Para uma transformação linear inversível $\mathbf{y} = \mathbf{B}\mathbf{x}$, tem-se:

$$I(y_1, y_2, \dots, y_m) = \sum_{i=1}^m H(y_i) - H(\mathbf{x}) - \log |\det(\mathbf{B})| \quad (2.24)$$

Pode-se mostrar que a informação mútua I pode ser expressa como:

$$I(y_1, y_2, \dots, y_m) = C - \sum_i J(y_i) \quad (2.25)$$

onde C é uma constante que não depende de \mathbf{B} . Portanto, a equação acima mostra que a minimização da informação mútua é equivalente à maximização da soma das não-gaussianidades das estimativas das componentes independentes.

Os algoritmos que usam essa abordagem baseiam-se em:

- Aproximações da Negentropia
- Aproximação das funções de densidade de probabilidade das componentes independentes

2.5 O Algoritmo *Natural ICA*

O algoritmo utilizado nessa dissertação para realizar a separação cega de fontes é o *Natural ICA*, da família dos algoritmos de maximização da verossimilhança. Baseia-

se nos métodos iterativos do tipo gradiente. Derivando a função da Eq. (2.20) em relação a \mathbf{B} , temos [4]:

$$\frac{1}{T} \frac{\partial \log L}{\partial \mathbf{B}} = [\mathbf{B}^T]^{-1} + E [\mathbf{g}(\mathbf{B}\mathbf{x}) \mathbf{x}^T] \quad (2.26)$$

onde $\mathbf{g}(\mathbf{x}) = [g_1(x_1) \cdots g_N(x_N)]^T$ é um vetor de funções, chamadas de funções *score* [4], definidas por:

$$g_i = \frac{1}{p_{x_i}(x_i)} \frac{\partial p_{x_i}(x_i)}{\partial x_i} \quad (2.27)$$

Portanto, usando a idéia dos algoritmos iterativos do tipo gradiente:

$$\Delta \mathbf{B} \propto [\mathbf{B}^T]^{-1} + E [\mathbf{g}(\mathbf{B}\mathbf{x}) \mathbf{x}^T] \quad (2.28)$$

Multiplicando à direita ambos os membros do lado direito da equação acima por $\mathbf{B}^T \mathbf{B}$ e sabendo que $\mathbf{y} = \mathbf{B}\mathbf{x}$, obtemos o chamado gradiente natural, dado por:

$$\Delta \mathbf{B} \propto (\mathbf{I} + E [\mathbf{g}(\mathbf{y}) \mathbf{y}^T]) \mathbf{B} \quad (2.29)$$

A Eq. (2.29) define o termo de atualização da matriz de separação utilizado pelo algoritmo *Natural ICA*. O emprego do gradiente natural simplifica a maximização da verossimilhança e torna a matriz \mathbf{B} melhor condicionada [4]. Podemos verificar que o algoritmo converge para $\mathbf{g}(\mathbf{y}) \mathbf{y}^T = \mathbf{I}$. Ou seja, mostra a decorrelação de y_k e $g(y_l), k \neq l$ que é uma extensão não-linear da necessária decorrelação entre os sinais. Para esta dissertação a função *score* selecionada foi a função

$$g(y) = \frac{\text{sign}(y)}{\sigma} \quad (2.30)$$

onde σ é o desvio padrão da componente y .

2.6 Separação Cega de Fontes no Domínio da Frequência

No domínio do tempo, os sinais que chegam aos microfones são o resultado das convoluções dos sinais com as respostas ao impulso correspondentes aos caminhos percorridos por um impulso das posições das fontes até as posições dos microfones, devido às reverberações e reflexões em obstáculos presentes no ambiente. Uma forma de simplificar a operação de convolução é utilizar a transformada de Fourier, que transforma a convolução em multiplicação.

A essa abordagem chamamos de Separação Cega de Fontes no Domínio da Frequência ou FDBSS (do inglês *Frequency Domain Blind Source Separation*). Ao

passar para o domínio da frequência, o problema da separação da mistura convolutiva se torna vários problemas de misturas instantâneas e, com isso, surge a possibilidade de se usar algoritmos mais simples para resolver cada um deles.

Porém, ocorrem os problemas de permutação e escalonamento, mencionados anteriormente, que, no domínio da frequência, se tornam muito mais graves do que eram no domínio do tempo. Principalmente, deve ser dada atenção ao problema de permutação, pois, após o processamento, as componentes do mesmo sinal devem ser agrupadas a fim de se reestabelecerem os sinais desejados.

A Fig. 2.2 mostra os estágios do método FDBSS, descritos a seguir.

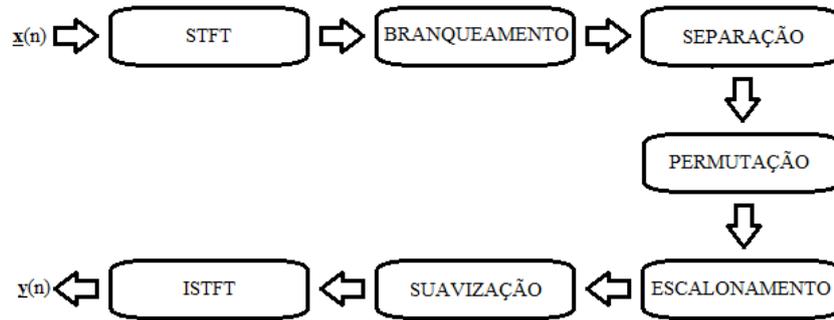


Figura 2.2: Estágios do FDBSS

2.6.1 Passando para o Domínio da Frequência

Para levar os sinais x_j , $j = 1, \dots, M$ do domínio do tempo para o domínio da frequência, utilizamos a STFT (do inglês *Short-Time Fourier Transform*), descrita na Eq. (2.31).

$$X_j(k, m) = \sum_n x_j(n) win_a(n - mJ) exp\left(\frac{-j2\pi kn}{K}\right), k = 0, \dots, K - 1 \quad (2.31)$$

onde $win_a(n)$ é a janela de amostragem de comprimento L e J é o salto da janela, que devem ser escolhidos para que não haja distorção do sinal após a transformação inversa. A janela normalmente utilizada é a janela de Hamming, definida como:

$$w_{Hamming}(n) = 0,5 \cos\left(1 - \cos\frac{2\pi n}{L}\right) \quad (2.32)$$

que atende à condição para reconstrução do sinal conhecida como COLA (do inglês *Constant Overlap-Add*) [1].

2.6.2 Branqueamento

Após passar os sinais para o domínio da frequência, é necessário realizar o branqueamento dos vetores das misturas em cada raia, o qual consiste em processar os sinais, tornando-os descorrelacionados, o que auxiliará a análise de componentes independentes [3]. O branqueamento também proporciona uma normalização das energias dos sinais, o que faz com que a convergência do algoritmo seja uniforme para cada uma das raias do domínio de frequência, tornando-a mais rápida e robusta.

Para isso, encontra-se uma matriz \mathbf{V} tal que $\mathbf{z} = \mathbf{V}\mathbf{x}$ e que:

$$E[\mathbf{z}\mathbf{z}^T] = E[\mathbf{V}\mathbf{x}(\mathbf{V}\mathbf{x})^T] = \mathbf{I} \quad (2.33)$$

ou seja:

$$\mathbf{V}E[\mathbf{x}\mathbf{x}^T]\mathbf{V}^T = \mathbf{V}\mathbf{E}_x\mathbf{V}^T = \mathbf{I}$$

Decompondo a matriz de autocorrelação de \mathbf{x} como $\mathbf{E}_x = \Sigma_x \Lambda_x \Sigma_x^T$, onde Σ_x é a sua matriz de autovetores e Λ_x é a sua matriz de autovalores, temos :

$$\mathbf{V}\mathbf{E}_x\mathbf{V}^T = \mathbf{V}\Sigma_x \Lambda_x \Sigma_x^T \mathbf{V}^T = \mathbf{I} \rightarrow \mathbf{V} = \Lambda_x^{-0,5} \Sigma_x^T \quad (2.34)$$

Ou seja, para o branqueamento do vetor de misturas \mathbf{x} , é necessário multiplicá-lo por $\Lambda_x^{-0,5} \Sigma_x^T$. Vale lembrar que a ordem em que os autovalores aparecem em Λ_x influencia na ordenação dos autovetores em Σ_x .

2.6.3 Permutação

Um dos grandes problemas na separação cega de fontes é a ambiguidade proporcionada pela permutação dos sinais no domínio da frequência. É mandatório que as frequências dos mesmos sinais estejam agrupadas ordenadamente a fim de se obter uma reconstrução correta dos mesmos no domínio do tempo.

Existem três formas clássicas de resolver o problema da permutação [1]:

- Usando informações sobre o sistema de misturas;
- Usando informações sobre o espectro de frequências dos sinais;
- Usando informações de tempo-frequência obtidas do algoritmo ICA multi-dimensional.

Nesta dissertação, utilizaremos o conceito de diferença entre tempos de chegada (TDOA, do inglês *Time Differential Of Arrival*), obtida através da primeira forma acima, a qual será descrita a seguir.

TDOA - *Time Difference Of Arrival*

O método de correção de permutação baseado em TDOA pertence ao grupo de algoritmos que utilizam as informações sobre o sistema de misturas. Ele se baseia na idéia do “modelo de campo próximo”, ou seja, assume que as fontes emissoras estão próximas das fontes receptoras. Esse conceito foi desenvolvido para ambientes não reverberantes mas funciona eficientemente nos ambientes reverberantes.

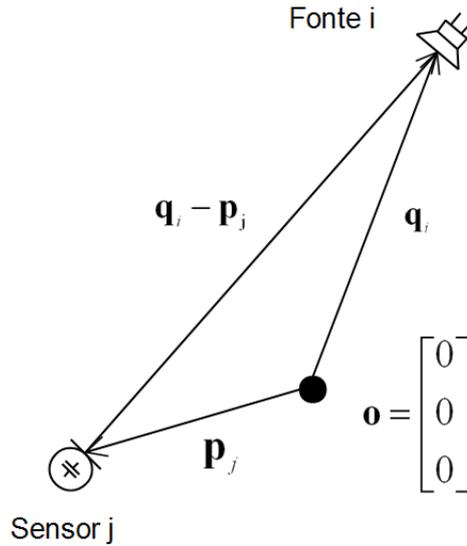


Figura 2.3: O Modelo do Campo Próximo

A Fig. 2.3 ilustra a idéia do campo próximo [1]. A partir de um vetor $\mathbf{o} = [0 \ 0 \ 0]^T$, o vetor origem, podemos definir \mathbf{q}_i , o vetor de posicionamento da i -ésima fonte, e \mathbf{p}_j , o vetor de posicionamento do j -ésimo microfone. Portanto a distância entre fonte e microfone é dada por $\|\mathbf{q}_i - \mathbf{p}_j\|$. Podemos também definir o TDOA. Sejam dois microfones quaisquer, (j, l) , o valor de TDOA em relação a i -ésima fonte pode ser obtido por:

$$\zeta_{jl}(i) = \tau_{ij} - \tau_{il} \quad (2.35)$$

Segundo [5], podemos modelar o sinal que chega no j -ésimo microfone a partir da i -ésima fonte como:

$$x_j(t) = \alpha_{ji} s_i(t - \tau_{ji}) \quad (2.36)$$

onde τ_{ji} é o atraso entre a fonte e o microfone, e α_{ji} é a atenuação do percurso. No domínio da frequência, a Eq. (2.36) pode ser escrita como:

$$x_j(f) = \alpha_{ji} e^{-j2\pi f \tau_{ji}} s_i(f) \rightarrow x_j(f) = h_{ji}(f) s_i(f) \quad (2.37)$$

Considerando que a atenuação seja inversamente proporcional à distância entre o

microfone e a fonte, e que o atraso é também proporcional à mesma medida, podemos reescrever $h_{ji}(f)$ da Eq. (2.37) como:

$$h_{ji}(f) = \frac{1}{\|\mathbf{q}_i - \mathbf{p}_j\|} e^{j2\pi f c^{-1}(\|\mathbf{q}_i - \mathbf{p}_j\|)} \quad (2.38)$$

onde c é a constante de velocidade de propagação do som no meio. Para o funcionamento deste algoritmo, é necessária a informação de mais de um sensor e a utilização destas em conjunto. Considerando um outro sensor, j' , podemos escrever a razão:

$$\begin{aligned} \frac{h_{ji}(f_k)}{h_{j'i}(f_k)} &= \frac{\|\mathbf{q}_i - \mathbf{p}_{j'}\|}{\|\mathbf{q}_i - \mathbf{p}_j\|} e^{j2\pi f_k c^{-1}(\|\mathbf{q}_i - \mathbf{p}_j\| - \|\mathbf{q}_i - \mathbf{p}_{j'}\|)} \\ \frac{h_{ji}(f_k)}{h_{j'i}(f_k)} &= \frac{\|\mathbf{q}_i - \mathbf{p}_{j'}\|}{\|\mathbf{q}_i - \mathbf{p}_j\|} e^{j2\pi f_k \zeta_{jj'}(i)} \end{aligned} \quad (2.39)$$

onde substituímos f por f_k pois estamos trabalhando com frequências discretas. O valor de $\zeta_{jj'}(i)$ é o TDOA da i -ésima fonte para o par de microfones j e j' . Vale lembrar que, segundo o critério de Nyquist [1], para evitar *aliasing* no domínio da frequência, é necessário que os sinais possuam banda máxima de $1/2T_s$, onde T_s é o período de amostragem dos sinais.

Para um arranjo de M microfones, obtemos $M(M-1)/2$ combinações. No entanto, somente $M-1$ destas combinações são distintas. Por isso, reescreveremos o valor de TDOA em função de um microfone de referência J , conforme a Eq. (2.40)

$$\zeta_j(i) = \tau_{ji} - \tau_{Ji} \quad (2.40)$$

Como $\frac{h_{ji}(f_k)}{h_{j'i}(f_k)}$ é uma razão de números complexos, podemos escrevê-la em notação polar:

$$\frac{h_{ji}(f_k)}{h_{j'i}(f_k)} = \left\| \frac{h_{ji}(f_k)}{h_{j'i}(f_k)} \right\| e^{j \arg\left(\frac{h_{ji}(f_k)}{h_{j'i}(f_k)}\right)}$$

onde a função $\arg(\cdot)$ é o argumento do número complexo. Substituindo a expressão acima em (2.39) obtemos o valor de TDOA:

$$\zeta_j(i) = \frac{\arg\left(\frac{h_{ji}(f_k)}{h_{Ji}(f_k)}\right)}{2\pi f_k} \quad (2.41)$$

Sendo d_{max} a maior distância entre dois microfones, esse algoritmo é válido para frequências tais que:

$$f \leq \frac{c}{2d_{max}} \quad (2.42)$$

Definimos, então, um vetor que possui todos os TDOAs de uma i -ésima fonte

para uma determinada frequência:

$$\zeta_i(k) = [\zeta_1(f_k) \dots \zeta_{M-1}(f_k)], i = 1, \dots, N; k = 1, \dots, K \quad (2.43)$$

Utilizando os $N \times K$ vetores $\zeta_i(k)$, podemos encontrar as posições das fontes. Uma vez que o problema de ICA esteja resolvido, os vetores $\zeta_i(k)$ formarão N *clusters*. Os N centróides desses *clusters* definem a localização dessas fontes. A Fig. 2.4 [1] exemplifica o método de clusterização para o caso de três fontes.

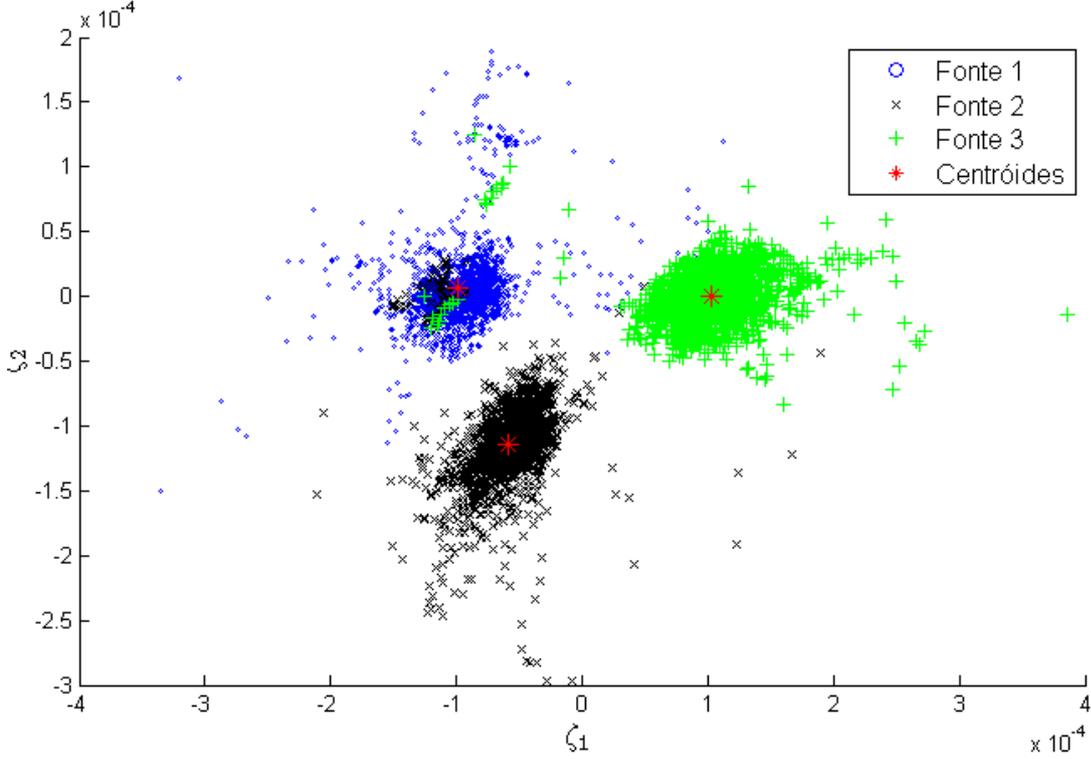


Figura 2.4: Exemplo de *clusters* para localização de fontes

Para a solução desse problema podemos utilizar algoritmos clássicos como o *K-means* [6]. A solução desse algoritmo fornecerá os N centróides $\mathbf{c}_\zeta(i)$. Para resolver o problema da permutação, definimos a matriz:

$$DIST_k = \begin{bmatrix} \|\zeta_1(k) - \mathbf{c}_\zeta(1)\|^2 & \dots & \|\zeta_1(k) - \mathbf{c}_\zeta(N)\|^2 \\ \dots & \ddots & \dots \\ \|\zeta_N(k) - \mathbf{c}_\zeta(1)\|^2 & \dots & \|\zeta_N(k) - \mathbf{c}_\zeta(N)\|^2 \end{bmatrix}, k = 1, \dots, K \quad (2.44)$$

Nessa matriz, cada linha corresponde a uma fonte permutada, enquanto que cada coluna corresponde a um centróide (uma fonte real). Cada elemento desta matriz é a distância de uma fonte permutada para uma fonte real. Agora, para cada uma

das k frequências, devemos encontrar a matriz de permutação \mathbf{P}_k .

Segundo [4], uma das formas de resolver este problema é verificar qual o elemento de menor valor de $DIST_k$. Atribuímos a esta linha a fonte real de sua coluna. Após essa atribuição, descartamos a linha e coluna selecionadas da matriz $DIST_k$ e buscamos o novo menor valor atribuindo esse novo valor a uma nova fonte. Faz-se esse procedimento até obtermos a matriz de permutação \mathbf{P}_k . Abaixo, vemos um exemplo desse procedimento:

$$DIST_k = \begin{bmatrix} 0.4 & 0.1 & 0.8 \\ 0.9 & 0.5 & 0.2 \\ 0.3 & 0.7 & 0.6 \end{bmatrix} \rightarrow \mathbf{P}_k = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} \quad (2.45)$$

2.6.4 Escalonamento

O escalonamento, conforme comentado anteriormente, visa encontrar uma matriz Λ (no caso Λ_k , pois será uma para cada raia no domínio da frequência), para que as fontes recuperadas estejam consistentes com as originais. Para esta dissertação, será utilizado o algoritmo MDP (do inglês *Minimum Distortion Principle*) [7].

Minimum Distortion Principle

Sejam \mathbf{T} , uma matriz que indica a distorção aceitável, escolhida de forma aleatória, e \mathbf{H}_k , a matriz mistura na frequência de índice k . Idealmente, Λ_k é dada por

$$\Lambda_k = \text{diag}(\mathbf{T}\mathbf{H}_k) \quad (2.46)$$

Admitimos que o trabalho do ICA já foi realizado e que o problema de permutação também está resolvido. Definimos $\mathbf{W}_{P_k}^{-1}$ como a matriz responsável por estas soluções. Então, podemos dizer que $\mathbf{W}_{P_k}^{-1} = \mathbf{H}_k$. Admitindo $\mathbf{T} = \mathbf{I}$ a fim de simplificar o problema, podemos reescrever a Eq. (2.46) como:

$$\Lambda_k = \text{diag}(\mathbf{W}_{P_k}^{-1}) \quad (2.47)$$

Utilizando o modelo de misturas instantâneas, vide Eq. (2.3), para cada raia de frequência e considerando que $\mathbf{W}_{P_k} = \mathbf{P}_k \mathbf{W}_k$, temos:

$$\mathbf{y}_k = \Lambda_k \mathbf{P}_k \mathbf{W}_k \mathbf{H}_k \mathbf{s}_k \mathbf{y}_k = \text{diag}(\mathbf{W}_{P_k}^{-1}) \mathbf{W}_{P_k} \mathbf{H}_k \mathbf{s}_k \mathbf{y}_k = \text{diag}(\mathbf{W}_{P_k}^{-1} \mathbf{W}_{P_k} \mathbf{H}_k) \mathbf{s}_k$$

e, portanto,

$$\mathbf{y}_k = \text{diag}(\mathbf{H}_k) \mathbf{s}_k \quad (2.48)$$

Com esta abordagem, para cada raia, temos $y_{ik} = h_{ii} s_{ik}$. Ou seja, o sinal recuperado y_i é uma versão filtrada da fonte original s_i . Podemos, então, perceber

que esse algoritmo resolve o escalonamento em cada raia de frequência, mas não o escalonamento global. Se uma das fontes originais estiver com uma intensidade muito maior que as demais, isso vai permanecer nos sinais recuperados.

2.6.5 Suavização

Antes de retornar para o domínio do tempo, pode ser necessário realizar a suavização dos sinais. Isso se deve à característica de circularidade inerente à DFT. O produto de dois sinais no domínio da frequência equivale a uma convolução circular no domínio do tempo. Para que se torne uma convolução linear de fato, o número de raias de frequência deve ser, pelo menos, igual à soma dos comprimentos das janelas de observação dos dois sinais. Caso não seja atendido, o sinal obtido, após usarmos a IDFT, será uma versão distorcida do sinal desejado.

No caso da separação cega de fontes, o filtro de misturas idealmente tem comprimento infinito. Isso faz com que ocorram distorções dos sinais resultantes no domínio do tempo [1]. Para melhorar essa resposta, [8] propõe o janelamento dos sinais no domínio do tempo. Com isso podemos controlar a resposta em frequência dos filtros obtidos, observando o efeito da janela no domínio da frequência.

Neste trabalho, utilizamos a janela de Hamming, definida em (2.32). A Eq. (2.49) mostra o espectro de frequência desta janela.

$$W_{\text{Hamming}}(k) = 0,25\delta(k-1) + 0,5\delta(k) + 0,25\delta(k+1) \quad (2.49)$$

Recalculando cada elemento da matriz separadora após a aplicação da janela, temos:

$$\mathbf{W}_{ij}(k) = 0,25\mathbf{W}_{ij}(k-1) + 0,5\mathbf{W}_{ij}(k) + 0,25\mathbf{W}_{ij}(k+1) \quad (2.50)$$

Como podemos observar na Eq. (2.50), ao tentarmos suavizar a resposta dos filtros, estamos alterando a matriz que é a solução da análise de componentes independentes. Portanto, é necessário observar o quanto o efeito da circularidade prejudica a qualidade dos sinais recuperados.

Utilizando essa técnica espera-se uma melhora nos valores da razão entre o sinal e artefatos (SAR), descrito na Seção 5.3, pois os artefatos devido à circularidade diminuem, enquanto que esperamos uma degradação nos valores da razão sinal interferência (SIR), também mencionado na Seção 5.3, pois a matriz solução da análise de componentes independentes foi alterada.

2.6.6 Voltando para o Domínio do Tempo

Uma vez resolvidos os problemas mencionados nas seções anteriores, devemos levar as componentes separadas na frequência, $Y(m, k)$ de volta para o domínio do tempo. Para isso, utilizamos a ISTFT, que é a STFT inversa, dada por:

$$y(n) = \sum_m \left\{ \sum_k Y(m, k) \exp\left(\frac{j2\pi kn}{K}\right) \right\} \quad (2.51)$$

$$y(n) = \sum_m \left\{ \text{win}_s(n - mJ) \sum_k Y(m, k) \exp\left(\frac{j2\pi kn}{K}\right) \right\} \quad (2.52)$$

onde $\text{win}_s(n)$ é a janela de síntese de comprimento L e J é o salto da janela, que, da mesma forma que na STFT, devem ser escolhidos para que não haja distorção dos sinais recuperados no domínio do tempo. A janela normalmente utilizada é a janela retangular, que também atende ao COLA.

Capítulo 3

Bancos de Filtros

Os bancos de filtros fazem parte dos sistemas chamados multitaxas. Neste capítulo apresentaremos, inicialmente, dois componentes importantes para essas estruturas: o decimador e o interpolador. Em seguida descreveremos alguns tipos de bancos de filtros, dando ênfase a dois tipos que serão utilizados no pré-processamento proposto nesta dissertação: o banco de filtros modulados por cossenos e o banco de filtros não-uniforme com decomposição diádica.

3.1 Sistemas Multitaxas

Normalmente, um sinal é processado de forma sequencial. Ou seja, o valor do sinal atual, juntamente com os valores de suas componentes atrasadas, é processado e, para cada nova amostra, o processo precisa ser repetido. Uma alternativa para isso é transformar o sinal em diversas componentes e processá-las paralelamente. Com isso, um bloco de novas amostras será necessário para que o processo seja repetido. Para isso é utilizada uma estrutura conhecida como estrutura em sub-bandas baseada em banco de filtros. Essa estrutura permite, também, que separemos o sinal em bandas distintas e realizemos a separação cega de fontes em cada uma delas distintamente.

Nessa parte serão mostrados dois elementos importantes para realizar a paralelização/serialização do sinal, o decimador e o interpolador, assim como suas propriedades. Também serão apresentadas diferentes implementações de bancos de filtros: o modulado por cosseno (*Cosine Modulated Filter Bank*) e o banco de filtros não-uniforme com estrutura diádica.

3.1.1 Decimador

O decimador, ilustrado na Fig. 3.1, é um bloco (operador) utilizado no processamento multitaxas, cuja relação entrada-saída é dada por:

$$y_d(n) = x(nM) \quad (3.1)$$

onde M é um valor inteiro. Esse operador só permite passar valores de $x(n)$ quando n é um múltiplo de M . Com isso o tempo (período) de apresentação das amostras do sinal fica M vezes maior em relação ao original, permitindo o uso de circuitos/aparelhos que operem em uma frequência M vezes menor do que a necessária para o processamento. Por isso, o decimador também é conhecido como *downsampler*.

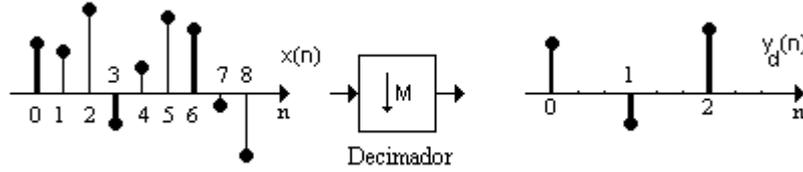


Figura 3.1: Decimação do sinal $x(n)$ com $M = 3$

A relação entrada-saída do decimador no domínio da frequência é dada por[9]:

$$Y_d(e^{j\omega}) = \frac{1}{M} \sum_{l=0}^{M-1} X(e^{j(\frac{\omega_l \pi}{M})}) \quad (3.2)$$

Portanto, o espectro de frequência do sinal decimado $y_d(n)$, $Y_d(e^{j\omega})$, é uma versão expandida por um fator M do espectro de $x(n)$, cujo período é igual a $2\pi/M$. A Fig. 3.2 mostra o efeito do decimador num sinal cujo espectro é dado por:

$$X(e^{j\omega}) = \begin{cases} A, & \text{se } |\omega| \leq \omega_o/M \\ 0, & \text{se } \omega_o/M < |\omega| < \pi \end{cases} \quad (3.3)$$

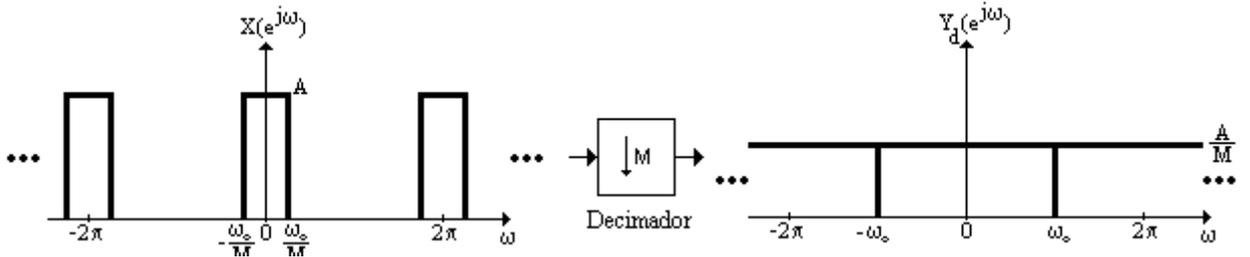


Figura 3.2: Efeito do decimador na resposta em frequência de $x(n)$

3.1.2 Interpolador

A relação entrada-saída do bloco interpolador é definida como:

$$y_i(n) = \begin{cases} x(n/L), & n \text{ múltiplo inteiro de } L \\ 0, & \text{outros valores de } n \text{ inteiro} \end{cases} \quad (3.4)$$

O interpolador de ordem L apresenta na sua saída uma amostra do sinal de entrada $x(n)$, seguida de $L - 1$ amostras com valor zero. Isso faz com que a amostra de $x(n)$ na saída do interpolador dure um período L vezes menor, daí o fato do interpolador ser também conhecido como *upsampler*, aumentando assim a frequência de amostragem do sinal. Isso pode ser verificado na Fig. 3.3 para $L = 4$.

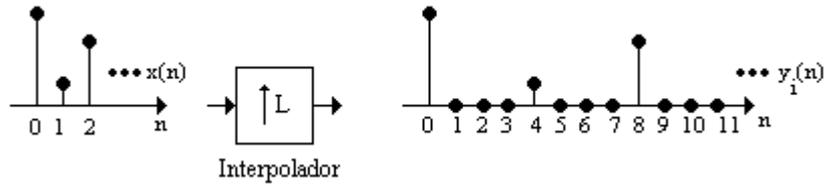


Figura 3.3: Interpolação do sinal $x(n)$ com $L = 4$

O espectro de frequência da saída do interpolador, $Y_i(e^{j\omega})$, é dado por [9]:

$$Y_i(e^{j\omega}) = X(e^{j\omega L}) \quad (3.5)$$

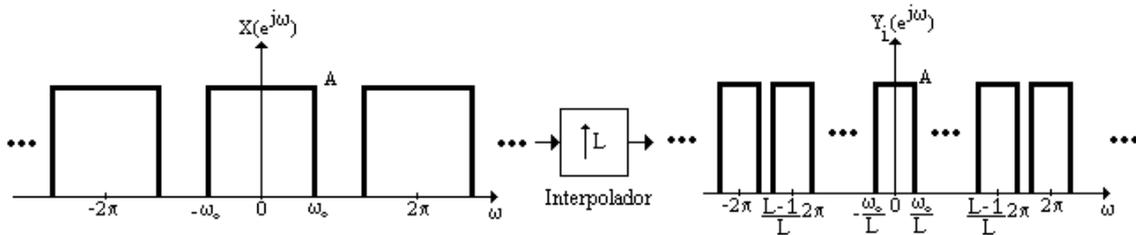


Figura 3.4: Efeito do interpolador no espectro de $x(n)$

3.2 Banco de Filtros

O papel de um banco de filtros consiste em gerar componentes de um sinal em diferentes bandas de frequência, ou seja, dividir o espectro de frequência de um sinal.

Um banco de M filtros é geralmente composto por um filtro passa-baixas, um filtro passa-alta e $M - 2$ filtros passa-banda com larguras de banda de passagem idênticas. Idealmente, eles deveriam ter uma resposta plana e cobrir todo o espectro de frequência do sinal [10]. Essas características podem ser vistas na Fig. 3.5 para um sistema com 3 sub-bandas.

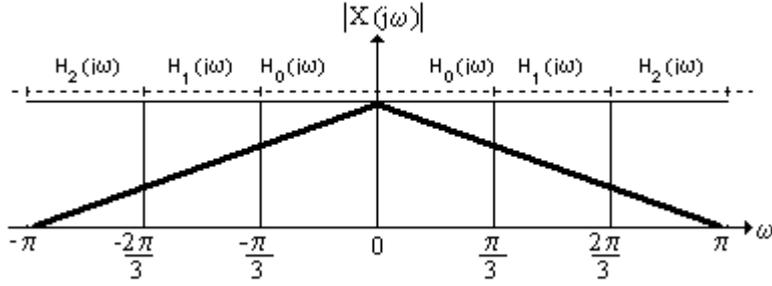


Figura 3.5: Espectro de freqüência dividido em subbandas

Também podemos verificar que a saída de cada filtro do banco ocupa uma faixa de frequência M vezes menor do que o sinal original. Por isso, podemos decimar os sinais saídos de cada filtro do banco por um fator $L \leq M$. Sem a decimação desses sinais, o número total de amostras geradas pelo banco de filtros seria M vezes maior que o total do sinal original, o que aumentaria o custo computacional do processamento.

Para $L = M$, o sistema é dito criticamente decimado. Nesse sistema, o número total de amostras provenientes dos decimadores, que estão cascateados com os filtros do banco, será igual ao total de amostras do sinal original. No caso de $L > M$, não será possível em geral restaurar o sinal original, pois haverá perda de informação.

As implementações dos bancos de filtros de análise (que decompõem o sinal original em M componentes) e de síntese (que restauram o sinal original a partir das suas M componentes em subbandas) estão mostradas na Fig. 3.6.

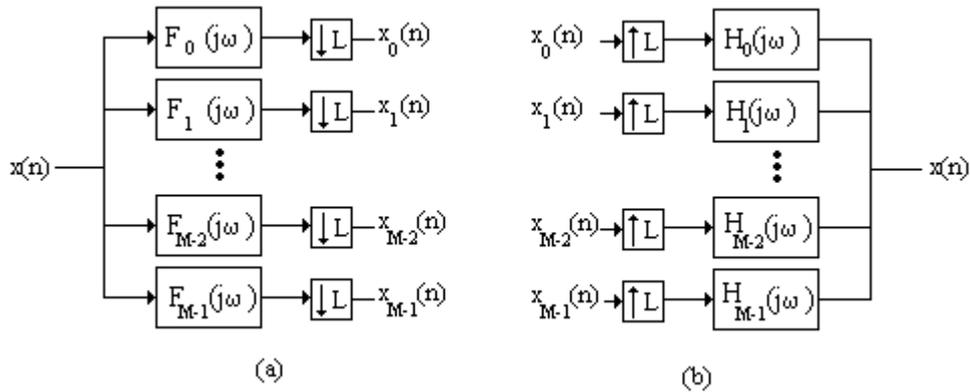


Figura 3.6: Bancos de filtros de (a) análise e (b) síntese

3.2.1 Banco de Filtros Modulados Por Cosseno

Na técnica de modulação por cossenos, geram-se $2M$ filtros que, depois de serem multiplicados por exponenciais complexas, serão combinados aos pares, formando o banco de filtros com M subbandas.

Assim como no banco de filtros baseado em DFT [9], cada filtro de subbanda do banco, $P_k(e^{j\omega})$, $0 \leq k \leq 2M - 1$, se relaciona a um filtro protótipo, $P_0(e^{j\omega})$, da seguinte forma:

$$P_k(e^{j\omega}) = P_0(e^{j(\omega - \frac{k\pi}{M})}) \quad (3.6)$$

O filtro protótipo é um filtro FIR cuja resposta em frequência é simétrica em relação a $\omega = 0$, é formado por coeficientes reais e é um passa-baixas de frequência de corte $\omega_p = \pi/2M$. A representação no domínio da frequência do filtro protótipo e dos $2M$ filtros das sub-bandas do banco de filtros baseado em DFT estão na Fig. 3.7.

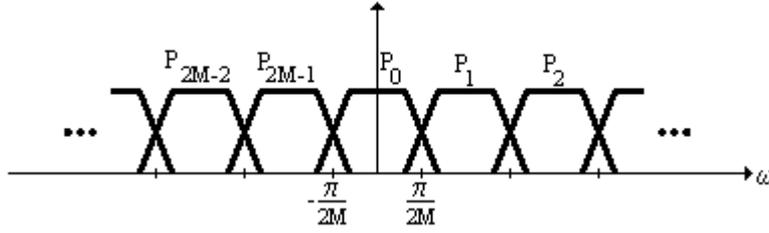


Figura 3.7: Filtros formadores do banco de filtros modulado por cosseno

Para formar o banco de M filtros modulados por cosseno bastaria combinar os filtros $P_k(e^{j\omega})$ e $P_{2M-k}(e^{j\omega})$. Porém isso faria com que as sub-bandas do banco não tivessem faixas de passagem de mesma largura. Para resolver esse problema, deslocamos a resposta em frequência do filtro protótipo da Fig. 3.7 de $\pi/2M$. Então definimos um novo conjunto de filtros $S_k(e^{j\omega})$, $0 \leq k \leq 2M - 1$ como:

$$S_k(e^{j\omega}) = P_k(e^{j(\omega - \frac{\pi}{2M})}) \quad (3.7)$$

Desta forma, as frequências centrais dos filtros $S_k(e^{j\omega})$ e $S_{2M-1-k}(e^{j\omega})$, $0 \leq k \leq M - 1$, para $-\pi \leq \omega < \pi$, têm a mesma distância em relação a $\omega = 0$; portanto, são filtros conjugados. Definindo as funções:

$$U_k(e^{j\omega}) = c_k P_0(e^{j\omega} e^{j\frac{(2k+1)\pi}{2M}}) \quad (3.8)$$

$$V_k(e^{j\omega}) = c_k^* P_0(e^{j\omega} e^{-j\frac{(2k+1)\pi}{2M}}) \quad (3.9)$$

o k -ésimo filtro de análise, $H_k(z)$, é definido como:

$$H_k(e^{j\omega}) = a_k U_k(e^{j\omega}) + a_k^* V_k(e^{j\omega}) \quad (3.10)$$

Para coincidirem com as bandas passantes dos bancos de filtros de análise e de síntese, os filtros $F_k(e^{j\omega})$ são descritos como:

$$F_k(e^{j\omega}) = b_k U_k(e^{j\omega}) + b_k^* V_k(e^{j\omega}) \quad (3.11)$$

O propósito de se introduzirem as constantes a_k e b_k é cancelar a sobreposição entre sub-bandas vizinhas nos bancos de análise e síntese. Também, visa-se eliminar a distorção de fase, promovida pela ação dos dois bancos de filtros, fazendo com que ela seja linear. A função da constante c_k é propiciar aos filtros $U_k(z)$ e $V_k(z)$ a mesma característica linear de fase do filtro protótipo $P_o(z)$. Considerando os fatores mencionados, os valores de a_k , b_k e c_k ficam [11]:

$$a_k = b_k^* = e^{j\theta_k} \quad (3.12)$$

$$c_k = e^{-j\frac{\pi(k+0,5)N}{2M}} \quad (3.13)$$

sendo

$$\theta_k = (-1)^k \pi/4 \quad (3.14)$$

Com isso, podemos achar as equações dos filtros que compõem os bancos de síntese e de análise. Calculando a primeira parcela da equação (3.10), utilizando as equações (3.12) e (3.13), obtemos:

$$a_k U_k(e^{j\omega}) = e^{j\theta_k} e^{-j\frac{\pi(k+0,5)N}{2M}} \sum_{n=0}^N p_0(n) e^{-j\frac{(2k+1)\pi}{2M}n} e^{-j\omega n} \quad (3.15)$$

Fazendo o mesmo com a segunda parcela de (3.10):

$$a_k^* V_k(e^{j\omega}) = e^{-j\theta_k} e^{j\frac{\pi(k+0,5)N}{2M}} \sum_{n=0}^N p_0(n) e^{j\frac{(2k+1)\pi}{2M}n} e^{-j\omega n} \quad (3.16)$$

Portanto, obtemos a seguinte resposta em frequência para o k -ésimo filtro de análise:

$$H_k(e^{j\omega}) = \sum_{n=0}^N 2p_0(n) \cos\left(\frac{\pi(k+0,5)(n+0,5N)}{M} + \theta_k\right) e^{-j\omega n} \quad (3.17)$$

ou seja, sua resposta ao impulso, $h_k(n)$, é dada por:

$$h_k(n) = 2p_0(n) \cos\left(\frac{\pi(k+0,5)(n+0,5N)}{M} + \theta_k\right) \quad (3.18)$$

Analogamente, para encontrarmos os filtros de síntese, basta verificar que b_k é complexo conjugado de a_k e, portanto:

$$f_k(n) = 2p_0(n) \cos\left(\frac{\pi(k+0,5)(n+0,5N)}{M} - \theta_k\right) \quad (3.19)$$

A resposta em frequência do banco de filtros de análise pode se vista na Fig. 3.8.

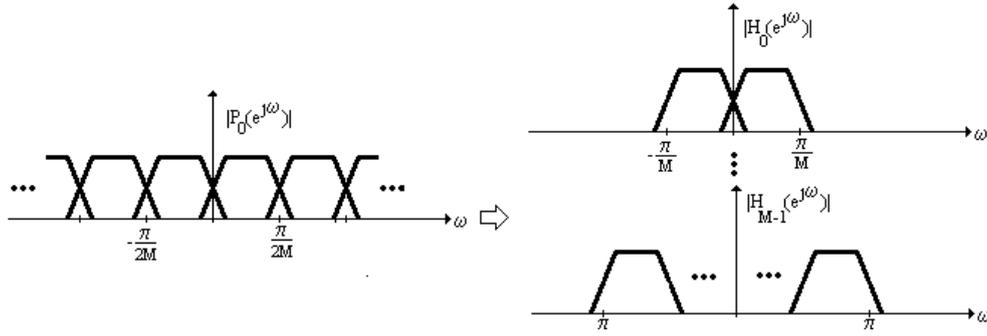


Figura 3.8: Banco de Filtros Modulados por Cossenos

Uma vez conhecidas as equações dos filtros, verificamos que um banco de filtros com boa seletividade depende de um bom projeto do filtro protótipo $p_0(n)$. Mesmo com a resolução dos problemas de sobreposição de filtros adjacentes e de fase não-linear através das constantes a_k , b_k e c_k , existem outros fatores que prejudicam o desempenho dos bancos. Um deles é a distorção que o sinal sofre durante a sua filtragem pelos bancos de análise e de síntese. É necessário que $p_0(n)$ seja projetado de tal forma que a distorção provocada pelos bancos de filtros seja a menor possível. Um dos motivos é a existência da banda de transição após a frequência de corte do filtro. No encontro das respostas em frequência de dois filtros vizinhos, uma região de ondulação aparece nessa região de transição. Outro motivo é que não existe filtro cuja banda passante seja totalmente plana. Uma tentativa de fazer o projeto do filtro protótipo é o método de janelamento [12]. Porém esse janelamento promove as perturbações na magnitude da frequência do filtro. Um modo de reduzir a distorção no sinal introduzida pelos bancos de filtros é minimizar a função custo dada por [11]:

$$\varphi = \alpha \left[\int_0^{\pi/M} \left(|P_0(e^{j\omega})|^2 + |P_0(e^{j(\omega-\pi/M)})|^2 - 1 \right)^2 d\omega \right] + (1-\alpha) \left[\int_{\delta+\pi/2M}^{\pi} \left(|P_0(e^{j\omega})|^2 \right) d\omega \right] \quad (3.20)$$

onde $0 < \alpha < 1$ e δ é uma constante que define a largura da faixa de transição do filtro protótipo.

3.2.2 Banco de Filtros Não-Uniforme com Decomposição Diádica

Um banco de filtros não-uniforme é uma estrutura que permite obter uma decomposição com maior detalhamento em partes do espectro do sinal. Normalmente é

usado quando uma determinada faixa de frequências contém informações mais importantes que as demais. Podemos citar como exemplo a audição humana, onde a sua resolução decai com o aumento da frequência. Outro exemplo são os algoritmos de compressão de imagens, que transmitem as bandas que contêm maior informação, em detrimento de outras que apresentam menos interferência na resolução da imagem.

Uma estrutura bastante conhecida para se obter um banco não-uniforme é o arranjo em árvore utilizando-se de banco de filtros de 2 canais, decimadores e interpoladores. Essa estrutura baseia-se nas propriedades, conhecidas como identidades nobres, dos decimadores e dos interpoladores [9], mostradas na Fig. 3.9.

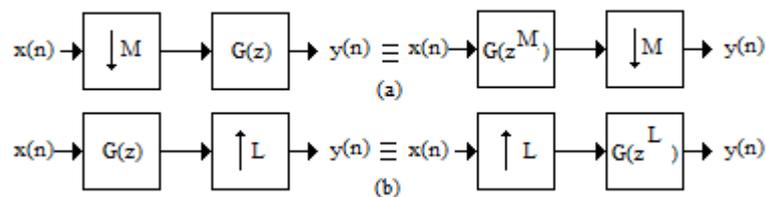


Figura 3.9: Identidades Nobres do (a) Decimador e (b) Interpolador

Essas propriedades mostram que podemos fazer com que a frequência de corte de um filtro $G(z)$, ω_{pG} , se reduza à frequência ω_{pG}/M , invertendo-se a ordem de decimadores ou interpoladores e dos filtros. Com esses conceitos, são propostas estruturas de bancos de filtros de análise e síntese, ilustradas na Fig. 3.10 para 4 bandas não-uniformes.

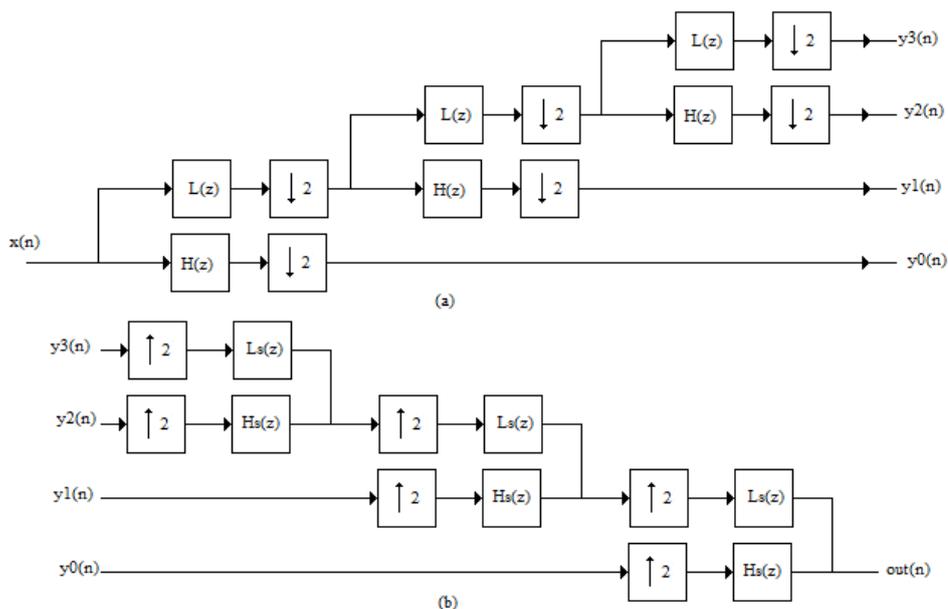


Figura 3.10: Estruturas de banco de filtros não uniforme de (a) análise e (b) síntese

Podemos representar os bancos da Fig. 3.10 pelas estruturas equivalentes da Fig. 3.11.

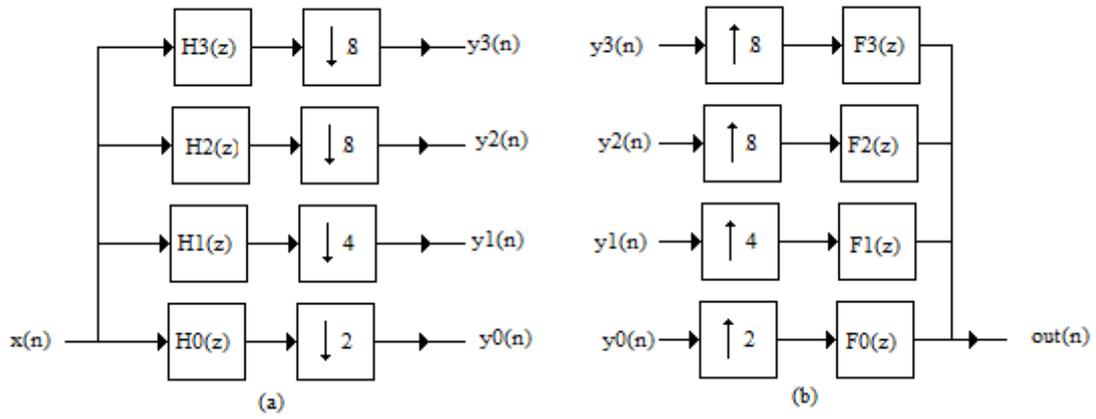


Figura 3.11: Estruturas equivalentes dos bancos de (a) análise e (b) síntese

Se $L(z)$ e $H(z)$ formarem um par de filtros passa-baixas e passa-altas, respectivamente, com frequência de corte em torno de $\pi/2$, e tais que satisfaçam as propriedades de reconstrução perfeita dadas em [9], temos $out(n) = x(n - \Delta)$, sendo Δ um atraso introduzido pelo processamento. Respostas em frequências ilustrativas dos filtros de análise e síntese dos bancos não-uniformes de 4 canais podem ser vistas na Fig. 3.12.

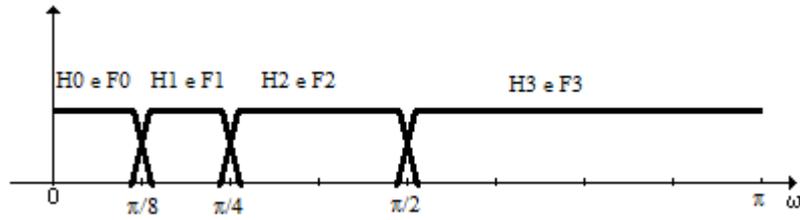


Figura 3.12: Respostas em frequência de bancos de filtros de análise e síntese não-uniformes

Capítulo 4

Separação Cega de Fontes em Subbandas

Conforme observado no Capítulo 2, o fator motivante para fazermos a separação cega no domínio da frequência (FDBSS) é o fato de a convolução de um sinal com a resposta ao impulso de um ambiente reverberante tornar-se, nesse domínio, o produto das transformadas de Fourier dos mesmos. Os sinais das misturas que chegam aos microfones são “janelados” e os sinais obtidos com esse janelamento são levados para o domínio da frequência, através do cálculo das suas FFTs, onde é aplicado o algoritmo de separação em cada uma das raia. O tamanho da janela utilizada é normalmente determinado empiricamente ou fixado em um valor entre 128 e 2048. Pode-se também utilizar uma estimativa do comprimento da resposta ao impulso do sistema fonte-microfone para se escolher o tamanho da janela. No entanto, conforme discutido em [13], há uma limitação no aumento do tamanho da janela, devido à quantidade de dados necessários para se fazer as estimativas das estatísticas utilizadas no método de separação em cada raia. O compromisso na escolha do tamanho da janela, de tal forma a considerar toda a parte significativa da resposta ao impulso do sistema fonte-microfone e ao mesmo tempo manter uma quantidade de dados apropriada para as estimativas das estatísticas, é apontado em [13] como a causa do desempenho ruim dos métodos de separação cega de fontes em ambientes com reverberações longas.

Neste trabalho, consideramos um outro fator limitante para o aumento do tamanho da janela: a não-estacionariedade do sinal de voz. A medida em que o tamanho da janela é aumentado, a premissa de que as estatísticas utilizadas pelo método de separação podem ser aproximadas considerando estacionariedade dos sinais naquele trecho torna-se menos verdadeira. Com isso, há uma piora no desempenho do algoritmo. Para atenuar esse problema, propomos a utilização de janelas de diferentes tamanhos em diferentes faixas de frequência. Essa proposta se baseia em dois fatos: o primeiro é que as componentes de baixas frequências de um sinal de voz se mantêm

“estacionárias” por um tempo maior dos que as de altas frequências; o segundo é que o tempo de reverberação diminui com a frequência em ambientes reais. Este segundo argumento será discutido na próxima seção. Na Seção 4.2 apresentaremos o método de separação cega proposto para melhorar os resultados do algoritmo FDBSS em ambientes reverberantes.

4.1 Características da Reverberação em Salas

No final do século XIX, W. C. Sabine estudou os efeitos da reverberação em ambientes fechados. Ele mediu, em diversos ambientes, o intervalo de tempo em que um som é interrompido, considerando-se o limiar de audibilidade aproximadamente igual a 60 dB (ou seja, a pressão sonora atinge um milionésimo do seu valor inicial), daí utilizar-se como símbolo de tempo de reverberação T_{60} . Empiricamente chegou-se a equação conhecida como Equação de Sabine:

$$T_{60} = \frac{4 \ln(10^6) V}{c S \alpha} \quad (4.1)$$

onde c é a velocidade de propagação do som no ar, V é o volume do ambiente, S , sua área interna total e α é o valor médio da constante de absorção do ambiente. Como esperado, o tempo de reverberação depende do tamanho da sala e da absorção das ondas sonoras pelas paredes e móveis. Forte absorção encurta o tempo de reverberação, enquanto fraca absorção tende a prolongá-lo. Porém, a observação mais importante para o nosso trabalho é que o coeficiente de absorção, que depende dos materiais dos diversos obstáculos que refletirão as ondas sonoras, não será igual para todas as frequências. O mecanismo de absorção sonora se baseia na existência de poros e interstícios nos materiais. A eficiência dos materiais absorventes sonoros se situa nas altas frequências. Aumentando-se a sua espessura, melhora-se o seu comportamento nas frequências mais baixas. Atualmente são desenvolvidos simuladores da propagação de sons em salas considerando suas geometrias e materiais.

Em [14], são descritos aspectos práticos da reverberação em ambientes fechados. A Fig. 4.1 apresenta um gráfico que ilustra valores típicos do tempo de reverberação em função da frequência para salas “reais”. As curvas traçadas em linha contínua indicam os limites superior e inferior do tempo de reverberação de um simulador “ideal” em todo o espectro audível. Salas típicas absorvem altas frequências muito mais do que as baixas. Isso faz com que os tempos de reverberação nessas salas sejam mais longos na faixa de baixas frequências do que nas de médias e altas frequências. A linha pontilhada mostra a relação tempo de reverberação e frequência de uma sala simulada que seria percebida como natural. Uma vez que as pessoas tendem a esperar esse comportamento de reverberação típico nos sons que estão

ouvindo, simulações de salas que se afastem deste padrão dependente da frequência são sentidas como não-naturais.

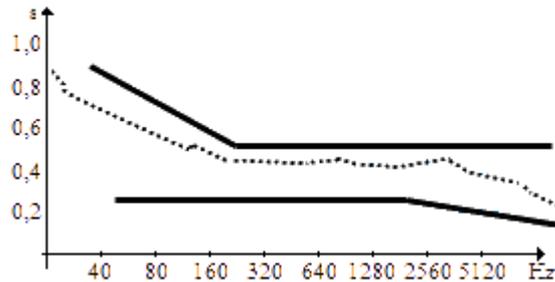


Figura 4.1: Características de tempo de reverberação e frequência de um ambiente fechado

4.2 Separação de Fontes por Subbanda

Com base na característica do tempo de reverberação em função da frequência descrita na seção anterior, propomos neste trabalho decompor os sinais das misturas em diferentes faixas de frequência e, em seguida, aplicar o método de separação cega de fontes no domínio da frequência em cada subbanda, separadamente.

Podem-se então utilizar diferentes tamanhos de janelas nas diferentes subbandas, explorando assim as características de reverberação com a frequência. Como, em ambientes naturais, o tempo de reverberação é maior em baixas frequências, espera-se que seja necessário utilizar janelas maiores nessa faixa de forma a modelar adequadamente o efeito da reverberação. Já nas mais altas frequências, podem-se utilizar janelas menores, reduzindo a complexidade computacional do algoritmo de separação e melhorando o seu desempenho.

Outro fator que nos estimula a utilizar janelas de tamanhos distintos vem da não-estacionariedade dos sinais de voz. Essa característica pode dificultar a convergência do algoritmo de separação de fontes, já que esse se baseia em estimativas de estatísticas das misturas em cada raia de frequência, sendo essas dependentes do tamanho da janela empregada. Ao utilizarmos uma janela muito longa, aumentamos o intervalo de observação e a chance de que não-estacionariedades ocorram naquelas amostras. Em altas frequências, a não estacionariedade dos sinais é mais acentuada, levando à conclusão de que janelas de tamanhos menores são mais apropriadas para essa faixa de frequências. Já as componentes de baixas frequências dos sinais naturalmente variam mais lentamente, sendo possível a utilização de janelas mais longas. O uso de diferentes tamanhos de janelas nas diferentes faixas de

frequência permite que o algoritmo de separação em cada raia processe sinais “mais estacionários”, melhorando a sua eficiência, principalmente em ambientes com alta reverberação.

Esses fatores nos levam a propor a utilização de estruturas de bancos de filtros como etapa de pré-processamento no método de separação cega de fontes.

Conforme apresentado no Capítulo 2, para se realizar o processamento no domínio da frequência, os sinais das misturas passam pela STFT, que seleciona parte do sinal por meio de uma janela. Uma vez no domínio da frequência, é realizada a análise das componentes independentes em cada uma das raias de frequência. Em seguida, são feitas as correções de permutação e escalonamento antes dos sinais estimados voltarem ao domínio do tempo via transformada STFT inversa. A Fig. 4.2 ilustra, de forma simplificada, as etapas desse processamento.

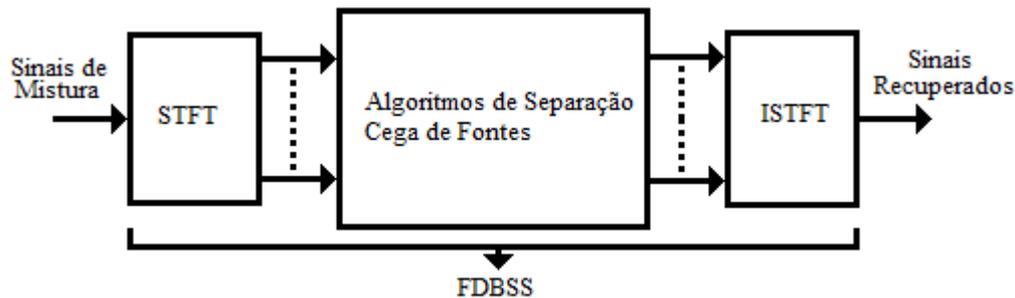


Figura 4.2: Etapas do método FDBSS

No método de separação de fontes proposto neste trabalho, utilizamos bancos de filtros de análise para fazer a decomposição das misturas em faixas reduzidas de frequência. Em seguida, realizamos a separação das fontes independentemente em cada uma das bandas de frequência, utilizando o algoritmo FDBSS. Os sinais resultantes são então alinhados entre raias, corrigindo possíveis permutações. Este processo é feito através do cálculo de correlações entre os sinais nas subbandas. A última etapa corresponde à obtenção das estimativas finais a partir das suas componentes nas subbandas, através da utilização de filtros de síntese.

A Fig. 4.3 ilustra o método proposto, utilizando os bancos de análise e síntese modulados por cosseno descritos no Capítulo 3. Nesse caso, a decomposição dos sinais é feita com filtros de análise com faixas de passagem de mesma largura, conhecidos como bancos de filtros uniformes. Portanto, utilizam-se fatores de decimação iguais nas diferentes subbandas. Apesar de os bancos de filtros poderem ser projetados para reconstrução perfeita (ou quase-perfeita) do sinal decomposto quando se emprega decimação pelo fator máximo (conforme descrito no Capítulo 2), é sabido que quando se aplicam modificações diferentes aos sinais nas diversas subbandas o cancelamento de *aliasing* entre subbandas passa a não acontecer. Para contornar este problema, devem-se reduzir os fatores de decimação, utilizando-se bancos de

filtros com decimação não-crítica. Nos experimentos apresentados no Capítulo 5, foram utilizadas estruturas com decimação máxima e não-máxima para se avaliar o efeito do *aliasing* na separação. Também foi avaliado o desempenho do algoritmo de separação de acordo com o número de subbandas, com a seletividade do banco de filtros e com o tamanho da janela utilizada em cada subbanda, em ambientes com diferentes tempos de reverberação.

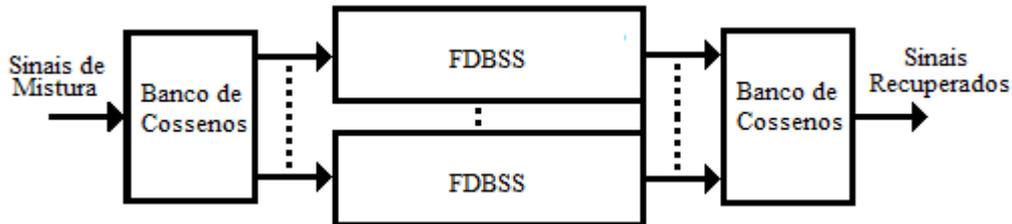


Figura 4.3: FDBSS em subbandas com bancos de filtros modulados por cosseno

Uma estrutura alternativa para a separação em subbandas utiliza bancos de filtros não-uniformes para a decomposição dos sinais em bandas de larguras diferentes. Dadas as características dos sinais de voz, que concentram a maior parte da energia nas baixas frequências, e da reverberação em ambientes naturais, descrita na seção anterior, a decomposição do tipo diádica se mostra apropriada. Essa decomposição foi apresentada em detalhes no Capítulo 2. A Fig. 4.4 ilustra o algoritmo FDBSS em subbandas utilizando bancos de filtros não-uniformes. Nesse algoritmo, a separação na banda de altas frequências é feita numa taxa maior do que nas de baixas, uma vez que a largura do filtro passa-altas do banco de análise é maior e, portanto, o fator de decimação máximo é igual a 2. Já nas bandas de frequências mais baixas, podem-se usar fatores de decimação maiores (inversamente proporcionais às larguras das bandas de passagem dos filtros de análise). Os tamanhos das janelas empregados em cada subbanda também podem ser diferentes, utilizando-se janelas mais curtas nas frequências mais altas.

Um cuidado adicional que se deve tomar na reconstrução dos sinais na estrutura com bancos de filtros não-uniformes, além de correções de possíveis permutações, é a compensação de atrasos diferentes introduzidos nos sinais das diferentes subbandas, sendo esses atrasos maiores nas subbandas de frequências mais baixas, cujos filtros de análise e síntese equivalentes são mais longos. Tomando-se esse cuidado, pode-se empregar um banco de síntese apropriado para se fazer a correta reconstrução dos sinais. Como vimos no Capítulo 2, os filtros de análise e síntese podem ser projetados para reconstrução perfeita (PR). No entanto, a modificação dos sinais nas subbandas pelos filtros de separação em geral gera *aliasing* nas estimativas das fontes, mesmo se utilizando bancos PR. Para minimizar este problema, reduzimos os fatores de decimação nos nossos experimentos, conforme apresentado no Capítulo 5.

Nas simulações também foram avaliados os efeitos do tamanho da janela e do número de subbandas no desempenho do algoritmo em diferentes condições de reverberação.

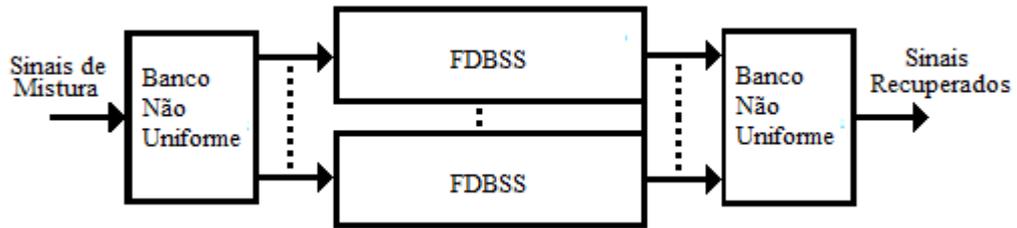


Figura 4.4: FDBSS em subbandas com bancos de filtros não-uniformes

Capítulo 5

Resultados das Simulações

Neste capítulo serão apresentados os resultados das simulações realizadas com o método de separação cega de fontes em subbandas proposto no Capítulo 4. O ambiente e as fontes de áudio utilizados nas simulações serão descritos na próximas seções. Em seguida, mostraremos as figuras de mérito utilizadas para a avaliação dos sinais obtidos nas simulações. Finalmente, apresentaremos os resultados obtidos com os algoritmos FDBSS convencional e em subbandas, utilizando os bancos de filtros modulados por cosseno e não-uniforme, com e sem decimação crítica, conforme apresentado no Capítulo 4. Utilizando os tamanhos de janela que resultaram no melhor desempenho da separação em cada subbanda, obtivemos uma estrutura em subbandas ótima para cada tipo de banco de filtros. Os resultados desta estrutura foi comparado aos do algoritmo FDBSS convencional para ambientes com diferentes tempos de reverberação.

5.1 Ambiente de Simulação

Para a realização das simulações desta dissertação foi utilizado o método *Image Source Model* [15], [16] e [17], para que, a partir de arquivos de áudio gravados em estúdio, fossem gerados os sons semelhantes aos gravados numa sala convencional.

A sala simulada tem como dimensões 4,45 m de altura, 3,55 m de comprimento e 2,5 m de largura. A sala é composta por paredes que não contêm portas, nem janelas. As paredes e os revestimentos de teto e piso são do mesmo material, escolhido de forma a obter um determinado tempo de reverberação T_{60} . Com o tempo de reverberação especificado, o algoritmo calcula os coeficientes de cada um dos lados do ambiente [1].

Dois microfones são utilizados para detectar os sinais e o “centro de gravidade” do arranjo encontra-se a 2 m da parede posterior e a 1,5 m da parede esquerda. Ambos os microfones encontram-se a 1,6 m de altura e distam 4 cm entre si.

Serão utilizadas duas fontes sonoras que se encontram a um raio de 1 m do centro

do arranjo. Uma das fontes forma um ângulo de 45° e a outra de 120° em relação a uma linha imaginária paralela à parede traseira da sala, passando pelo centro do arranjo de microfones [1]. A Fig. 5.1 mostra o arranjo utilizado nas simulações.

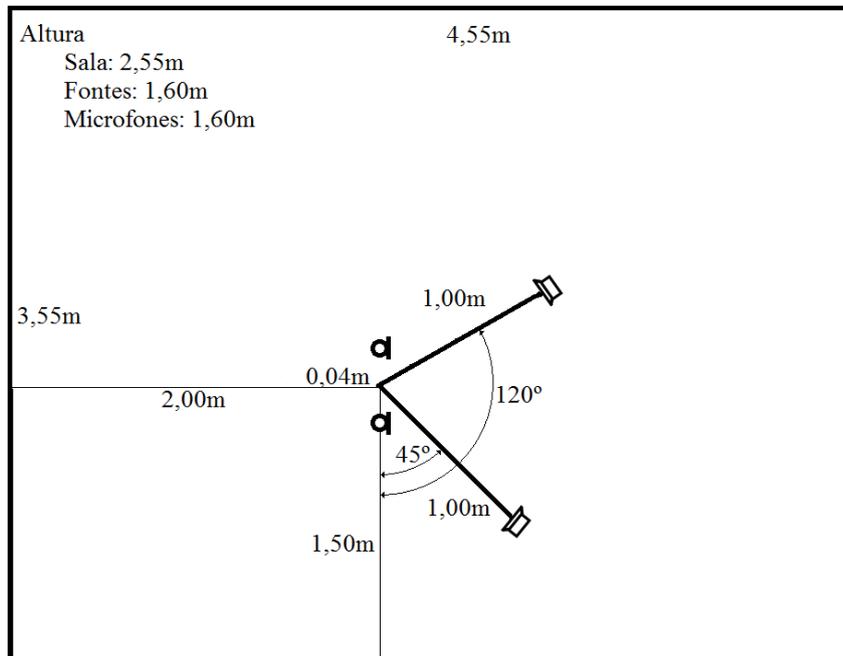


Figura 5.1: Arranjo da sala utilizada nas simulações

5.2 Arquivos de Áudio

As fontes empregadas nas simulações foram extraídas da base de dados SASSEC [18], sendo usados trechos de 10 segundos de duração cada, amostrados a 16 kHz, gravados em 2 canais com 16 bits por amostra. Estes sinais foram reamostrados para que a frequência de amostragem caísse para 8 kHz [1]. Foram usados dois arquivos de audio um com voz masculina e outro com voz feminina.

5.3 Avaliação de Desempenho

Para avaliar a eficiência do processamento dos sinais de entrada foram utilizadas figuras de mérito. As figuras calculam a razão de energias [19] entre as componentes presentes no sinal obtido. Essas razões são a *SDR*, razão sinal-distorção, a *SIR*, razão sinal-interferência, e a *SAR*, razão sinal-artefatos.

Considerando \hat{s}_j como o sinal recuperado a partir do sinal original desejado, s_j , podemos decompô-lo em:

$$\hat{s}_j = s_{target} + e_{interf} + e_{artif} + e_{noise} \quad (5.1)$$

onde s_{target} é uma versão da fonte real s_j com uma distorção (filtragem) aceitável [19], e_{interf} é a componente da estimativa correspondente a interferências das outras fontes, e_{noise} é a componente devido a ruídos dos sensores e e_{artif} é a componente da estimativa que corresponde à presença de artefatos (erros diferentes das fontes não desejadas e dos ruídos do ambiente) na estimativa.

Para obter essas componentes, conforme demonstrado em [19], devemos decompor \hat{s}_j em projeções ortogonais. Esse sinal é decomposto por três “projetores ortogonais” [19]: um formado pelo sinal original, outro formado pelos sinais exceto o original e o último pelo produto dos sinais e suas componentes de ruído.

Uma vez obtidas estas componentes, calculamos as figuras de mérito através da expressões:

$$SDR = 10 \log_{10} \left(\frac{\|s_{target}\|^2}{\|e_{interf} + e_{artif} + e_{noise}\|^2} \right) \quad (5.2)$$

$$SIR = 10 \log_{10} \left(\frac{\|s_{target}\|^2}{\|e_{interf}\|^2} \right) \quad (5.3)$$

$$SAR = 10 \log_{10} \left(\frac{\|s_{target} + e_{noise} + e_{interf}\|^2}{\|e_{artif}\|^2} \right) \quad (5.4)$$

Para sinais de áudio, a precisão dos resultados dependem, entre outros fatores, do número de bits por amostra [19].

5.4 FDBSS

Realizaremos, inicialmente, as simulações utilizando o algoritmo FDBSS em banda cheia, conforme a Fig. 4.2. Nessas simulações utilizamos, para o domínio da frequência uma FFT de 2048 pontos. Para essa e todas as demais simulações deste capítulo, o algoritmo de separação de fontes utilizado foi o *Natural ICA*, mencionado na Seção 2.5. As figuras de mérito SIR 1, SAR 1 e SDR 1 referem-se ao sinal de voz feminina enquanto SIR 2, SAR 2 e SDR 2 referem-se ao sinal de voz masculina.

Para um ambiente com tempo de reverberação de 130 ms, variamos o tamanho da janela, conforme descrito na Tabela 5.1.

Da mesma forma, para um tempo de reverberação de 300 ms, obtivemos os resultados presentes na Tabela 5.2.

Podemos verificar que os melhores resultados de *SIR* e *SDR* foram obtidos com a janela de 1024 pontos tanto para a reverberação de 130 ms quanto a reverberação de 300 ms, enquanto que a janela de 2048 amostras forneceu os melhores desempenhos para a *SAR*. Também observamos que para o maior tempo de reverberação os resultados da separação foram significativamente piores, indicando a dificuldade do

Tabela 5.1: FDBSS para Reverberação de 130 ms

Número de bins da FFT: 2048				
Resultados em dB				
Janela	256	512	1024	2048
SDR 1	6,056	7,793	16,609	17,925
SDR 2	7,170	11,308	16,429	15,446
SIR 1	6,159	7,931	18,547	23,459
SIR 2	7,429	11,862	19,228	22,267
SAR 1	23,295	23,472	21,108	19,370
SAR 2	20,271	20,802	19,713	16,485

Tabela 5.2: FDBSS para Reverberação de 300 ms

Número de bins da FFT: 2048				
Resultados em dB				
Janela	256	512	1024	2048
SDR 1	6.753	3.676	5.888	7.614
SDR 2	4.707	4.288	6.412	7.293
SIR 1	7.661	3.965	6.340	9.244
SIR 2	5.108	4.714	7.007	9.399
SAR 1	14.683	17.055	16.845	13.147
SAR 2	16.414	15.841	16.123	11.919

problema.

5.5 FDBSS em Subbandas com Banco Modulado por Cosseno

Nestas simulações, realizamos a separação de fontes separadamente em cada uma das subbandas de um banco de filtros modulado por cosseno, conforme a Fig.4.3. Para essas simulações, os filtros protótipos utilizados foram projetados conforme a equação (3.20). Mantivemos a janela de amostragem do STFT em 1024 pontos com passo de deslocamento igual à metade do comprimento da janela. As Figs. 5.2, 5.3, 5.4 e 5.5 mostram as respostas em frequência dos filtros de análise de um banco com 2, 4, 8 e 16 subbandas, respectivamente.

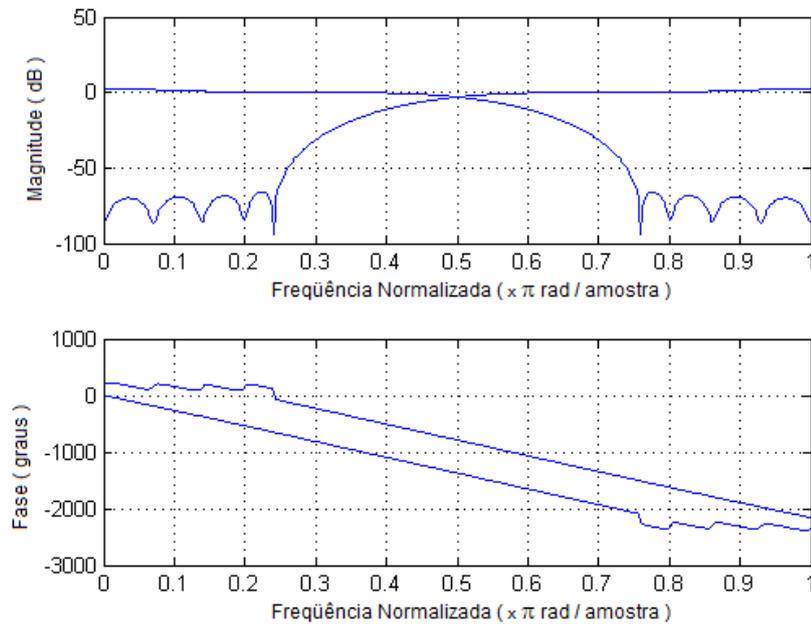


Figura 5.2: Respostas em frequência dos filtros modulados por cossenos com 2 bandas

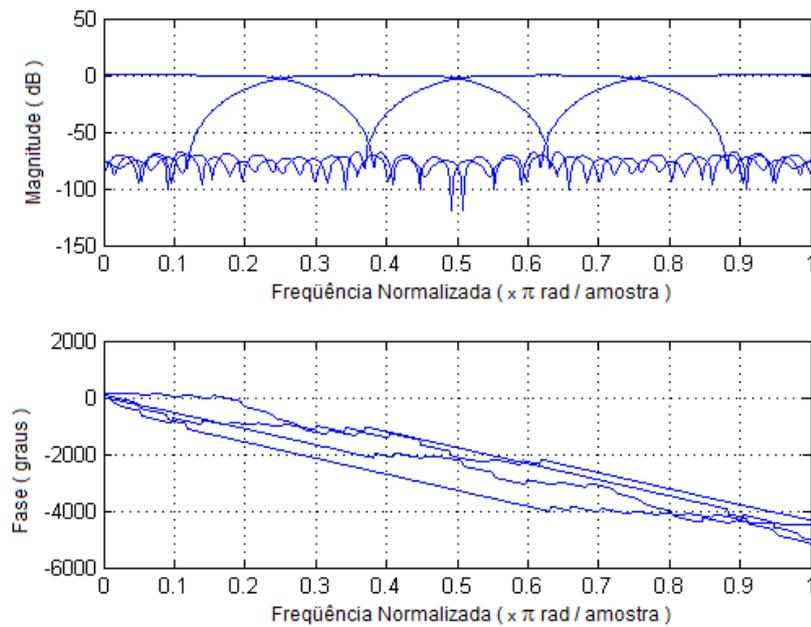


Figura 5.3: Respostas em frequência dos filtros modulados por cossenos com 4 bandas

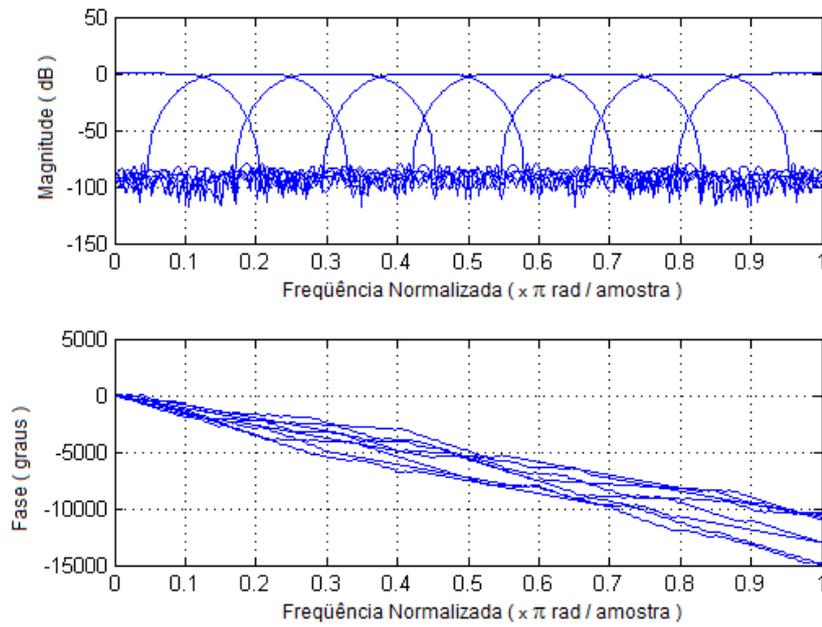


Figura 5.4: Respostas em frequência dos filtros modulados por cossenos com 8 bandas

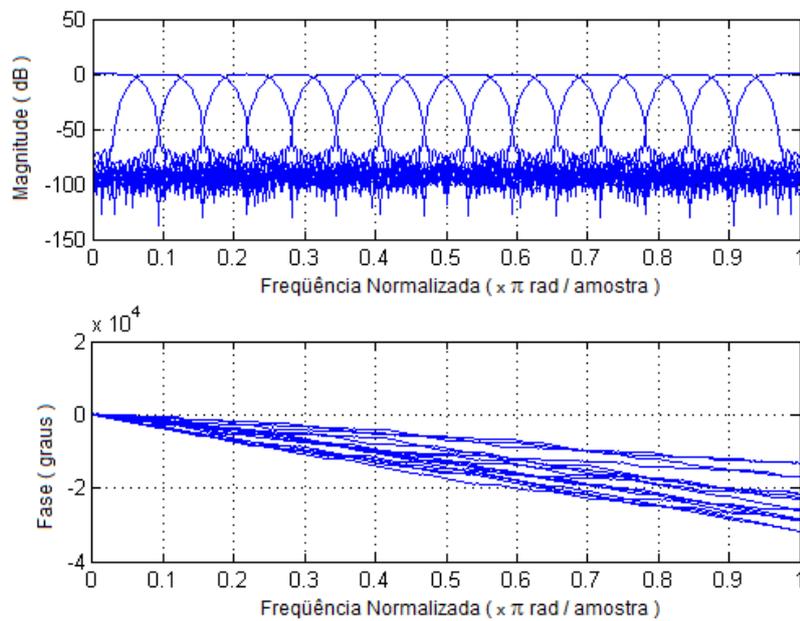


Figura 5.5: Respostas em frequência dos filtros modulados por cossenos com 16 bandas

Inicialmente utilizamos o banco de filtros criticamente decimado, ou seja, com fator de decimação igual ao número de subbandas. Para um tempo de reverberação igual a 130 ms, obtivemos os resultados das Tabelas 5.3, 5.4, 5.5 e 5.6 para estruturas de 2, 4, 8 e 16 subbandas, respectivamente. Em todas as tabelas apresentadas neste trabalho, os resultados da coluna “Completo” são os valores obtidos com os sinais estimados após todo o processamento.

Tabela 5.3: FDBSS com banco de filtros modulado por cosseno com 2 bandas

Reverberação: 130 ms			
Fator de decimação 2			
Número de bandas: 2			
Comprimento do protótipo: 32			
Resultados em dB			
Banda	1	2	Completo
SDR 1	17,954	17,954	17,885
SDR 2	17,482	17,482	17,455
SIR 1	24,671	24,671	24,737
SIR 2	23,140	23,140	23,176
SAR 1	19,009	19,009	18,903
SAR 2	18,880	18,880	18,831

Tabela 5.4: FDBSS com banco de filtros modulado por cosseno com 4 bandas

Reverberação: 130 ms					
Fator de decimação 4					
Número de bandas: 4					
Comprimento do protótipo: 64					
Resultados em dB					
Banda	1	2	3	4	Completo
SDR 1	11,275	11,275	10,162	10,097	10,321
SDR 2	12,837	12,837	11,792	11,780	11,769
SIR 1	15,105	15,105	16,547	15,961	13,511
SIR 2	19,828	19,828	19,874	20,643	16,755
SAR 1	13,729	13,729	11,391	11,508	13,347
SAR 2	13,850	13,850	12,571	12,422	13,517

Tabela 5.5: FDBSS com banco de filtros modulado por cosseno com 8 bandas

Reverberação: 130 ms
Fator de decimação 8
Número de bandas: 8
Comprimento do protótipo: 128
Resultados em dB

Banda	1	2	3	4	5	6	7	8	Completo
SDR 1	5,673	5,673	5,065	7,903	8,740	7,734	8,663	9,103	4,450
SDR 2	7,175	7,175	3,087	11,096	8,886	10,418	10,514	11,244	6,189
SIR 1	8,175	8,175	9,425	14,050	12,455	11,603	13,601	13,965	7,660
SIR 2	10,64	10,64	5,786	16,488	13,674	15,824	15,077	16,777	11,075
SAR 1	9,874	9,874	7,516	9,279	11,384	10,317	10,528	10,991	7,956
SAR 2	10,13	10,13	7,449	12,674	10,822	12,006	12,516	12,759	8,221

Tabela 5.6: FDBSS com banco de filtros modulado por cosseno com 16 bandas

Reverberação: 130 ms
Fator de decimação 16
Número de bandas: 16
Comprimento do protótipo: 256
Resultados em dB

Banda	1	2	3	4	5	6	7	8	
SDR 1	1,769	1,769	6,124	3,472	5,846	4,371	2,877	5,008	
SDR 2	4,655	4,655	3,414	3,234	5,434	8,930	4,425	6,611	
SIR 1	3,170	3,170	8,924	5,669	8,449	7,163	4,853	6,988	
SIR 2	6,496	6,496	4,985	5,453	7,194	11,645	6,431	9,083	
SAR 1	9,071	9,071	9,879	8,527	9,887	8,375	8,476	10,162	
SAR 2	10,146	10,146	9,790	8,301	10,966	12,546	9,633	10,743	
Banda	9	10	11	12	13	14	15	16	Completo
SDR 1	6,665	6,752	3,600	4,541	4,888	2,025	2,704	2,198	-0,608
SDR 2	7,493	7,870	9,014	6,459	4,447	6,782	6,413	6,196	2,339
SIR 1	8,497	8,971	5,200	6,871	7,222	3,168	3,839	3,763	1,930
SIR 2	10,074	10,498	11,908	8,504	6,099	9,122	8,416	8,443	6,553
SAR 1	11,872	11,248	9,859	9,171	9,453	10,091	10,589	8,913	5,081
SAR 2	11,386	11,670	12,414	11,285	10,397	11,085	11,322	10,713	5,275

Igualmente, para o tempo de reverberação de 300 ms, realizamos simulações com os bancos de filtros modulados por cosseno de 2, 4, 8 e 16 subbandas, cujos resultados encontram-se nas Tabelas 5.7, 5.8, 5.9 e 5.10, respectivamente.

Tabela 5.7: FDBSS com banco de filtros modulado por cosseno com 2 bandas

Reverberação: 300 ms			
Fator de decimação 2			
Número de bandas: 2			
Comprimento do protótipo: 32			
Resultados em dB			
Banda	1	2	Completo
SDR 1	5,169	5,169	5,057
SDR 2	5,777	5,777	5,698
SIR 1	5,960	5,960	5,836
SIR 2	6,960	6,960	6,849
SAR 1	13,935	13,935	13,910
SAR 2	12,798	12,798	12,843

Tabela 5.8: FDBSS com banco de filtros modulado por cosseno com 4 bandas

Reverberação: 300 ms					
Fator de decimação 4					
Número de bandas: 4					
Comprimento do protótipo: 64					
Resultados em dB					
Banda	1	2	3	4	Completo
SDR 1	5,436	5,436	5,263	3,534	5,235
SDR 2	6,397	6,397	8,196	8,895	6,186
SIR 1	6,943	6,943	8,201	5,820	6,674
SIR 2	8,619	8,619	11,355	13,424	8,296
SAR 1	11,563	11,563	8,958	8,424	11,576
SAR 2	10,931	10,931	11,370	10,976	10,934

Tabela 5.9: FDBSS com banco de filtros modulado por cosseno com 8 bandas

Reverberação: 300 ms
Fator de decimação 8
Número de bandas: 8
Comprimento do protótipo: 128
Resultados em dB

Banda	1	2	3	4	5	6	7	8	Completo
SDR 1	2,820	2,820	2,489	5,377	4,875	3,959	2,702	4,818	1,110
SDR 2	4,513	4,513	1,870	8,895	6,557	8,879	6,645	9,211	3,804
SIR 1	4,576	4,576	5,354	8,628	6,968	5,971	4,159	6,708	3,192
SIR 2	6,821	6,821	3,718	12,395	9,515	12,991	9,291	13,656	6,725
SAR 1	8,900	8,900	6,760	8,718	9,845	9,246	9,564	10,184	7,004
SAR 2	9,182	9,182	8,010	11,709	10,080	11,224	10,535	11,328	7,742

Tabela 5.10: FDBSS com banco de filtros modulado por cosseno com 16 bandas

Reverberação: 300 ms
Fator de decimação 16
Número de bandas: 16
Comprimento do protótipo: 256
Resultados em dB

Banda	1	2	3	4	5	6	7	8	
SDR 1	1,191	1,191	4,844	0,379	5,373	1,77	2,494	3,006	
SDR 2	4,216	4,216	1,606	2,041	4,530	8,28	4,058	4,148	
SIR 1	2,511	2,511	6,942	1,746	7,750	3,46	4,274	4,528	
SIR 2	5,812	5,812	2,893	3,862	6,144	11,044	5,717	5,914	
SAR 1	8,939	8,939	9,811	8,291	9,799	8,303	8,606	9,608	
SAR 2	10,346	10,346	9,316	8,191	10,554	11,900	10,073	9,900	
Banda	9	10	11	12	13	14	15	16	Completo
SDR 1	4,811	4,941	2,064	2,756	4,243	0,790	2,732	0,642	-2,237
SDR 2	6,556	7,050	8,054	5,063	4,074	6,501	6,326	5,093	1,915
SIR 1	6,409	6,864	3,222	4,059	6,293	1,642	3,914	1,753	0,165
SIR 2	8,693	9,327	10,602	6,784	5,619	8,761	8,478	6,939	5,575
SAR 1	10,821	10,218	10,063	10,057	9,404	10,548	10,439	9,329	4,407
SAR 2	11,211	11,423	11,944	10,741	10,364	10,962	10,983	10,501	5,423

Os resultados das simulações com o algoritmo FDBSS em subbandas apresentados nas tabelas acima demonstram que, para ambientes com maior tempo de reverberação, melhores resultados são obtidos empregando-se um número maior de subbandas. Enquanto que, para 130 ms a estrutura com 2 subbandas obteve os melhores resultados, para 300 ms a estrutura com 4 subbandas foi a que apresentou melhor desempenho. Como nessas simulações foi empregada decimação crítica, a medida em que se aumenta o número de subbandas, o efeito de *aliasing* se torna mais acentuado.

Para verificarmos a influência da decimação nos resultados, fizemos as mesmas simulações com o banco de filtros modulado por cosseno com fator de decimação igual a 2 para todas as bandas. As Tabelas 5.11, 5.12 e 5.13 mostram os resultados obtidos utilizando bancos de 4, 8 e 16 subbandas, respectivamente, para o ambiente com tempo de reverberação igual a 130 ms.

Tabela 5.11: FDBSS com banco de filtros modulado por cosseno com 4 bandas

Reverberação: 130 ms					
Fator de decimação 2					
Número de bandas: 4					
Comprimento do protótipo: 64					
Resultados em dB					
Banda	1	2	3	4	Completo
SDR 1	18,623	18,623	14,616	10,498	18,501
SDR 2	17,845	17,845	15,995	12,665	17,723
SIR 1	26,339	26,339	19,086	17,128	26,464
SIR 2	23,095	23,095	23,743	23,098	23,090
SAR 1	19,438	19,438	16,589	11,645	19,267
SAR 2	19,406	19,406	16,812	13,099	19,236

Tabela 5.12: FDBSS com banco de filtros modulado por cosseno com 8 bandas

Reverberação: 130 ms									
Fator de decimação 2									
Número de bandas: 8									
Comprimento do protótipo: 128									
Resultados em dB									
Banda	1	2	3	4	5	6	7	8	Completo
SDR 1	18,672	18,672	16,789	16,386	15,901	12,690	15,550	17,387	18,316
SDR 2	18,172	18,172	16,737	18,340	14,871	13,701	18,443	18,624	18,083
SIR 1	25,953	25,953	22,598	20,996	22,805	19,358	19,659	23,834	25,154
SIR 2	23,713	23,713	22,179	25,749	23,838	24,247	26,031	28,134	23,688
SAR 1	19,582	19,582	18,136	18,265	16,914	13,794	17,731	18,521	19,338
SAR 2	19,612	19,612	18,224	19,221	15,478	14,118	19,286	19,146	19,498

Tabela 5.13: FDBSS com banco de filtros modulado por cosseno com 16 bandas

Reverberação: 130 ms									
Fator de decimação 2									
Número de bandas: 16									
Comprimento do protótipo: 256									
Resultados em dB									
Banda	1	2	3	4	5	6	7	8	
SDR1	13,626	13,626	19,326	17,353	15,024	17,479	17,442	17,835	
SDR2	18,139	18,139	17,194	16,783	16,988	18,624	18,633	17,634	
SIR1	15,061	15,061	29,305	23,603	18,329	23,407	21,189	24,874	
SIR2	24,151	24,151	22,511	21,152	21,707	25,828	25,812	24,338	
SAR1	19,267	19,267	19,791	18,547	17,822	18,779	19,855	18,806	
SAR2	19,407	19,407	18,730	18,793	18,805	19,552	19,567	18,693	
Banda	9	10	11	12	13	14	15	16	Completo
SDR1	15,263	15,433	16,436	15,644	15,724	16,894	17,991	15,663	14,358
SDR2	14,025	16,979	19,361	19,290	16,265	17,789	19,522	17,751	18,270
SIR1	20,844	23,450	21,008	18,699	20,445	21,840	25,420	18,036	15,868
SIR2	22,613	27,205	28,391	27,835	21,497	27,560	28,263	25,213	23,536
SAR1	16,705	16,199	18,334	18,667	17,550	18,599	18,870	19,489	19,789
SAR2	14,695	17,420	19,947	19,951	17,844	18,280	20,151	18,623	19,823

Da mesma forma, para o tempo de reverberação de 300 ms, os resultados para os bancos com 4, 8 e 16 estão mostrados nas Tabelas 5.14, 5.15 e 5.16.

Tabela 5.14: FDBSS com banco de filtros modulado por cosseno com 4 bandas

Reverberação: 300 ms					
Fator de decimação 2					
Número de bandas: 4					
Comprimento do protótipo: 64					
Resultados em dB					
Banda	1	2	3	4	Completo
SDR1	4,328	4,328	5,852	4,0184	4,177
SDR2	4,939	4,939	9,626	9,8523	4,937
SIR1	5,051	5,051	6,854	5,9033	4,859
SIR2	6,058	6,058	11,210	14,368	6,004
SAR1	13,651	13,651	13,531	9,544	13,784
SAR2	12,336	12,336	15,089	11,902	12,529

Tabela 5.15: FDBSS com banco de filtros modulado por cosseno com 8 bandas

Reverberação: 300 ms									
Fator de decimação 2									
Número de bandas: 8									
Comprimento do protótipo: 128									
Resultados em dB									
Banda	1	2	3	4	5	6	7	8	Completo
SDR1	3,138	3,138	5,155	5,565	6,199	4,779	4,159	6,378	2,836
SDR2	-3,254	-3,254	5,005	10,173	8,536	11,149	9,645	12,707	-2,615
SIR1	3,690	3,690	6,017	6,387	7,026	6,531	5,134	7,183	3,266
SIR2	-2,997	-2,997	6,155	11,771	10,357	14,988	12,531	15,730	-2,396
SAR1	13,913	13,913	13,571	14,097	14,597	10,441	12,288	14,855	14,767
SAR2	13,913	13,913	12,281	15,571	13,574	13,600	13,018	15,818	14,849

Tabela 5.16: FDBSS com banco de filtros modulado por cosseno com 16 bandas

Reverberação: 300 ms									
Fator de decimação 2									
Número de bandas: 16									
Comprimento do protótipo: 256									
Resultados em dB									
Banda	1	2	3	4	5	6	7	8	
SDR 1	2,855	2,855	9,178	2,747	14,239	6,188	8,344	6,935	
SDR 2	8,551	8,551	1,405	3,912	10,913	13,168	7,623	9,144	
SIR 1	3,366	3,366	11,638	3,550	16,578	6,931	9,355	8,693	
SIR 2	10,833	10,833	2,053	5,148	11,960	15,668	8,277	10,305	
SAR 1	14,044	14,044	13,107	12,057	18,136	15,024	15,647	12,262	
SAR 2	12,781	12,781	12,094	11,133	17,876	16,872	16,764	15,828	
Banda	9	10	11	12	13	14	15	16	Completo
SDR 1	6,044	7,203	2,356	2,566	10,179	6,498	7,109	-1,754	3,178
SDR 2	8,920	12,393	13,478	9,501	10,453	14,361	12,466	8,873	8,273
SIR 1	6,761	9,073	3,077	3,213	11,850	7,268	8,3765	-1,246	3,392
SIR 2	10,912	15,694	15,943	11,847	11,909	18,369	14,500	10,777	8,927
SAR 1	15,048	12,271	12,247	12,845	15,411	15,138	13,665	11,495	18,002
SAR 2	13,602	15,247	17,220	13,570	16,178	16,624	16,887	13,721	17,337

De acordo com as simulações, a decimação pelo fator 2 proporcionou uma melhora nítida para o ambiente com pouca reverberação. Verificamos, também, que o melhor resultado foi obtido usando 4 bandas.

Um fato que ficou evidente nos resultados foi a maior eficiência nas bandas de maiores frequências em relação aos resultados obtidos com o sinal Completo. No entanto, como o sinal de voz apresenta mais energia nas bandas de menor frequência, estas acabam fazendo com que os resultados finais tendam a seguir os obtidos nessas bandas. Esse resultado sugere que, se agruparmos as bandas de forma não uniforme, poderemos obter resultados melhores, como veremos na próxima seção.

Verificamos também a influência do tamanho da janela nos resultados. Para isso, mantivemos o número de bandas do filtro igual a 4. Os resultados para tempo de reverberação igual a 130 ms encontram-se nas Tabelas 5.17, 5.18, 5.19 e 5.20.

Tabela 5.17: FDBSS com banco de filtros modulado por cosseno com janela de 256 pontos

Tamanho da janela: 256					
Reverberação: 130 ms					
Resultados em dB					
Banda	1	2	3	4	Completo
SDR 1	6,503	12,255	3,172	13,058	7,785
SDR 2	11,185	13,142	7,071	18,825	11,256
SIR 1	6,624	15,537	10,127	13,539	7,954
SIR 2	11,809	16,772	15,664	20,481	11,974
SAR 1	22,968	15,13	4,553	23,038	22,596
SAR 2	20,201	15,701	7,834	23,852	19,697

Tabela 5.18: FDBSS com banco de filtros modulado por cosseno com janela de 512 pontos

Tamanho da janela: 512					
Reverberação: 130 ms					
Resultados em dB					
Banda	1	2	3	4	Completo
SDR 1	16,952	13,108	4,017	16,9	17,161
SDR 2	17,114	13,93	6,708	19,61	16,509
SIR 1	19,167	18,044	10,769	19,319	19,729
SIR 2	20,782	20,638	16,171	24,675	20,121
SAR 1	20,989	14,857	5,397	20,646	20,711
SAR 2	19,589	15,01	7,334	21,245	19,032

Tabela 5.19: FDBSS com banco de filtros modulado por cosseno com janela de 1024 pontos

Tamanho da janela: 1024					
Reverberação: 130 ms					
Resultados em dB					
Banda	1	2	3	4	Completo
SDR 1	16,811	11,55	3,208	16,171	16,507
SDR 2	17,444	13,849	6,951	17,611	16,685
SIR 1	21,333	18,011	10,449	22,026	21,317
SIR 2	24,411	22,741	16,911	26,411	23,798
SAR 1	18,734	12,731	4,491	17,504	18,28
SAR 2	18,435	14,472	7,501	18,235	17,642

Tabela 5.20: FDBSS em bandas de banco de filtros de cossenos FDBSS com banco de filtros modulado por cosseno com janela de 2048 pontos

Tamanho da janela: 2048					
Reverberação: 130 ms					
Resultados em dB					
Banda	1	2	3	4	Completo
SDR 1	10,418	9,618	2,94	13,25	10,893
SDR 2	12,453	12,326	6,495	15,109	12,573
SIR 1	14,07	16,789	10,244	19,02	14,824
SIR 2	20,61	20,823	16,089	25,137	20,681
SAR 1	13,036	10,632	4,226	14,641	13,284
SAR 2	13,211	13,024	7,106	15,577	13,339

Como esperado, verificamos que as bandas de maior frequência obtiveram melhores resultados com janelas mais curtas. Já as janelas mais longas tiveram melhor performance nas baixas frequências. A fim de obter um melhor desempenho global, utilizamos o comprimento de janela que apresentou melhor resultado em cada banda. A Tabela 5.21 mostra esses resultados.

Tabela 5.21: FDBSS com banco de filtros modulado por cosseno com janelas distintas

Reverberação: 130 ms					
Resultados em dB					
JANELA	1024	1024	512	1024	
Banda	1	2	3	4	Completo
SDR 1	16,811	11,55	4,017	16,9	16,533
SDR 2	17,444	13,849	6,708	19,61	16,703
SIR 1	21,333	18,011	10,769	19,319	21,39
SIR 2	24,411	22,741	16,171	24,675	23,781
SAR 1	18,734	12,731	5,397	20,646	18,282
SAR 2	18,435	14,472	7,334	21,245	17,669

Simulações semelhantes foram realizadas para reverberação de 300 ms. As Tabelas 5.22, 5.23, 5.24 e 5.25 apresentam os resultados obtidos para diferentes tamanhos de janelas.

Tabela 5.22: FDBSS com banco de filtros modulado por cosseno com janela de 256 pontos

Tamanho da janela: 256					
Reverberação: 300 ms					
Resultados em dB					
Banda	1	2	3	4	Completo
SDR 1	2,797	5,605	-0,536	4,395	3,745
SDR 2	3,948	7,432	5,752	11,385	3,867
SIR 1	3,062	6,741	2,206	4,741	3,97
SIR 2	4,429	8,896	10,575	12,3	4,199
SAR 1	16,821	12,82	4,806	16,812	18,163
SAR 2	15,076	13,392	7,852	18,843	16,606

Tabela 5.23: FDBSS com banco de filtros modulado por cosseno com janela de 512 pontos

Tamanho da janela: 512					
Reverberação: 300 ms					
Resultados em dB					
Banda	1	2	3	4	Completo
SDR 1	5,939	7,822	0,247	5,777	6,648
SDR 2	6,697	9,941	5,507	11,217	6,43
SIR 1	6,459	9,423	3,083	6,691	7,13
SIR 2	7,392	11,753	12,2	13,547	7,001
SAR 1	16,297	13,4	5,175	13,837	17,199
SAR 2	15,726	14,892	6,807	15,223	16,315

Tabela 5.24: FDBSS com banco de filtros modulado por cosseno com janela de 1024 pontos

Tamanho da janela: 1024					
Reverberação: 300 ms					
Resultados em dB					
Banda	1	2	3	4	Completo
SDR 1	4,601	5,697	0,345	5,244	5,26
SDR 2	5,759	8,73	6,218	10,585	5,501
SIR 1	5,481	7,567	4,32	6,594	6,251
SIR 2	7,294	11,493	14,112	14,059	6,888
SAR 1	13,045	10,959	3,932	11,835	13,089
SAR 2	11,762	12,301	7,153	13,344	11,94

Tabela 5.25: FDBSS com banco de filtros modulado por cosseno com janela de 2048 pontos

Tamanho da janela: 2048					
Reverberação: 300 ms					
Resultados em dB					
Banda	1	2	3	4	Completo
SDR 1	4,747	4,964	0,401	5,018	5,489
SDR 2	6,5	8,177	5,78	9,816	6,234
SIR 1	6,182	7,584	4,56	6,796	7,026
SIR 2	8,977	11,547	13,338	14,026	8,354
SAR 1	11,189	9,1	3,808	10,582	11,53
SAR 2	10,637	11,15	6,815	12,056	10,957

Do mesmo modo que fizemos anteriormente, selecionamos os tamanhos de janela que apresentaram melhor performance para cada subbanda para reverberação de 300 ms. A Tabela 5.26 apresenta os resultados.

Tabela 5.26: FDBSS com banco de filtros modulado por cosseno com janelas diferentes

Reverberação: 300 ms					
Resultados em dB					
JANELA	2048	512	1024	512	
Banda	1	2	3	4	Completo
SDR 1	4,747	7,822	0,345	5,777	5,605
SDR 2	6,5	9,941	6,218	11,217	6,439
SIR 1	6,182	9,423	4,32	6,691	7,106
SIR 2	8,977	11,753	14,112	13,547	8,466
SAR 1	11,189	13,4	3,932	13,837	11,722
SAR 2	10,637	14,892	7,153	15,223	11,3

Independente do tempo de reverberação utilizado, houve uma ligeira melhora nos resultados das simulações utilizando-se o banco de filtros modulado por cosseno. Existe pouca energia no sinal de voz nas altas frequências, o que justifica a melhora não ser tão expressiva nas métricas adotadas. Porém, os resultados sugerem que devemos agrupar melhor as baixas frequências.

5.6 FDBSS em Subbandas com Banco de Filtros Não-Uniforme

Utilizamos, como alternativa ao banco de filtros modulados por cossenos, o banco de filtros não uniforme apresentado no Capítulo 3, para que as bandas dos sinais fossem arranjadas de formas diferentes. A característica deste banco mostra que ele pode privilegiar algumas bandas em detrimento de outras. No caso, devemos privilegiar as bandas de baixas frequências, já que nestas se encontra a maior parte da energia da voz. A Fig. 4.4 mostra a nova proposta para o processamento dos sinais.

Usamos a estrutura em árvore diádica mostrada na Fig. 3.10, que apresenta 4 bandas distintas com decimações distintas. O filtro protótipo utilizado foi o filtro de Johnstone 32D.

Da mesma forma que foi feito com a estrutura de banco de filtros modulado por cosseno, alteramos o tamanho das janelas, porém sempre mantendo o seu deslocamento igual à metade do comprimento das mesmas. As Tabelas 5.27, 5.28, 5.29 e 5.30 mostram os resultados das simulações para tempo de reverberação igual a 130 ms.

Tabela 5.27: FDBSS em Subbandas com Banco Não-Uniforme com janela de 256 pontos

Tamanho de janela: 256
 Reverberação: 130 ms
 Resultados em dB

BANDA	1	2	3	4	Completo
SDR 1	12,342	5,388	8,543	4,123	8,670
SDR 2	11,852	1,045	10,484	7,811	6,962
SIR 1	16,458	9,943	13,798	10,787	15,574
SIR 2	16,973	2,855	17,365	16,905	12,119
SAR 1	14,569	7,681	10,258	5,525	9,780
SAR 2	13,535	7,534	11,560	8,470	8,802

Tabela 5.28: FDBSS em Subbandas com Banco Não-Uniforme com janela de 512 pontos

Tamanho de janela: 512
 Reverberação: 130 ms
 Resultados em dB

BANDA	1	2	3	4	Completo
SDR 1	9,587	4,338	8,622	4,544	7,019
SDR 2	9,940	1,170	10,288	7,558	6,280
SIR 1	13,031	8,554	14,55	11,827	12,41
SIR 2	14,46	3,26	18,083	17,159	11,857
SAR 1	12,415	6,975	10,052	5,719	8,744
SAR 2	11,719	7,030	11,145	8,146	7,963

Tabela 5.29: FDBSS em Subbandas com Banco Não-Uniforme com janela de 1024 pontos

Tamanho de janela: 1024
 Reverberação 130 ms
 Resultados em dB

BANDA	1	2	3	4	Completo
SDR 1	5,732	3,247	7,432	4,727	3,978
SDR 2	5,737	0,670	9,205	7,120	3,528
SIR 1	8,692	7,374	13,195	12,022	8,834
SIR 2	9,094	3,088	17,140	16,873	8,385
SAR 1	9,343	6,099	8,974	5,888	6,232
SAR 2	8,912	6,101	10,050	7,693	5,835

Tabela 5.30: FDBSS em Subbandas com Banco Não-Uniforme com janela de 2048 pontos

Tamanho de janela: 2048
 Reverberação: 130 ms
 Resultados em dB

BANDA	1	2	3	4	Completo
SDR 1	3,188	1,630	5,904	4,914	1,592
SDR 2	3,327	-0,096	8,227	6,196	1,545
SIR 1	6,240	5,521	11,934	11,361	6,213
SIR 2	6,231	2,684	15,837	16,752	5,703
SAR 1	7,082	4,983	7,420	6,337	4,361
SAR 2	7,375	5,031	9,165	6,687	4,683

Identicamente, fizemos simulações, alterando o tamanho das janelas, para um tempo de reverberação de 300 ms. As Tabelas 5.31, 5.32, 5.33 e 5.34 apresentam os resultados dessas simulações.

Tabela 5.31: FDBSS em Subbandas com Banco Não-Uniforme com janela de 256 pontos

Tamanho de janela: 256
 Reverberação: 300 ms
 Resultados em dB

BANDA	1	2	3	4	Completo
SDR 1	4,606	2,365	5,815	0,862	3,225
SDR 2	6,024	-0,192	7,890	5,725	3,910
SIR 1	5,528	4,587	8,506	3,733	4,490
SIR 2	7,949	1,142	11,584	11,076	6,054
SAR 1	12,856	7,635	9,743	5,549	10,523
SAR 2	11,132	7,254	10,602	7,548	8,967

Tabela 5.32: FDBSS em Subbandas com Banco Não-Uniforme com janela de 512 pontos

Tamanho de janela: 512
 Reverberação: 300 ms
 Resultados em dB

BANDA	1	2	3	4	Completo
SDR 1	5,221	2,194	5,181	1,238	3,455
SDR 2	6,602	0,053	7,640	5,497	4,622
SIR 1	7,096	4,773	7,998	4,727	5,665
SIR 2	9,446	1,96	12,014	11,601	7,972
SAR 1	10,547	6,931	9,032	5,077	8,489
SAR 2	10,253	6,687	9,879	7,010	7,960

Tabela 5.33: FDBSS em Subbandas com Banco Não-Uniforme com janela de 1024 pontos

Tamanho de janela: 1024
 Reverberação: 300 ms
 Resultados em dB

BANDA	1	2	3	4	Completo
SDR 1	3,664	1,011	3,478	0,952	1,887
SDR 2	4,571	-0,185	6,837	5,992	2,685
SIR 1	6,320	4,106	7,014	4,953	5,204
SIR 2	6,926	1,839	11,229	12,371	5,916
SAR 1	7,972	5,364	6,806	4,361	5,756
SAR 2	9,157	6,292	9,116	7,373	6,477

Tabela 5.34: FDBSS em Subbandas com Banco Não-Uniforme com janela de 2048 pontos

Tamanho de janela: 2048
 Reverberação: 300 ms
 Resultados em dB

BANDA	1	2	3	4	Completo
SDR 1	1,087	-0,099	1,352	1,223	-0,584
SDR 2	2,117	-0,472	6,148	5,031	0,549
SIR 1	3,231	2,842	4,401	5,032	2,262
SIR 2	4,521	2,026	10,877	12,013	3,805
SAR 1	6,868	4,798	5,669	4,744	4,623
SAR 2	7,147	5,233	8,271	6,267	4,839

Independentemente do tempo de reverberação, verificamos que não houve melhoria com a variação do tamanho da janela de amostragem. Verificamos que a estrutura não-uniforme, da forma que foi implementada, estaria sofrendo influência das diferentes decimações feitas em cada estágio, sendo mais uma “fonte de ruído” na separação.

Por isso, resolvemos adotar outra abordagem de banco de filtros não uniformes, exibido na Fig 3.10. Usando as identidades nobres, apresentadas anteriormente na Fig. 3.9, deslocamos todos os decimadores para a direita dos filtros de análise, assim como todos os interpoladores para esquerda dos filtros de síntese. Em seguida, somente decimamos e interpolamos os sinais por 2 ao final da filtragem. A Fig. 5.6 descreve essa estrutura alternativa.

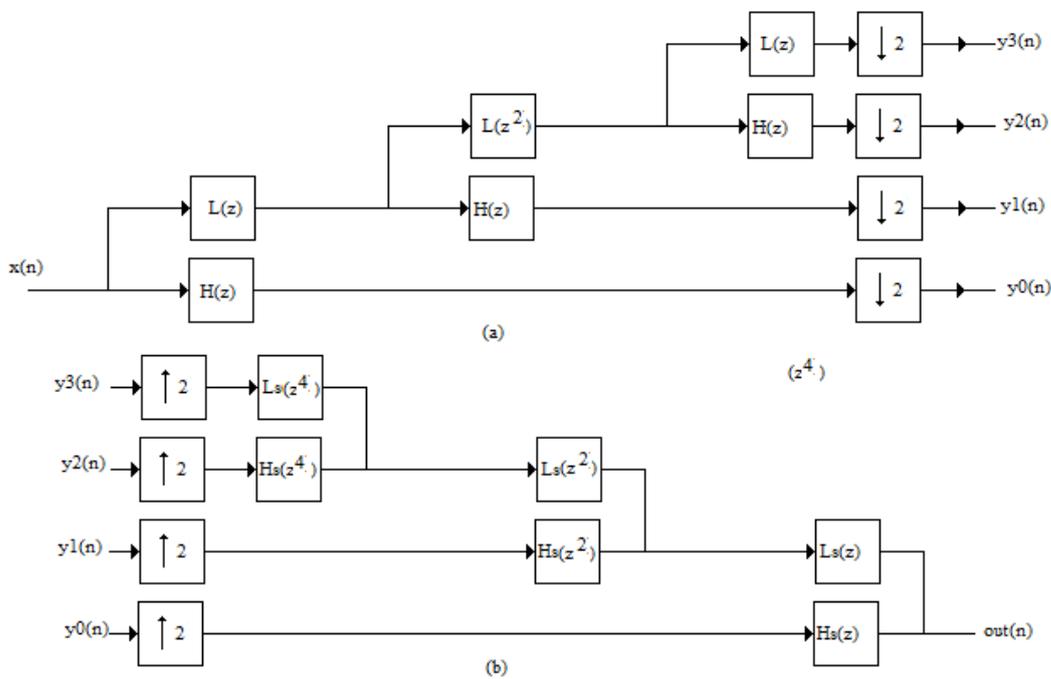


Figura 5.6: Banco de Filtros Não-Uniformes com Decimação Não-Crítica

Com essa nova estrutura, fizemos as mesmas simulações variando o comprimento da janela de amostragem, inicialmente para o tempo de reverberação de 130 ms. Os resultados obtidos encontram-se nas Tabelas 5.35, 5.36, 5.37 e 5.38.

Tabela 5.35: FDBSS em Subbandas com Banco Não-Uniforme com Decimação Não-Crítica com janela de 256 pontos

Tamanho de janela: 256
 Reverberação: 130 ms
 Resultados em dB

BANDA	1	2	3	4	Completo
SDR 1	6,684	18,513	12,201	4,123	7,561
SDR 2	10,435	11,048	12,985	7,811	10,900
SIR 1	6,796	20,596	14,929	10,787	7,697
SIR 2	10,966	11,676	17,568	16,906	11,481
SAR 1	23,473	22,739	15,650	5,525	23,327
SAR 2	20,164	20,041	14,919	8,470	20,224

Tabela 5.36: FDBSS em Subbandas com Banco Não-Uniforme com Decimação Não-Crítica com janela de 512 pontos

Tamanho de janela: 512
 Reverberação 130 ms
 Resultados em dB

BANDA	1	2	3	4	Completo
SDR 1	13,588	19,244	13,459	4,544	14,025
SDR 2	16,464	15,960	15,423	7,558	16,176
SIR 1	13,429	26,086	17,226	11,827	15,028
SIR 2	19,321	18,622	21,996	17,159	18,867
SAR 1	21,288	20,262	15,907	5,719	21,013
SAR 2	19,684	19,410	16,531	8,145	19,589

Tabela 5.37: FDBSS em Subbandas com Banco Não-Uniforme com Decimação Não-Crítica com janela de 1024 pontos

Tamanho de janela: 1024
 Reverberação: 130 ms
 Resultados em dB

BANDA	1	2	3	4	Completo
SDR 1	17,098	17,894	12,538	4,728	16,629
SDR 2	17,705	16,687	14,460	7,120	16,944
SIR 1	21,530	25,552	17,340	12,022	21,165
SIR 2	25,279	22,352	23,513	16,877	23,829
SAR 1	19,070	18,723	14,364	5,888	18,545
SAR 2	18,553	18,088	15,056	7,693	17,958

Tabela 5.38: FDBSS em Subbandas com Banco Não-Uniforme com Decimação Não-Crítica com janela de 2048 pontos

Tamanho de janela: 2048
 Reverberação: 130 ms
 Resultados em dB

BANDA	1	2	3	4	Completo
SDR 1	10,537	14,056	10,089	4,914	10,071
SDR 2	12,880	13,173	12,440	6,197	12,382
SIR 1	13,606	21,34	15,821	11,361	13,8
SIR 2	20,883	19,303	20,959	16,752	20,627
SAR 1	13,675	14,987	11,552	6,337	14,176
SAR 2	13,665	14,437	13,133	6,687	14,533

A Tabela 5.39 apresenta a simulação utilizando o tamanho de janela que gerou o melhor resultado em cada uma das bandas para reverberação de 130 ms.

Tabela 5.39: FDBSS em Subbandas com Banco Não-Uniforme com Decimação Não-Crítica com tamanhos de janela distintos

Reverberação: 130 ms
 Resultados em dB

JANELA	1024	512	512	512	
BANDA	1	2	3	4	Completo
SDR 1	17,098	19,244	13,460	4,544	16,954
SDR 2	17,705	15,960	15,423	7,559	17,104
SIR 1	21,529	26,086	17,227	11,827	21,139
SIR 2	25,278	18,622	21,996	17,159	22,863
SAR 1	19,070	20,262	15,907	5,719	19,075
SAR 2	18,553	19,409	16,531	8,145	18,466

Os resultados presentes nas Tabelas 5.40, 5.41, 5.42 e 5.43 são de simulações variando o comprimento da janela de amostragem para tempo de reverberação de 300 ms.

Tabela 5.40: FDBSS em Subbandas com Banco Não-Uniforme com Decimação Não-Crítica com janela de 256 pontos

Tamanho de janela: 256
 Reverberação: 300 ms
 Resultados em dB

BANDA	1	2	3	4	Completo
SDR 1	2,938	6,770	5,275	0,862	3,558
SDR 2	3,558	4,931	7,889	5,725	4,123
SIR 1	3,166	7,505	6,208	3,733	3,715
SIR 2	3,883	5,903	9,312	11,076	4,358
SAR 1	17,558	15,556	13,345	5,549	19,586
SAR 2	16,470	12,902	13,906	7,548	18,244

Tabela 5.41: FDBSS em Subbandas com Banco Não-Uniforme com Decimação Não-Crítica com janela de 512 pontos

Tamanho de janela: 512
 Reverberação: 300 ms
 Resultados em dB

BANDA	1	2	3	4	Completo
SDR 1	5,147	6,763	7,207	1,238	5,569
SDR 2	5,648	5,130	10,261	5,497	5,998
SIR 1	5,615	7,71	8,627	4,727	5,869
SIR 2	6,226	6,379	12,008	11,601	6,42
SAR 1	16,102	14,522	13,313	5,077	18,323
SAR 2	15,623	12,052	15,326	7,010	17,222

Tabela 5.42: FDBSS em Subbandas com Banco Não-Uniforme com Decimação Não-Crítica com janela de 1024 pontos

Tamanho de janela: 1024
 Reverberação: 300 ms
 Resultados em dB

BANDA	1	2	3	4	Completo
SDR 1	3,712	5,770	6,402	0,952	4,262
SDR 2	5,105	3,905	9,571	5,992	5,424
SIR 1	4,508	7,431	7,973	4,953	4,826
SIR 2	6,825	5,486	12,238	12,371	6,438
SAR 1	12,792	11,470	12,223	4,361	14,642
SAR 2	10,777	10,140	13,206	7,372	13,130

Tabela 5.43: FDBSS em Subbandas com Banco Não-Uniforme com Decimação Não-Crítica com janela de 2048 pontos

Tamanho de janela: 2048
 Reverberação: 300 ms
 Resultados em dB

BANDA	1	2	3	4	Completo
SDR 1	6,487	7,471	5,920	1,224	6,878
SDR 2	7,420	4,688	8,896	5,031	7,134
SIR 1	8,221	9,904	8,227	5,033	8,3
SIR 2	9,672	6,599	12,289	12,014	8,732
SAR 1	11,923	11,572	10,380	4,744	13,020
SAR 2	11,794	10,033	11,804	6,267	12,797

Os dados da Tabela 5.44 são referentes à simulação feita com a janela cujo comprimento obteve melhor performance por subbanda para tempo de reverberação de 300 ms.

Tabela 5.44: FDBSS em Subbandas com Banco Não-Uniforme com Decimação Não-Crítica com comprimentos de janela diferentes

Reverberação: 300 ms
 Resultados em dB

JANELA	2048	2048	512	512	
BANDA	1	2	3	4	Completo
SDR 1	6,487	7,472	7,207	1,238	6,916
SDR 2	7,420	4,688	10,261	5,497	7,229
SIR 1	8,221	9,904	8,626	4,727	8,291
SIR 2	9,672	6,599	12,008	11,601	8,734
SAR 1	11,923	11,572	13,313	5,077	13,180
SAR 2	11,794	10,034	15,326	7,010	13,110

Das tabelas acima, podemos perceber que, apenas alterando a forma que as bandas foram reagrupadas, obtivemos uma performance melhor com a estrutura não-uniforme em relação ao banco de filtros modulado por cosseno com janelas de comprimentos iguais para todas as subbandas. Ao selecionarmos o melhor comprimento de janela em cada banda, percebemos a influência de utilizar um filtro mais curto para as bandas de alta frequência, o que se mostrou apropriado principalmente para o ambiente mais reverberante. Os sinais recuperados, utilizando a estrutura não-uniforme otimizada, apresentaram, tanto qualitativa como quantitativamente, avaliações melhores do que os obtidos da forma convencional. Isso demonstra que um pré-processamento adequado pode melhorar o desempenho da separação de fontes.

Capítulo 6

Conclusões

Nesta dissertação foi feito um estudo da separação cega de sinais de voz em ambientes reverberantes. Foi proposta a utilização de bancos de filtros no pré-processamento dos sinais antes de se realizar a separação dos sinais. Foi avaliada a influência do banco de filtros de análise e do tamanho da janela utilizada no algoritmo de separação empregado, para dois ambientes com diferentes tempo de reverberação.

Apresentamos, no Capítulo 2, a análise de componentes independentes. Vimos os tipos de misturas existentes e o problema das ambiguidades inerentes ao problema. Apontamos as premissas para a análise de componentes independentes e comentamos as três abordagens principais deste problema, dando um maior enfoque ao algoritmo *Natural ICA*, utilizado nas simulações realizadas neste trabalho. Em seguida, caracterizamos o problema e descrevemos a sua solução no domínio da frequência, exibindo os passos desse processamento.

No Capítulo 3 apresentamos um resumo dos principais conceitos de processamento em multitaxas. Caracterizamos o funcionamento dos blocos fundamentais para esse processamento: o decimador e o interpolador. Descrevemos duas estruturas bem conhecidas de bancos de filtros: o banco modulado por cosseno e o banco não-uniforme com estrutura em árvore diádica.

No Capítulo 4, apresentamos o método de separação cega de fontes em subbandas proposto neste trabalho. Discutimos os fatores que devem ser considerados para a escolha do tamanho da janela de amostragem, que recai num compromisso entre modelar corretamente o ambiente e manter as características estatísticas do sinal de voz, considerando a sua não-estacionariedade. Caracterizamos o fenômeno de reverberação em locais fechados, onde pudemos ver a relação entre o tempo de reverberação e as componentes de diferentes frequências dos sinais. Em seguida, introduzimos a separação das fontes em subbandas, indicando a motivação do tema desta dissertação, que é o uso de estruturas de bancos de filtros, aliado ao uso de janelas de tamanhos distintos por subbanda, para melhorar o desempenho de métodos de separação de fontes. Também apresentamos as estruturas que foram

utilizadas neste trabalho.

Nas simulações apresentadas no Capítulo 5, inicialmente detalhamos o ambiente das simulações e os sinais utilizados. Também apresentamos as figuras de mérito empregadas neste trabalho para avaliação dos métodos. Realizamos simulações utilizando a abordagem FDBSS original, encontrando a melhor janela de amostragem, tanto para um ambiente de baixa quanto para um de mais alta reverberação, obtendo os sinais estimados das fontes, que foram utilizados como referência para as demais simulações.

Em seguida, apresentamos os resultados da realização em paralelo da separação cega de fontes em bandas distintas do banco de filtros. Avaliamos a influência do fator de decimação e do número de bandas utilizadas. Verificamos que um baixo fator de decimação proporcionou melhores resultados para ambientes de baixa reverberação. Observamos também que, em algumas bandas, as figuras de mérito tiveram resultados superiores ao resultado da separação completa. Isso indicou que, rearrumando as bandas de outra forma, poderíamos obter resultados finais melhores.

Uma vez selecionado o número de subbandas utilizadas, alteramos o tamanho da janela de amostragem a fim de verificar seu efeito. Observamos que havia uma ligeira melhora ao se utilizar janelas diferentes para diferentes bandas. Isso se deve à maior energia dos sinais nas bandas de baixas frequências. Motivados pelo fato de haver melhor performance em algumas subbandas em relação ao resultado final, utilizamos a estrutura de banco de filtros não-uniforme diádica.

Inicialmente, simulamos o algoritmo utilizando a estrutura em árvore com decimação crítica. Percebemos que o resultado não foi satisfatório devido à interferência proporcionada pela decimação em excesso, que gerou uma espécie de ruído para a separação. Substituímos, então, o banco de filtros não-uniforme por outro com decimação por 2 em todas as subbandas. Realizamos testes para janelas de amostragem de diversos tamanhos e selecionamos aquelas que apresentaram melhores resultados nas diferentes subbandas. Esta abordagem, utilizando um comprimento de janela diferente por banda, apresentou sinais com qualidade melhor do que os utilizados como referência até então.

Portanto, podemos concluir deste trabalho:

- A validade em realizar o pré-processamento antes da separação das fontes.
- O uso de banco de filtro no pré-processamentos proporcionou superiores em relação à abordagem original.

Como direcionamento deste trabalho:

- Uso de outras estruturas de banco de filtros.

- Substituição do algoritmo de separação por algoritmos de banda larga, conforme os citados em [2].
- Uso de algoritmo de estimação do tempo de reverberação em cada subbanda, para auxiliar na seleção ótima do tamanho da janela.

Referências Bibliográficas

- [1] LAPORTE, L. V. M. *Algoritmos de Separação Cega de Sinais de Áudio no Domínio da Frequência em Ambientes Reverberantes: Estudo e Comparações*. Dissertação de mestrado, COPPE - Universidade Federal do Rio de Janeiro, Outubro 2010.
- [2] PETRAGLIA, M. R., BATALHEIRO, P. B., HADDAD, D. B. “Métodos de Separação Cega de Fontes”. In: *Tutoriais do XVII Congresso Brasileiro de Automática, - CBA 2008*, Campinas: Book Editora, cap. 2, pp. 133 – 157, 2008.
- [3] HYVÄRINEN, A., KARHUNEN, J., OJA, E. *Independent Component Analysis*. John Wiley & Sons, 2001.
- [4] MAKINO, S., ARAKI, S., SAWADA, H. “Frequency-Domain Blind Source Separation”. In: S. Makino, T. Lee, H. S. (Ed.), *Blind Speech Separation*, Springer, cap. 2, pp. 47–78, 2007.
- [5] BENESTY, J., CHEN, J., HUANG, Y. “Microphone Array Signal Processing”. cap. 3, pp. 39–66, Springer, 2008.
- [6] DUDA, R. O., HART, P. E., STORK, D. G. *Pattern Classification*. Wiley Interscience, 2000.
- [7] MATSUOKA, K., NAKASHIMA, S. “Minimal Distortion Principle for Blind Source Separation”, *Proc. ICA*, pp. 722–727, Dezembro 2001.
- [8] SAWADA, H., MUKAI, R., DE LA K. DE RYHOVE, S., et al. “Spectral Smoothing for Frequency-Domain Blind Source Separation”. In: *International Workshop on Acoustic Echo and Noise Control (IWAENC’03)*, pp. 311–314, Kyoto, Japan, Setembro 2003.
- [9] VAIDYANATHAN, P. *Multirate Systems and Filter Banks*. Prentice Hall, 1993.
- [10] DINIZ, P. S. R. *Adaptive Filtering, Algorithms and Practical Implementation, 2nd. edition*. Kluwer Academic Publishers, 2002.

- [11] HAYKIN, S. *Adaptive Filter Theory, 3rd. edition.* Prentice Hall, 1996.
- [12] OPPENHEIN, A. V., SCHAFER, R. W. *Discrete-Time Signal Processing.* Prentice-Hall International, 1989.
- [13] ARAKI, S., MAKINO, S., NISHIKAWA, T., et al. “The Fundamental Limitation of Frequency-Domain Blind Source Separation for Convulsive Mixtures of Speech”, *IEEE Trans. Speech Audio Proc.*, v. 11, n. 2, pp. 109–116, March 2003.
- [14] <http://www.cara.de/ENU/index.html>. Acessado em Setembro de 2012.
- [15] ALLEN, J., BERKELEY, D. “Image Method for Efficiently Small Room Acoustics”, *Journal of the Acoustic Society of America*, v. 65, n. 4, pp. 943–950, April 1979.
- [16] LEHMANN, E. A., JOHANSSON, A. M. “Prediction of Energy Decay in Room Impulse Responses Simulated with an Image-Source Model”, *The Journal of the Acoustic Society of America*, v. 124, n. 1, pp. 269–277, July 2008.
- [17] LEHMANN, E. A., JOHANSSON, A. M., NORDHOLM, S. “Reverberation-Time Prediction Method for Room Impulse Responses Simulated with the Image-Source Model”. In: *2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, July 2006.
- [18] VINCENT, E., SAWADA, H., BOFILL, P., et al. “First Stereo Audio Source Separation Evaluation Campaign: Data, Algorithms and Results”. In: *Proc. Int. Conf. on Independent Component Analysis and Signal Separation*, pp. 552–559, a, a 2007.
- [19] VICENT, E., GRIBONVAL, R., FÉVOTTE, C. “Performance Measurement in Blind Audio Source Separation”, *IEEE Transactions On Audio, Speech And Language Processing*, v. 14, n. 4, pp. 1462–1469, July 2006.