



Universidade do Estado do Rio de Janeiro
Centro de Tecnologia e Ciências
Faculdade de Engenharia

Ronaldo Alencar da Rocha

Separação Cega de Sinais de Fala Utilizando Detectores de Voz

Rio de Janeiro

2014

Ronaldo Alencar da Rocha

Separação Cega de Sinais de Fala Utilizando Detectores de Voz



Dissertação apresentada, como requisito parcial para obtenção do título de Mestre em Ciências, ao Programa de Pós-Graduação em Engenharia Eletrônica, da Universidade do Estado do Rio de Janeiro. Área de concentração: Sistemas Inteligentes.

Orientador: Prof. Dr. Paulo Bulkool Batalheiro

Rio de Janeiro

2014

Ronaldo Alencar da Rocha

Separação Cega de Sinais de Fala Utilizando Detectores de Voz

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre em Ciências, ao Programa de Pós-Graduação em Engenharia Eletrônica, da Universidade do Estado do Rio de Janeiro. Área de concentração: Sistemas Inteligentes.

Aprovado em:

Banca Examinadora:

Prof. Dr. Paulo Bulkool Batalheiro, D.Sc. (Orientador)

Faculdade de Engenharia - UERJ

Prof. Dr. Pedro Henrique Gouvêa Coelho

Faculdade de engenharia - UERJ

Prof. Dr. Diego Barreto Haddad

Centro Federal de Educação Tecnológica Celso Suckow da Fonseca - CEFET

Rio de Janeiro

2014

AGRADECIMENTO

Gostaria de agradecer:

Primeiramente à Deus pois sem Ele nada seria possível, principalmente por me fortalecer, incentivar, cuidar e por colocar na minha vida pessoas maravilhosas.

À minha família, familiares e amigos, pela compreensão e paciência, dos quais sempre recebi incentivo e apoio, nos bons e maus momentos, que em muitas vezes foram essenciais para a evolução deste trabalho, em particular aos meus pais, Valdir Tomaz da Rocha e Terezinha Alencar de Oliveira, meus primeiros professores, e aos meus irmãos Claudio de Oliveira e Antonia Aila, cuja compreensão e carinho foram de excepcional valia nesta jornada, fornecendo-me todo o alicerce para permanecer nesta jornada.

Aos colegas do curso de pós-graduação pela ajuda e apoio em especial às minhas amigas Letícia Guimarães e Luciane Barreira que sempre me incentivaram, apoiaram e ajudaram sempre com muita paciência e carinho.

Ao meu orientador Paulo, por ter me guiado nesta jornada de forma tão dedicada, compartilhando seu vasto e inestimável conhecimento da teoria aqui apresentada, sempre me fornecendo ideias e sugestões valiosas para a elaboração deste trabalho, também pela paciência e principalmente por ter acreditado em mim.

Por fim, a todos os professores da UERJ responsáveis pelo meu crescimento intelectual e profissional, em particular os professores Jorge Amaral e Tiago Roux por me ajudar, apoiar e compartilhar seu conhecimento com dedicação, atenção e paciência. Também a todos os funcionários pelo suporte extra classe durante esta jornada, à Universidade do Estado do Rio de Janeiro (UERJ) por fornecer toda a infra-estrutura necessária e ao Programa de Pós-Graduação em Engenharia Eletrônica (PEL) pela oportunidade.

Meu sincero obrigado a todos! E sem vocês o término desta etapa não seria possível.

*A persistência
é o menor caminho do êxito.*

Charles Chaplin

RESUMO

Rocha, Ronaldo Alencar da. *Separação Cega de Sinais de Fala Utilizando Detectores de Voz*. 119 f. Dissertação (Mestrado em Engenharia Eletrônica) - Faculdade de Engenharia, Universidade do Estado do Rio de Janeiro (UERJ), Rio de Janeiro, 2014.

Neste trabalho contemplamos o emprego de detectores de voz como uma etapa de pré-processamento de uma técnica de separação cega de sinais implementada no domínio do tempo, que emprega estatísticas de segunda ordem para a separação de misturas convolutivas e determinadas. Seu algoritmo foi adaptado para realizar a separação tanto em banda cheia quanto em sub-bandas, considerando a presença e a ausência de instantes de silêncio em misturas de sinais de voz. A ideia principal visa detectar trechos das misturas que contenham atividade de voz, evitando que o algoritmo de separação seja acionado na ausência de voz, promovendo ganho de desempenho e redução do custo computacional.

Palavras-chave: Separação cega de fontes, detectores de voz, estruturas em sub-bandas, processamento multitaxas.

ABSTRACT

In this work we contemplate the use of detector's voice as a preprocessing step of a blind source separation technique implemented in the time domain, employing second order statistics in the separation of convolutive and determined mixtures. This algorithm is adapted to perform the separation both in fullband and in subbands, considering the presence and the absence of a moments of silence in mixtures of voice signals. The main idea aims at detect portions of the mixtures containing voice activity, avoiding that the separation algorithm is triggered in the absence of voice, promoting performance gain and reduced computational cost.

Keywords: Blind source separation, voice detectors, subbands.

LISTA DE FIGURAS

Figura 1 - Ilustração do problema de <i>cocktail party</i> [3].	16
Figura 2 - Aplicação de BSS para medir o eletrocardiograma fetal/materno a partir de sensores na superfície corporal [3].	16
Figura 3 - Método de BSS para aquisição do eletroencefalograma/magnetoencefalograma [3].	17
Figura 4 - Ilustração de um esquema simples de BSS.	21
Figura 5 - Problemas comuns em separação cega de fontes [23].	23
Figura 6 - Estrutura geral do sistema de mistura e separação para o problema de separação cega.	24
Figura 7 - Estrutura de um sistema de separação para misturas não lineares.	25
Figura 8 - Distribuição supergaussiana (em vermelho), gaussiana (em azul) e subgaussiana (em verde).	28
Figura 9 - Estrutura de um sistema de separação cega completo [19].	33
Figura 10- Divisão do sinal em blocos sobrepostos [40].	36
Figura 11- Ilustração de correlação para o caso para $Q = P = 2$ [34].	40
Figura 12- Ilustração da redução computacional realizada, empregando-se os S primeiros elementos da primeira coluna de \mathbf{W}_{pq} para um sistema TITO [34].	43
Figura 13- O parâmetro $\beta(i, m)$ numa implementação <i>offline</i> [47].	43
Figura 14- O parâmetro $\beta(i, m)$ numa implementação <i>online</i> [47].	44
Figura 15- O parâmetro $\beta(i, m)$ numa implementação <i>block online</i> [47].	44
Figura 16- Ilustração de um cenário acústico para um sistema TITO [40].	46
Figura 17- Desempenho do ALED empregado em mistura com ausência de trechos de silêncio.	51
Figura 18- Desempenho do ALED empregado em mistura com presença de trechos de silêncio.	52
Figura 19- Modelo lógico do LSED.	53
Figura 20- Desempenho do LSED empregado em mistura com presença de trechos de silêncio.	54
Figura 21- Desempenho do SMBVAD empregado em mistura com presença de trechos de silêncio.	59
Figura 22- Ilustração de funcionamento do decimador para $FD = 2$	61

Figura 23- Ilustração de funcionamento do expansor para $FD = 2$	61
Figura 24- Ilustração de bancos de filtros de análise e síntese respectivamente.	62
Figura 25- Ilustração de um sistema multitaxa completo em estrutura uniforme.	63
Figura 26- Resposta em frequência dos filtros $G_k(z)$	64
Figura 27- Ilustração de um sistema multitaxa completo em estrutura não uniforme maximamente decimado para $M = 4$	65
Figura 28- Representação equivalente de um sistema multitaxa completo em estrutura não uniforme maximamente decimado para $M = 4$	66
Figura 29- Resposta em frequência dos filtros $G_k(z)$ em estrutura não uniforme em árvore para $M = 4$	67
Figura 30- Resposta em frequência do filtro protótipo.	69
Figura 31- Ilustração de BSS em sub-bandas para um sistema TITO.	70
Figura 32- Ilustração para o k -ésimo canal da configuração linear TITO para BSS em sub-bandas.	71
Figura 33- Sala acústica virtual [63].	75
Figura 34- Mistura obtidas para $T_{60} = 25ms$: (a) Mistura 1 e (b) Mistura 2.	76
Figura 35- Evolução da SIR média para $M = 1, 2$ e 4 , $FD_k = \lceil M/2 \rceil$ para $T_{60} = 25ms$ e $S_k = 256/FD_k$, SD em azul e CD em vermelho : (a) Mistura 1 e (b) Mistura 2.	80
Figura 36- Evolução da SIR média para $M = 1, 2$ e 4 , $FD_k = \lceil M/2 \rceil$ para $T_{60} = 100ms$ e $S_k = 768/FD_k$, SD em azul e CD em vermelho : (a) Mistura 1 e (b) Mistura 2.	80
Figura 37- Evolução da SIR média para $M = 1, 3, 4$ e 5 , $DIV = \lceil DIVmax/2 \rceil$ para $T_{60} = 25ms$ e $S_k = 256/FD_k$, SD em azul e CD em vermelho : (a) Mistura 1 e (b) Mistura 2.	89
Figura 38- Evolução da SIR média para $M = 1, 3, 4$ e 5 , $DIV = \lceil DIVmax/2 \rceil$ para $T_{60} = 100ms$ e $S_k = 768/FD_k$, SD em azul e CD em vermelho : (a) Mistura 1 e (b) Mistura 2.	89
Figura 39- Evolução da SIR média para a implementação em banda cheia e para a versão mesclada em estrutura uniforme com $FD = M4$, para $T_{60} = 25ms$ e S_k reduzido, SD em azul e CD em vermelho : (a) Mistura 1 e (b) Mistura 2.	95
Figura 40- Sala acústica real [65].	96
Figura 41- Mistura obtidas: (a) Mistura 3 e (b) Mistura 4.	97

Figura 42- Evolução da SIR média para $M = 1, 2$ e 4 , $FD_k = \lceil M/2 \rceil$ para $S_k = 2048/FD_k$, SD em azul e CD em vermelho : (a) Mistura 3 e (b) Mistura 4. . . 99

LISTA DE TABELAS

Tabela 1 - Critério de atualização de p	50
Tabela 2 - Critério de atualização de p	51
Tabela 3 - Outros parâmetros empregados no ALED.	52
Tabela 4 - Relação entre a sub-banda e a respectiva faixa de frequência.	54
Tabela 5 - Parâmetros empregados no LSED.	54
Tabela 6 - Parâmetros empregados no SMBVAD.	58
Tabela 7 - Resultados médios finais da SIR, SAR e SDR, para diferentes valores de FD e para $S_k = 256/FD_k$ e $768/FD_k$ para a Mistura 1.	78
Tabela 8 - Resultados médios finais da SIR, SAR e SDR, para diferentes valores de FD e para $S_k = 256/FD_k$ e $768/FD_k$ para a Mistura 2.	79
Tabela 9 - Número de blocos empregados no algoritmo de separação, para diferentes valores de M e FD para a Mistura 1.	81
Tabela 10- Número de blocos empregados no algoritmo de separação, para diferentes valores de M e FD para a Mistura 2.	82
Tabela 11- Número de multiplicações por iteração em implementação comum e em implementação eficiente, para diferentes valores de M e FD para a Mistura 1.	83
Tabela 12- Número de multiplicações por iteração em implementação comum e em implementação eficiente, para diferentes valores de M e FD para a Mistura 2.	84
Tabela 13- Resultados médios finais da SIR, SAR e SDR, para diferentes valores de FD e para $S_k = 256/FD_k$ para a Mistura 1 em estrutura não uniforme.	85
Tabela 14- Resultados médios finais da SIR, SAR e SDR, para diferentes valores de FD e para $S_k = 256/FD_k$ para a Mistura 2 em estrutura não uniforme.	86
Tabela 15- Resultados médios finais da SIR, SAR e SDR, para diferentes valores de FD e para $S_k = 768/FD_k$ para a Mistura 1 em estrutura não uniforme.	87
Tabela 16- Resultados médios finais da SIR, SAR e SDR, para diferentes valores de FD e para $S_k = 768/FD_k$ para a Mistura 2 em estrutura não uniforme.	88
Tabela 17- Número de blocos empregados no algoritmo de separação, para diferentes valores de M e FD para a Mistura 1.	90
Tabela 18- Número de blocos empregados no algoritmo de separação, para diferentes valores de M e FD para a Mistura 2.	91

Tabela 19- Tabela Complexidade Computacional para a Mistura 1 - $S_k = 256/FD_k$ e $S_k = 768/FD_k$	92
Tabela 20- Tabela Complexidade Computacional para a Mistura 2 - $S_k = 256/FD_k$ e $S_k = 768/FD_k$	93
Tabela 21- Tabela apresentando a estratégia para reduzir a complexidade computacional para a Mistura 1 - $S = 256$ e 768	94
Tabela 22- Tabela apresentando a estratégia para reduzir a complexidade computacional para a Mistura 2 - $S = 256$ e 768	94
Tabela 23- Resultados médios finais da SIR, SAR e SDR, para diferentes valores de FD e para $S_k = 2048/FD_k$ para a Mistura 3.	98
Tabela 24- Resultados médios finais da SIR, SAR e SDR, para diferentes valores de FD e para $S_k = 2048/FD_k$ para a Mistura 4.	98
Tabela 25- Número de blocos empregados no algoritmo de separação, para diferentes valores de M e FD para a Mistura 3.	100
Tabela 26- Número de blocos empregados no algoritmo de separação, para diferentes valores de M e FD para a Mistura 4.	100
Tabela 27- Número de multiplicações por iteração em implementação comum e em implementação eficiente, para diferentes valores de M e FD para a Mistura 3.	101
Tabela 28- Número de multiplicações por iteração em implementação comum e em implementação eficiente, para diferentes valores de M e FD para a Mistura 4.	101
Tabela 29- Resultados médios finais da SIR, SAR e SDR, para diferentes valores de FD e para $S_k = 2048/FD_k$ para a Mistura 3 em estrutura não uniforme.	102
Tabela 30- Resultados médios finais da SIR, SAR e SDR, para diferentes valores de FD e para $S_k = 2048/FD_k$ para a Mistura 4 em estrutura não uniforme.	102
Tabela 31- Número de blocos empregados no algoritmo de separação, para diferentes valores de M e FD para a Mistura 3.	103
Tabela 32- Número de blocos empregados no algoritmo de separação, para diferentes valores de M e FD para a Mistura 4.	103
Tabela 33- Tabela Complexidade Computacional para a Mistura 3 - $S_k = 2048/FD_k$. ..	104
Tabela 34- Tabela Complexidade Computacional para a Mistura 4 - $S_k = 2048/FD_k$. ..	104

LISTA DE SIGLAS

ALED	Adaptive Linear Energy-Based Detector
BSS	Blind Source Separation
CMFB	Cosine Modulated Filter Bank
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
ICA	Independent Component Analysis
LSED	Linear Sub-Band Energy Detector
MIMO	Multiple-Input Multiple-Output
PDF	Probability Density Function
PR	Perfect Reconstruction
SAR	Source-to-Artifact-Ratio
SDR	Source-to-Distortion-Ratio
SIR	Source-to-Interference Ratio
SMBVAD	Statistical Model-Based Voice Activity Detector
SNR	Signal-to-Noise Ratio
TDD	Time-Delayed Decorrelation
TITO	Two Input Two Output
VADs	Voice Activity Detectors

SUMÁRIO

	INTRODUÇÃO	15
1	ABORDAGEM GERAL SOBRE OS MÉTODOS DE SEPARAÇÃO CEGA DE FONTES	20
1.1	Introdução	20
1.2	Configurações de Mistura	20
1.3	Sistemas de mistura e de separação	20
1.4	Misturas Lineares e Instantâneas	22
1.5	Misturas Lineares e Convolutiva	24
1.6	Misturas Não Lineares	25
1.7	Conceitos e Propriedades de Métodos de Separação Cega de Fontes	26
1.7.1	<u>Independentes estatisticamente das fontes entre si</u>	27
1.7.2	<u>Não Branqueamento (<i>nonwhiteness</i>)</u>	27
1.7.3	<u>Não gaussianidade (<i>nongaussianity</i>)</u>	28
1.7.4	<u>Não Estacionária (<i>nonstationarity</i>)</u>	28
1.7.5	<u>Banda Estreita versus Banda Larga</u>	29
1.8	Problemas dos Métodos de Separação Cega de Fontes	29
1.8.1	<u>Permutação</u>	30
1.8.2	<u>Escalamento</u>	30
1.8.3	<u>Filtragem</u>	30
1.9	Métodos de avaliação de desempenho	30
1.10	Métodos de Separação Cega de Fontes	31
1.10.1	<u>Técnicas de Análise de Componentes Independentes</u>	32
2	MÉTODO DE SEPARAÇÃO CEGA DE FONTES NO DOMÍNIO DO TEMPO	33
2.1	Introdução	33
2.2	Formulação Matricial para Misturas Convolutivas	34
2.3	Matriz de Correlação	38
2.4	Função de Custo	39
2.4.1	<u>Otimização da Função de Custo</u>	40
2.4.2	<u>Gradiente Natural</u>	41

2.4.3	<u>Implementações do algoritmo: <i>offline</i>, <i>online</i> e <i>block online</i></u>	42
2.4.4	<u>Fator de aprendizagem adaptativo</u>	45
2.4.5	<u>Inicialização dos filtros de separação</u>	45
3	DETECTORES DE ATIVIDADE DE VOZ	47
3.1	Introdução	47
3.2	Detector Linear e Adaptativo de Energia (ALED)	48
3.3	Detector Linear de Energia em Sub-bandas (LSED)	52
3.4	Modelo Estatístico de Detecção de Voz (SMBVAD)	55
4	BANCO DE FILTROS E BSS EM SUB-BANDAS	60
4.1	Introdução	60
4.2	Sistemas Multitaxas	60
4.2.1	<u>Decimador</u>	60
4.2.2	<u>Expansor</u>	60
4.2.3	<u>Banco de filtros</u>	61
4.3	Banco de Filtros em Estrutura Uniforme	62
4.4	Banco de Filtros em Estrutura Não Uniforme	64
4.5	Bancos de Filtros Modulados por Cosseno	68
4.6	Implementação de BSS em sub-bandas	69
5	COMPLEXIDADE COMPUTACIONAL E RESULTADOS EXPERIMENTAIS	72
5.1	Introdução	72
5.2	Complexidade computacional	72
5.3	Resultados experimentais	74
5.3.1	<u>Considerações iniciais</u>	74
5.3.2	<u>Experimento 1</u>	77
5.3.3	<u>Experimento 2</u>	84
5.3.4	<u>Experimento 3</u>	94
5.3.5	<u>Experimento 4</u>	96
5.3.6	<u>Experimento 5</u>	101
	CONCLUSÃO	105
	APÊNDICE	106
	REFERÊNCIAS	112

INTRODUÇÃO

A ideia de recuperação de sinais combinados (misturados) no processo de transmissão, a partir de combinações (misturas) obtidas por um certo número limitado de sensores, durante muitos anos foi dada como impossível. Todavia nas últimas décadas, ocorreram grandes avanços em sistemas de recuperação de informações como as técnicas avançadas de processamento de sinais, que permitem um melhor tratamento das informações obtidas no receptor de tal forma a explorar ao máximo as suas características originais, uma vez que contemplam na análise, diversos fatores como: distorções, ruídos, atenuações e interferências, que em geral, são responsáveis por fornecerem versões corrompidas dos sinais capturados pelos sensores. É bom salientar que a existência de interferências capturadas em um sensor, significa a presença de mistura entre diferentes fontes, que em alguns casos, como em gravações de músicas, tais misturas podem ser propositais [1] e [2].

A célebre expressão *cocktail party*, é um dos problemas passíveis de resolução pelas técnicas de Separação Cega de Fontes (BSS, do inglês *Blind Source Separation*). Tal problema se refere à capacidade humana de discernir em uma festa, os sons de interesse, como por exemplo, o som de uma determinada conversa, a partir do ruído difuso do ambiente [2]. Um exemplo bastante discutido sobre esta expressão é o caso de uma reunião ou uma festa. Neste caso, existem vários tipos de fontes, algumas delas possivelmente correlacionadas, que perturbam a identificação e compreensão de uma fonte específica. Situações como estas são apresentadas ao cérebro humano frequentemente, todavia devido à grande capacidade de processamento do cérebro humano, não é detectada nenhuma dificuldade para compreender um determinado interlocutor. Um esquema geral do tipo de informação proveniente de fontes diferentes que são captadas por sensores é ilustrada na Figura 1 [3].

A área de processamento de sinais biomédicos é outro problema típico de Separação Cega de Fontes, pois uma vez que o principal foco é obter o máximo de informações da maneira mais indolor, não invasiva e confiável possível, sobre a situação da saúde do paciente, a ideia de empregar técnicas que utilizam sensores sobre a superfície corporal, é bastante atrativa pelo fato de ser menos invasivo (e por conseguinte indolor). Na engenharia biomédica dois casos bastante estudados são: o eletrocardiograma materno/fetal (Figura 2 [3]) e o eletroencefalograma (ou magnetroencefalograma, Figura 3 [3]), onde para ambos os casos a utilização de técnicas de separação de fontes é fundamental para separar as fontes existentes nos sinais detectados pelos sensores [1], [4], [5], [6] e [7].

Além dos casos supracitados, as técnicas de separação de fontes têm sido alvo de grande

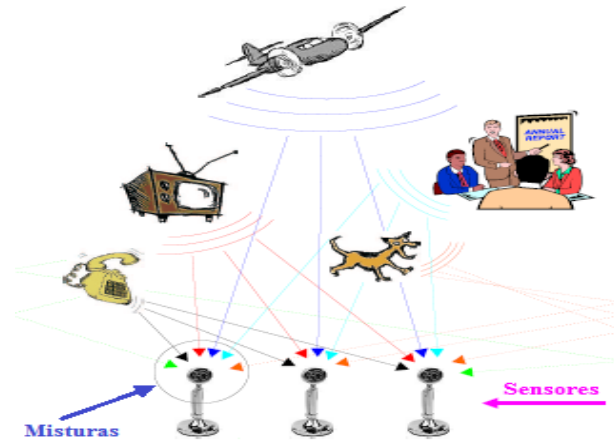


Figura 1 - Ilustração do problema de *cocktail party* [3].

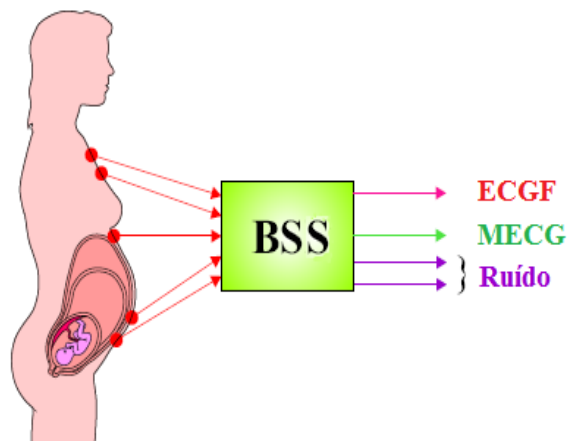


Figura 2 - Aplicação de BSS para medir o eletrocardiograma fetal/materno a partir de sensores na superfície corporal [3].

interesse da comunidade científica na última década, sendo empregadas em um grande número de aplicações tais como: sistemas de áudio [8], reconhecimento de fala [9], sensoriamento remoto, análise de sinais sísmicos [10], processamento de imagens [11] e comunicação digital [12], entre outros.

No âmbito de separação de sinais, costuma-se na literatura difundir o termo “cega” para fazer alusão ao fato de haver ausência de informações tanto das fontes quanto do sistema de mistura [2]. Em vista disso, as técnicas desenvolvidas para BSS não necessitam de nenhuma informação a priori das fontes [13].

Uma consideração fundamental em BSS é que as fontes são estatisticamente independentes [7]. E em geral considera-se também, distribuição não gaussiana (frequentemente supergaussiana) das fontes e distribuição uniforme dos coeficientes da matriz de mistura [2]

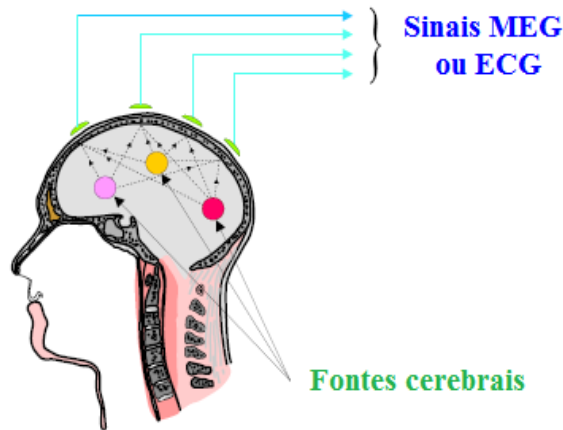


Figura 3 - Método de BSS para aquisição do eletroencefalograma/magnetoencefalograma [3].

e [14].

Dentre as configurações lineares de misturas, as mais desafiadoras são as convolutivas (foco desta dissertação), as quais, num contexto de fontes sonoras, levam em conta a reverberação de um ambiente ecoico. Nestes casos, tipicamente filtros de separação de resposta ao impulso finita (FIR, do inglês *Finite Impulse Response*) com milhares de coeficientes são necessários, tornando muito complexa a tarefa de separação. Para tentar resolver este problema, diversos métodos têm sido propostos na literatura. Enquanto alguns destes realizam a BSS no domínio do tempo [14], outros a efetuam no domínio da frequência [7], costumeiramente almejando melhoria de desempenho ou propondo redução da complexidade computacional [15].

A Análise de Componentes Independentes (ICA, do inglês *Independent Component Analysis*) é um método para encontrar estatisticamente fontes independentes a partir de misturas, utilizando estatísticas de alta ordem. De maneira simples, o método assume que as misturas são lineares e instantâneas. Todavia em um ambiente acústico real estas considerações não são válidas, necessitando de um modelo mais extensivo, porque os sinais das fontes não são misturas instantâneas, mas sim convolutivas, como por exemplo, na separação de sinais de áudio onde variações do método ICA devem ser utilizados [7] e [16].

A técnica de BSS escolhida é uma implementação no domínio do tempo, que emprega estatísticas de segunda ordem para a separação de misturas convolutivas e determinadas [14], considerando a presença e a ausência de instantes de silêncio em misturas de sinais de voz, utilizando detectores de voz como uma etapa de pré-processamento.

A escolha do sistema de separação pode ser orientada por meio de uma função de custo

proposta em [14], que depende da configuração da mistura, sendo frequentemente minimizada com métodos do tipo gradiente.

Para a avaliação do desempenho das diversas configurações apresentadas utilizam-se as métricas: Relação Fonte-Interferência (SIR, do inglês *Source-to-Interference Ratio*), Relação Fonte-Artefato (SAR, do inglês *Source-to-Artifact-Ratio*) e Relação Fonte-Distorção (SDR, do inglês *Source-to-Distortion-Ratio*), calculadas como descrito em [17], considerando em todos os casos os valores médios entre as métricas das duas estimativas.

Em aplicações de processamento de sinais de voz, a identificação dos intervalos (trechos) contendo a fala, existentes em um sinal de voz, costuma ser explorada, uma vez que a informação existe somente em alguns intervalos (ou regiões) nos quais existe a fala, a qual pode ser considerada em virtude disso, descontínua.

Nesta dissertação empregamos detectores de voz nas misturas, com a finalidade de aplicar o algoritmo de separação cega apenas nos intervalos identificados como trechos contendo a fala, promovendo desta forma uma melhoria de desempenho e redução de custo computacional, visto que o algoritmo de BSS é desativado nos trechos em que a atividade vocal não é detectada, denominados trechos de silêncio ou de pausa [15] e [18].

Estruturas multitaxas bem como banco de filtros são amplamente explorados na literatura dentre algumas aplicações podemos citar: filtragem adaptativa em sub-bandas, codificação e compressão de sinais de voz e imagens, redução e/ou expansão da taxa de amostragem, sistemas de antenas, entre outros [19], [20] e [21]. Visando explorar as vantagens dos sistemas multitaxas, adaptamos a técnica de BSS escolhida neste trabalho, e investigamos a influência do sistema multitaxa que permite separar o sinal em bandas distintas e aplicar a BSS em cada sub-banda isoladamente.

Em virtude da complexidade do tema abordado, estruturamos os objetivos deste trabalho da seguinte maneira, no Capítulo 1, introduzimos os conceitos básicos de separação cega de fontes, bem como as propriedades e problemas que costumam ser explorados na literatura, também contemplamos a forma de avaliação de desempenho empregada neste trabalho. No Capítulo 2, enfatizamos a técnica de BSS no domínio do tempo utilizada nesta dissertação, explicitamos os fundamentos teóricos e as formulações matemáticas empregadas na técnica escolhida. No Capítulo 3 apresentamos diferentes tipos de detectores de voz investigados neste trabalho, enfocando no emprego do detector de voz utilizado nesta dissertação como etapa de pré-processamento. No Capítulo 4 contemplamos a abordagem de sistemas multitaxas, neste contexto introduzimos as estruturas em sub-bandas que empregam bancos de filtros modulados por cosseno tanto em banco uniforme quanto em banco não uniforme,

também expomos a implementação de tais estruturas em sub-bandas como ferramenta para o algoritmo de separação cega de fontes no domínio do tempo. No Capítulo 5 descrevemos os testes realizados bem como os resultados obtidos a partir de simulações computacionais para a implementação *offline* e o custo computacional do algoritmo. Finalmente no Capítulo 6 contemplamos estratégias para melhorar o desempenho do algoritmo ou reduzir o custo computacional proporcionando a continuação deste trabalho. As observações obtidas no decorrer desta proposta são mencionadas na Conclusão deste trabalho.

1 ABORDAGEM GERAL SOBRE OS MÉTODOS DE SEPARAÇÃO CEGA DE FONTES

1.1 Introdução

Este capítulo apresenta os conceitos básicos sobre os métodos de separação cega de fontes, abordando inicialmente as configurações de misturas e os sistemas de mistura e separação. Também apresenta as propriedades e as dificuldades dos métodos de separação de fontes existentes na literatura bem como o método contemplado nesta dissertação.

1.2 Configurações de Mistura

A configuração dos tipos de misturas, de forma geral, pode ser classificada em dois grandes grupos: lineares e não-lineares. Dentre o grupo de misturas lineares pode-se subdividi-lo em dois casos: misturas instantâneas e misturas convolutivas. Uma mistura é considerada convolutiva devido aos atrasos que resultam da propagação do som através do espaço e do fenômeno de múltiplos percursos engendrado por reflexões do som em diferentes objetos (reverberação).

Outra classificação de misturas refere-se à relação entre o número de fontes e o número de misturas (ou sensores). Neste aspecto podemos dividi-las em três casos: sobredeterminadas (quando o número de misturas supera o número de fontes), determinadas (quando o número de misturas é igual ao número de fontes) e subdeterminadas (quando o número de misturas é inferior ao número de fontes) [15]. O aumento do número de misturas em relação ao número de fontes implica uma maior facilidade para as técnicas de separação, pois significa aumento da informação disponível (diversidade espacial) [2]. Esta dissertação contempla apenas os casos de misturas lineares convolutivas e determinadas.

1.3 Sistemas de mistura e de separação

Seja Q o número de fontes, P o número de sensores e $s_q(n)$ a q -ésima fonte, para $q = 1, \dots, Q$, onde n é o índice amostral. O vetor $s(n)$, que compreende todas as fontes $s_q(n)$ no instante n é dado por $s(n) = [s_1(n) \ s_2(n) \ \dots \ s_Q(n)]^T$, onde $(.)^T$ significa transposição. Analogamente, considerando que $x_p(n)$ seja a p -ésima mistura para $p = 1, \dots, P$, pode-se expressar o vetor $x(n)$ que contém os sinais capturados pelos P sensores no instante n , como: $x(n) = [x_1(n) \ x_2(n) \ \dots \ x_P(n)]^T$. A relação entre os vetores $x(n)$ e $s(n)$ para um sistema de

misturas lineares é dada por:

$$x(n) = \mathbf{H} * s(n), \quad (1)$$

em que “*” significa convolução (para o caso de misturas lineares instantâneas podemos substituir o símbolo de convolução pelo produto matricial usual) e \mathbf{H} é uma matriz de dimensão $P \times Q$ denominada matriz de mistura que oriunda do sistema de mistura, que pode ser definida como:

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1Q} \\ h_{21} & h_{22} & \cdots & h_{2Q} \\ \vdots & \vdots & \vdots & \vdots \\ h_{P1} & h_{P2} & \cdots & h_{PQ} \end{bmatrix}, \quad (2)$$

sendo os filtros h_{pq} , elementos da matriz \mathbf{H} , correspondem aos múltiplos caminhos percorridos pelo sinal até serem capturados pelos sensores, e determinam o quanto da fonte $s_q(n)$ está presente na mistura $x_p(n)$. Para sinais de áudio, estes filtros refletem a reverberação do ambiente. O esquema de um sistema de separação cega de fontes é apresentado na Figura 4 [2].

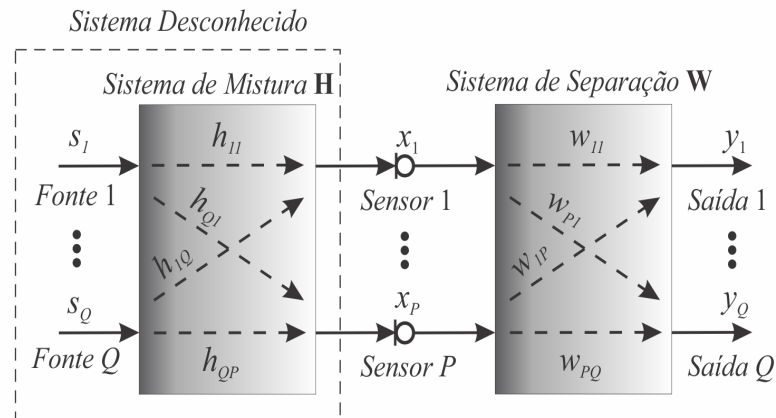


Figura 4 - Ilustração de um esquema simples de BSS.

Conforme mencionado anteriormente as técnicas de separação cega de fontes têm como objetivo recuperar as fontes existentes em uma mistura sem nenhum conhecimento a priori das fontes originais nem do sistema de mistura \mathbf{H} . Neste contexto, definimos ainda uma matriz \mathbf{W} e o sinal $y_q(n)$, que correspondem respectivamente à matriz de separação de

dimensão $Q \times P$ e à estimativa do sinal $s_q(n)$. Pode-se concatenar todas as saídas $y_q(n)$, em um vetor $y(n)$ que compreende todas as estimativas das fontes $y_q(n)$ no instante n por $y(n) = [y_1(n) \ y_2(n) \ \dots \ y_Q(n)]^T$. A relação entre $y(n)$ e \mathbf{W} é:

$$y(n) = \mathbf{W} * x(n). \quad (3)$$

Em geral, algumas considerações iniciais¹ sobre as fontes e misturas são feitas, tais como: independência estatísticas entre as fontes originais, não-estacionariedade, esparcidade, entre outras, logo nenhuma das técnicas de separação é absolutamente cega.

Considerando apenas sistemas causais, cabe ressaltar que caso a mistura seja instantânea, os coeficientes das matrizes dos sistemas de mistura e de separação, são reduzidas a constantes.

1.4 Misturas Lineares e Instantâneas

Para misturas lineares e instantâneas podemos substituir o símbolo de convolução da Equação (1) pelo produto matricial usual. Tal substituição deve-se ao fato do modelo para misturas lineares e instantâneas não modelar os atrasos que resultam da propagação do som através do espaço (atraso entre os sensores), nem o fenômeno de múltiplo caminhos (*multipath*) gerado por reflexões do som em diferentes objetos (reverberação). Por este motivo, considera-se que o modelo para misturas lineares e instantâneas é mais simples, em relação ao modelo de misturas lineares e convolutiva. Desta forma, podemos expressar a Equação (1) por:

$$x(n) = \mathbf{H}s(n). \quad (4)$$

Analogamente, a Equação (3) pode ser interpretada como:

$$y(n) = \mathbf{W}x(n) = \mathbf{W}\mathbf{H}s(n) = \mathbf{C}s(n), \quad (5)$$

em que a matriz $\mathbf{C}=\mathbf{W}\mathbf{H}$ condensa os efeitos de mistura e separação. A Equação (5) permite concluir o conceito de *separabilidade*² das fontes [22] que de modo geral, refere-se ao fato de existir uma matriz \mathbf{W} que torna possível separar as misturas. Para o caso em que a

¹Algumas destas considerações serão apresentadas nas próximas seções deste capítulo.

²Separabilidade de fontes é uma propriedade intrínseca das combinações lineares nos sinais medidos, ou seja, depende da estrutura da matriz \mathbf{H} [22].

matriz de mistura tenha posto Q , pode-se dizer que a mistura é separável, o que resulta na possibilidade de existir uma matriz de separação \mathbf{W} tal que a matriz \mathbf{C} seja uma matriz identidade, ou seja, $\mathbf{C}=\mathbf{I}$ tornando assim possível encontrar $s(n)$ a partir de $y(n)$, desde que \mathbf{W} seja escolhida adequadamente. Em [22] é apresentada uma possível solução para obter a separação das fontes nos casos em que \mathbf{W} tenha posto menor que Q .

Não obstante, mesmo que a mistura seja separável não é correto afirmar que a matriz \mathbf{C} será uma matriz identidade, uma vez que desconhecemos as fontes e a matriz de mistura, pode ocorrer, dentre os problemas mais usuais, que a ordem das fontes no vetor $s(n)$ não corresponda com a ordem das saídas em $y(n)$, tal problema é denominado permutação, ou então, que haja alguma ambiguidade de escalamento. A Figura 5 [23] apresenta estes dois problemas de separação de fontes (permutação e escalamento), cujas descrições serão apresentadas na seção 1.8.

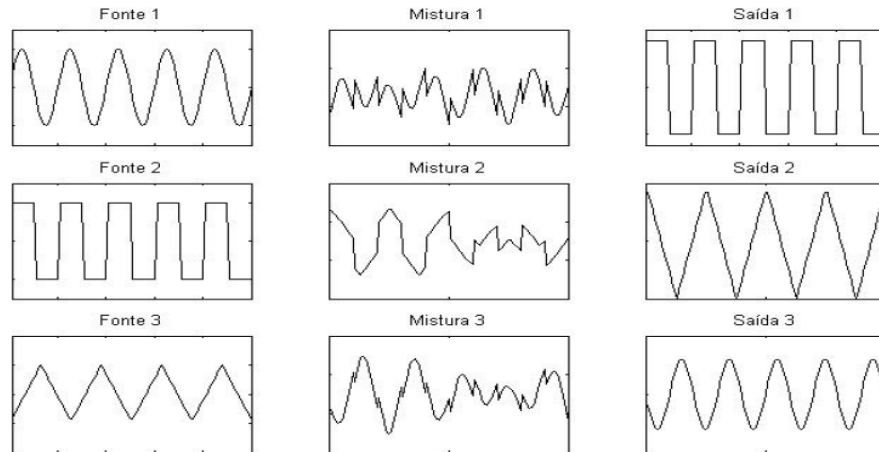


Figura 5 - Problemas comuns em separação cega de fontes [23].

Logo, faz-se necessário ao considerar a eficiência de separação não somente se a matriz \mathbf{C} é uma matriz identidade, mas também, devido a problemas de permutação e escalamento, quando uma matriz diagonal é encontrada. Desta maneira, procuramos uma matriz \mathbf{C} semelhante a matriz apresentada na Equação (6) em que $P = Q = 3$ e cuja ideia pode ser facilmente expandida para qualquer $P = Q$:

$$\mathbf{C} \approx \begin{bmatrix} 0 & a & 0 \\ 0 & 0 & b \\ c & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} \quad (6)$$

sendo que $abc \neq 0$ é uma solução aceitável para a separação das fontes, tanto quanto $\mathbf{C} \approx \mathbf{I}$

(o motivo de usar o símbolo de aproximação, refere-se ao fato de na prática, a igualdade entre a matriz \mathbf{C} e a matriz identidade não necessariamente costuma ocorrer, provavelmente devido à algum problema de separação cega de fontes, vide seção 1.8).

1.5 Misturas Lineares e Convolutiva

Devido à reverberação presente em um ambiente acústico, assumindo que as fontes são independentes entre si, pode-se considerar que os sinais das fontes $s_q(n)$ são filtrados por um sistema linear de mistura \mathbf{H} de múltiplas entradas e múltiplas saídas (MIMO, do inglês *Multiple-Input Multiple-Output*), antes de serem capturados pelos sensores. O esquema geral de um sistema de separação de fonte cega de fontes pode ser observado na Figura 6, através de uma adaptação de [14], considerando $P = Q$.

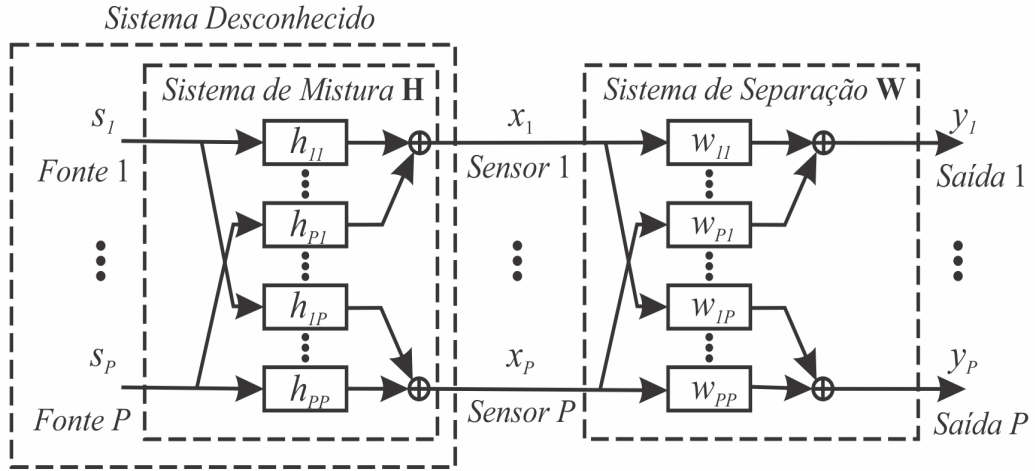


Figura 6 - Estrutura geral do sistema de mistura e separação para o problema de separação cega.

Desta forma o modelo de misturas lineares e convolutivas para o caso determinado, empregado nesta dissertação, para cada uma das p misturas, pode ser representado por [14]:

$$x_p(n) = \sum_{q=1}^P \sum_{\kappa=0}^{U-1} h_{qp}(n) s_q(n - \kappa), \quad (7)$$

onde $h_{qp}(\kappa)$, $\kappa = 0, \dots, U-1$ representa a resposta ao impulso de um filtro FIR de comprimento U que modela o caminho acústico (eco) da q -ésima fonte até a p -ésima mistura.

O objetivo das técnicas de BSS é encontrar os filtros que compõem o sistema de separação \mathbf{W} (conforme a Figura 6), que proporcionam obter, em suas saídas, versões filtradas das fontes originais, tendo acesso apenas aos sinais misturados. As saídas (ou estimativas)

do sistema de separação podem ser representadas por [14]:

$$y_q(n) = \sum_{p=1}^P \sum_{\kappa=0}^{S-1} w_{pq}(\kappa) x_p(n - \kappa), \quad (8)$$

onde $w_{pq}(\kappa)$, para $\kappa = 0, \dots, S - 1$, representa os coeficientes de um filtro FIR de separação de comprimento S , onde idealmente, para modelar todas as reflexões, o parâmetro S deve ser escolhido de tal maneira para que no mínimo seja igual a U , não obstante, considerando que as características do meio acústico são desconhecidas, a escolha de U nesta dissertação, foi encontrada de forma empírica, em outras palavras, diversas simulações do algoritmo foram necessárias de modo a encontrar um valor de U para que o desempenho da separação seja o melhor possível de modo a evitar o *overfitting* e o *underfitting* [14]. Conforme explicado anteriormente, a consideração de independência mútua das fontes, possibilita que as estimativas das fontes também sejam forçadas a ser estatisticamente independentes entre si.

Nesta dissertação empregamos um método de separação que por sua vez, utiliza estatísticas de segunda ordem, conforme veremos no capítulo seguinte, para estimar os P^2S coeficientes de $w_{pq}(\kappa)$.

1.6 Misturas Não Lineares

Métodos de BSS para misturas não lineares [24], [25] e [26], embora pouco discutidos na literatura, são desenvolvidos para modelar sistemas reais, e em geral, consistem em dividir o problema de separação em duas etapas, uma linear e outra não linear, conforme é apresentado na Figura 7, adaptação de [1].

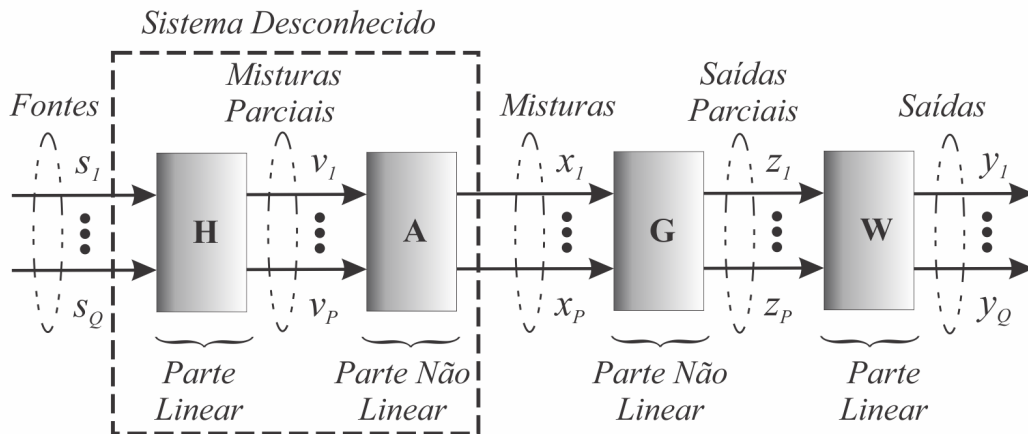


Figura 7 - Estrutura de um sistema de separação para misturas não lineares.

Em geral, costuma-se considerar misturas instantâneas para modelar o sistema apresentado na Figura 7, na qual a etapa de mistura pode ser subdividida em duas partes, a primeira linear e a segunda não linear, o que nos permite expressar as misturas por:

$$x(n) = \mathbf{A}(v(n)) = \mathbf{A}(\mathbf{H}s(n)), \quad (9)$$

sendo $s(n)$ e $x(n)$ os vetores que compreendem respectivamente os sinais das fontes e os sinais capturados pelos P sensores no instante n , \mathbf{H} a matriz de mistura de dimensão $P \times Q$ responsável pela parte linear do modelo descrito e \mathbf{A} um vetor que pode ser expresso por $\mathbf{A} = [a_1 \ a_2 \ \dots \ a_P]^T$, em que cada elemento representa uma função de transferência não linear.

Analogamente à etapa de mistura, a etapa de recuperação dos sinais das fontes, consiste em duas partes, na qual a primeira parte (não linear) almeja encontrar (estimar) uma função inversa não linear capaz de inverter o processo não linear realizado na etapa de mistura, e a segunda parte visa realizar um processo linear, similar às técnicas de BSS empregadas em misturas lineares (geralmente ICA [27]), podemos então expressar a q -ésima saída por:

$$y_q(n) = \sum_{p=1}^P w_{qp} g_p(x_p(n)), \quad (10)$$

em que $g_p(x_p(n))$ representa o p -ésimo elemento de \mathbf{G} .

Estimar $g_p(x_p(n))$ não é uma tarefa trivial devido a diversos fatores como: as não linearidades do sinal, a ausência de informação a priori e as distorções e ruídos captados pelos sensores, todavia, encontram-se na literatura alguns métodos que empregam redes neurais artificiais [28] e outros utilizam algoritmos genéticos [29] que visam estimar $g_p(x_p(n))$.

Em consequência disso, a complexidade computacional dos algoritmos para misturas não lineares é superior em relação aos algoritmos para misturas lineares devido ao elevado número de parâmetros que necessitam ser estimados. Além disso, o elevado número de parâmetros a ajustar associado à otimização de funções não lineares, que compreendem o vetor \mathbf{G} , podem algumas vezes levar o algoritmo de separação à mínimos locais [27].

1.7 Conceitos e Propriedades de Métodos de Separação Cega de Fontes

Nesta seção abordaremos os principais conceitos e propriedades amplamente explorados na literatura pelos métodos de separação cega para fontes de sinais acústicos, em particular, para sinais de fala. Isso porque os sinais de fala são sinais que possuem diversas características próprias, sendo que algumas destas podem auxiliar na separação de tais sinais

através de métodos de BSS.

1.7.1 Independentes estatisticamente das fontes entre si

Um pressuposto básico nos métodos de BSS na literatura é a independência estatística das fontes entre si. Em [1] temos que sinais das fontes são considerados independentes entre si, quando a Função de Densidade de Probabilidade (PDF, do inglês *Probability Density Function*) conjunta do vetor $s(n)$ é igual ao produto das PDFs marginais dos q -ésimos componentes do vetor $s(n)$. Podemos então expressar a relação entre as PDFs conjunta e marginais como:

$$p_{\bar{s}}(s) \triangleq \prod_{q=1}^Q p_{\bar{s}_q}(\bar{s}_q). \quad (11)$$

Tal consideração possibilita que as estimativas das fontes, ou seja, o vetor $y(n)$, oriundos do processo de separação das fontes, também sejam forçadas a apresentar a mesma relação entre as PDFs conjunta e marginais para que a separação seja assegurada [1]. Em outras palavras, considerar que as fontes originais são estatisticamente independentes entre si, faz com que as fontes estimadas também sejam estatisticamente independentes entre si. Sendo assim podemos expressar as PDFs das saídas estimadas como:

$$p_{\bar{y}}(y) \triangleq \prod_{q=1}^Q p_{\bar{y}_q}(\bar{y}_q). \quad (12)$$

O conceito de independência estatística das fontes entre si é amplamente utilizado em técnicas de separação que empregam misturas tanto instantâneas quanto convolutivas [30]. E na prática, significa considerar que os sinais de fala são originados por diferentes fontes em lugares (posições) distintos em um ambiente acústico [31].

1.7.2 Não Branqueamento (*nonwhiteness*)

Alguns métodos de BSS exploram o fato dos sinais de áudio apresentarem dependências temporais. Desta forma, embora as amostras dos sinais originados de diferentes fontes sejam consideradas independentes entre si, pode-se assumir que amostras de cada fonte de sinal não são independentes ao longo do eixo do tempo, caracterizando assim um sinal não branco (ou colorido) [32].

A propriedade de não branqueamento do sinal é explorada por métodos de BSS através

da diagonalização de matrizes de correlação calculadas a partir dos sinais das estimativas (saídas) com diferentes atrasos de tempo (*lags*) [33] e [34].

1.7.3 Não gaussianidade (*nongaussianity*)

As amostras de um sinal de fala possuem uma distribuição chamada na literatura de não gaussiana (ou supergaussiana), tal distribuição significa que os coeficientes espectrais de sinais de fala são bem distribuídos [35].

A propriedade de um sinal com distribuição considerada não gaussiana é explorada por métodos de BSS que empregam estatísticas de ordem superior, na literatura, em geral é empregada por algoritmos de Análise de Componentes Independentes (ICA).

Na Figura 8 são apresentados exemplos de distribuições supergaussiana, gaussiana e subgaussiana.

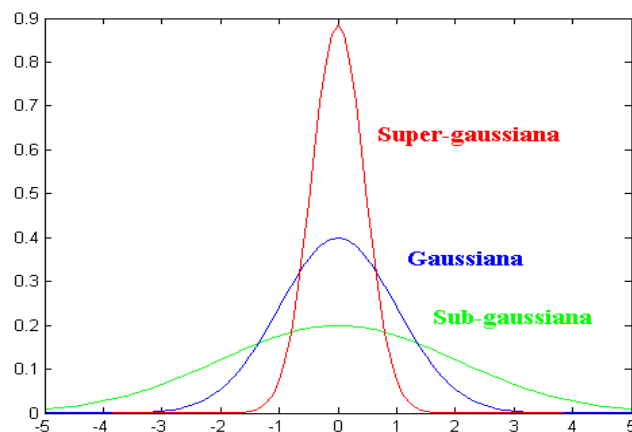


Figura 8 - Distribuição supergaussiana (em vermelho), gaussiana (em azul) e subgaussiana (em verde).

1.7.4 Não Estacionária (*nonstationarity*)

O sinal de fala é um sinal não-estacionário devido às modulações de amplitude inerentes aos trechos que contem a voz e trechos em que há combinações de voz e silêncio (ausência de voz) [31].

A propriedade de não estacionariedade do sinal é explorada por métodos de BSS através da diagonalização simultânea de matrizes de correlação de tempo curto (*short-time correlation*) das estimativas (saídas) em diferentes instantes de tempo [34].

1.7.5 Banda Estreita versus Banda Larga

Na literatura, com o propósito de resolver o problema de BSS, a análise realizada pelo algoritmo de separação pode ser feita em banda estreita ou banda larga³.

É comum que algoritmos de BSS implementados no domínio da frequência realizem a análise de separação independentemente em cada faixa de frequência, motivo pelo qual é denominado de “banda estreita”. Em geral, os algoritmos que empregam a análise em banda estreita empregam estatísticas de alta ordem para realizar a separação e costumam apresentar algumas limitações conforme identificadas em [36] e [37], dentre elas o problema de permutação devido ao branqueamento do sinal.

Por outro lado, os algoritmos de BSS, que realizam a análise de separação contemplando as dependências temporais do sinal, ou seja, através do não branqueamento do sinal, são denominados de banda larga e costumam empregar estatísticas de segunda ordem para realizar a separação.

No capítulo 2, veremos que para sinais de voz, alguns destes conceitos e propriedades serão exploradas pelo método de BSS empregado nesta dissertação, pois possibilitam diagonalizar a matriz de correlação de saída, considerando múltiplos blocos em diferentes instantes de tempo (TDD, do inglês *Time-Delayed Decorrelation*) [15] e [33].

1.8 Problemas dos Métodos de Separação Cega de Fontes

Nesta seção explicaremos alguns dos problemas típicos encontrados nos métodos de separação de fontes que podem ocorrer durante o processo de separação, uma vez que o conceito básico para tais métodos, baseia-se somente na suposição de independência mútua entre os sinais originais e que as fontes originais, bem como o sistema de mistura são desconhecidos podemos citar três problemas comuns encontrados na literatura: permutação, escalamento e filtragem.

Existe também, um problema não muito citado na literatura, que é o fato de métodos de separação cega proporcionarem uma alteração na fase, no sinal estimado, todavia para muitos casos tal problema pode ser desconsiderado [38].

Embora tais problemas resultem em restrições para algumas aplicações [39], cabe ressaltar que para um caso real em que não temos informação alguma sobre as fontes originais e cujo objetivo é estimar as fontes a partir das misturas, é indiferente saber a ordem em que as fontes são estimadas ou a amplitude dos sinal estimado, uma vez que na maioria das

³Na literatura, encontra-se também a expressão “banda cheia” com o mesmo sentido de “banda larga”.

aplicações a informação relevante está na forma de onda [40], sendo que esta última para algumas aplicações ainda pode estar comprometida pelo problema de filtragem.

1.8.1 Permutação

Uma vez que desconhecemos a ordem no vetor $s(n)$ das fontes originais, não se pode determinar a ordem dos sinais estimados em $y(n)$, em outras palavras, não é possível localizar a posição na saída do sistema de separação de uma certa fonte que foi estimada, sem nenhuma informação sobre as fontes originais.

1.8.2 Escalamento

As variâncias (energias) das componentes independentes não podem ser encontradas, visto que não conhecemos as fontes originais nem o sistema de mistura. Isso pode proporcionar um problema denominado na literatura conhecido como escalamento (ou ambiguidade do escalamento ou ainda fator de escala), tal problema é ocorre pois caso uma fonte $s_q(n)$ seja multiplicada por um fator k , e se os elementos da q -ésima coluna da matriz de mistura \mathbf{H} , sejam multiplicados por um fator $1/k$, a mistura $x(n)$ que é a única informação disponível é a mesma, tornando assim impossível de determinar o valor de k .

1.8.3 Filtragem

Os sinais estimados são versões distorcidas (filtradas) das fontes originais, logo a escolha dos filtros utilizados, bem como o comprimento dos mesmos, são fundamentais para se obter um bom resultado no sinal estimado, tornando-o mais semelhante ao sinal original.

1.9 **Métodos de avaliação de desempenho**

Na literatura, diversas formas são empregadas para avaliar o desempenho do método de separação, nesta dissertação, utilizaremos três métricas bastante comuns, a Relação Fonte-Interferência (SIR), a Relação Fonte-Artefato (SAR) e a Relação Fonte-Distorção (SDR) conforme proposto em [17]. Na qual para todos os casos, são mensurados, nesta dissertação, os valores médios entre as métricas das duas estimativas encontradas.

De maneira suscita a ideia apresentada em [17] decompõem as saídas estimadas da seguinte maneira:

$$y_q(n) = [s_{target}]_q(n) + [e_{interf}]_q(n) + [e_{noise}]_q(n) + [e_{artif}]_q(n), \quad (13)$$

em que s_{target} é o sinal desejado com alguma distorção tolerável (em outras palavras, é a fonte original, considerando que a mesma é conhecida), e_{interf} é o erro resultante das interferências provocadas pelas demais fontes, e_{noise} é o erro resultante do ruído dos sensores (microfones), e_{artif} é o erro resultante dos artefatos (ou seja, é o erro proveniente de todas as distorções do sinal não resultantes de interferências nem de ruído), calculados conforme [17].

Dentre as métricas empregadas, conforme apresentado em [17] podemos definir:

- SIR como a métrica empregada para mensurar a razão entre o sinal desejado e as distorções causadas pelo ruído;
- SAR como a métrica empregada para mensurar a razão entre o sinal de saída (contendo além da fonte original, interferências e ruído) e os artefatos;
- SDR como a métrica empregada para mensurar a razão entre o sinal desejado e todas as aludidas distorções (interferências, ruído e artefatos).

Tais métricas são apresentadas em [17] respectivamente como:

$$SIR_q \triangleq 10 \log_{10} \frac{\|[s_{target}]_q(n)\|^2}{\|[s_{interf}]_q(n)\|^2}, \quad (14)$$

$$SAR_q \triangleq 10 \log_{10} \frac{\|[s_{target}]_q(n) + [e_{interf}]_q(n) + [e_{noise}]_q(n)\|^2}{\|[e_{artif}]_q(n)\|^2}, \quad (15)$$

$$SDR_q \triangleq 10 \log_{10} \frac{\|[s_{target}]_q(n)\|^2}{\|[e_{interf}]_q(n) + [e_{noise}]_q(n) + [e_{artif}]_q(n)\|^2}. \quad (16)$$

1.10 Métodos de Separação Cega de Fontes

Uma vez que consideramos que existe uma matriz \mathbf{W} capaz de resolver o problema de separação das misturas, é preciso encontrar algum método para determiná-la. Para tal função, uma diversidade de métodos são amplamente empregados na literatura, tanto no domínio do tempo [14], que será empregado nesta dissertação, quanto no domínio da frequência [7]. Em geral, algoritmos de separação atuam na maximização (ou minimização) de uma função de custo.

A seguir, abordaremos uma visão geral dos conceitos básicos de uma técnica no domínio da frequência amplamente explorada na literatura, em geral, para misturas lineares e instantâneas, e no capítulo seguinte um detalhamento da técnica no domínio do tempo empregada nesta dissertação, que visa a separação cega de fontes para misturas lineares e convolutivas, empregando sinais de voz.

1.10.1 Técnicas de Análise de Componentes Independentes

As técnicas de BSS que empregam a Análise de Componentes Independentes, têm como objetivo encontrar a matriz de separação \mathbf{W} , de forma a obter nas saídas estimadas, sinais estatisticamente independentes, para isso, tais métodos costumam fazer as seguintes considerações:

- As fontes originais devem ser estatisticamente independentes;
- As fontes devem ser não-gaussianas ou no máximo, apenas uma fonte pode ser gaussiana.

Em geral, as técnicas de ICA empregam estatísticas de alta ordem para resolver o problema de separação e são aplicadas em misturas lineares e instantâneas e também em sistemas determinados, todavia adaptações podem ser feitas para aplicar tais técnicas tanto para misturas lineares e convolutivas, quanto para sistemas subdeterminados e overdeterminados [41].

É comum técnicas de ICA, utilizarem a informação mútua [42] ou a kurtosis [43] como funções de custo. Duas técnicas comuns de BSS que empregam técnicas de ICA encontradas na literatura são: Infomax [42] e FastICA [43].

2 MÉTODO DE SEPARAÇÃO CEGA DE FONTES NO DOMÍNIO DO TEMPO

2.1 Introdução

Neste capítulo descreveremos um método de separação cega de fontes no domínio do tempo utilizado nesta dissertação, que emprega estatísticas de segunda ordem para a separação de misturas convolutivas e determinadas proposto em [14], o qual é bastante atrativo pela robustez aos problemas de branqueamento das estimativas e de permutação das saídas quando comparados a métodos no domínio da frequência. Além disso, utilizamos nesta dissertação sinais de áudio, que devido à reverberação do ambiente, proporcionam misturas conhecidas como misturas convolutivas. Este método no domínio do tempo é conhecido na literatura como um método em banda cheia, todavia no capítulo 4 faremos uma adaptação neste método para utilizá-lo em uma abordagem em sub-banda.

O método proposto em [14] parte do princípio de que as fontes originais são estatisticamente independentes e conforme mencionado anteriormente, tal consideração força as saídas a também ser estatisticamente independentes. E assim, as fontes originais são filtradas por um sistema MIMO, apresentado na Figura 9 (adaptação de [19]), em que o sistema de mistura \mathbf{H} é desconhecido e cujo objetivo do algoritmo de BSS é encontrar o sistema de separação \mathbf{W} . Na Figura 9, vemos como o sistema de separação é atualizado pelo algoritmo de BSS.

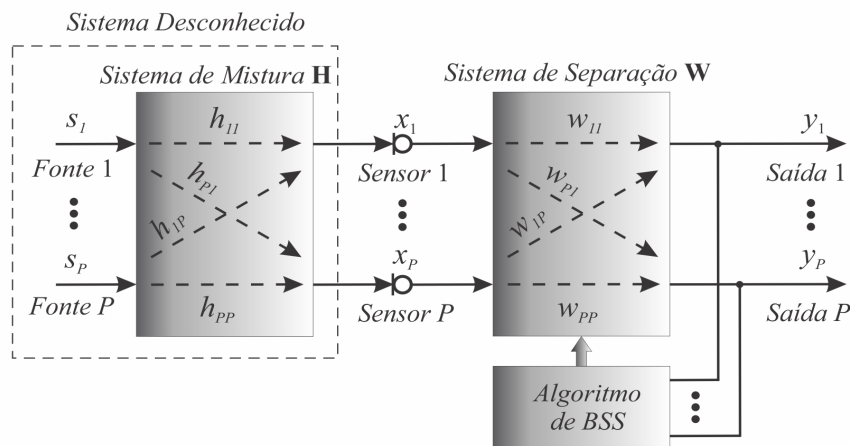


Figura 9 - Estrutura de um sistema de separação cega completo [19].

2.2 Formulação Matricial para Misturas Convolutivas

O algoritmo de BSS proposto em [14], empregado nesta dissertação, bem como o sistemas de mistura e separação podem ser representados na forma matricial, com a finalidade de facilitar tanto a notação quanto a dedução do algoritmo. Assim, podemos adaptar a Equação (8) para representar os sinais estimados $y_q(n)$, para $q = 1, \dots, Q$, do sistema de separação no formato matricial, para um determinado instante de tempo n , conforme apresentado na Equação (17):

$$\mathbf{y}_q(n) = \sum_{p=1}^P \mathbf{x}_p^T(n) \mathbf{w}_{pq}, \quad (17)$$

em que \mathbf{x}_p representa um vetor que contém as últimas S amostras (em relação ao instante n) capturadas pelo p -ésimo sensor, em outra palavras:

$$\mathbf{x}_p(n) = \begin{bmatrix} x_p(n) & x_p(n-1) & \cdots & x_p(n-S+1) \end{bmatrix}^T, \quad (18)$$

e \mathbf{w}_{pq} representa um vetor que contém os S coeficientes do filtro que modela o caminho acústico (eco) do p -ésimo sensor até a q -ésima fonte estimada, ou seja:

$$\mathbf{w}_{pq} = \begin{bmatrix} w_{pq}(0) & w_{pq}(1) & \cdots & w_{pq}(S-1) \end{bmatrix}^T. \quad (19)$$

Para generalizar a formulação podemos introduzir dois novos parâmetros:

- O parâmetro D que representa o número de atrasos de tempo considerados para o cálculo das correlações presentes na função de custo, como veremos na seção seguinte, explorando as dependências temporais dos sinais, e por conseguinte, a propriedade de não branqueamento do sinal. O parâmetro de D deve ser escolhido de tal forma que $1 \leq D \leq S$ [13], [14] e [34];
- O parâmetro N que representa o comprimento dos blocos dos sinais das saídas resultante de um particionamento dos mesmos, que são empregados para calcular as matrizes de correlação das saídas em diferentes instantes de tempo, explorando assim, a propriedade de não estacionariedade do sinal. Em que $N \geq PD$ é a condição necessária para se obter uma matriz de correlação $\mathbf{R}_{\mathbf{y}\mathbf{y}}$ de posto completo, conforme poderá ser analisado na seção seguinte [13] e [34].

Podemos estender a Equação (17) para incluir os blocos de comprimento N , e assim

temos:

$$\mathbf{y}_q(m) = \begin{bmatrix} y_q(mS) & y_q(mS - 1) & \cdots & y_q(mS + N - 1) \end{bmatrix}^T, \quad (20)$$

sendo m o índice de tempo do bloco analisado e o vetor $\mathbf{y}_q(m)$ contém um bloco referente ao particionamento da q -ésima saída. Podemos então, associar a saída em função das amostras capturadas pelos sensores e dos filtros de separação da seguinte forma:

$$\mathbf{y}_q(m) = \sum_{p=1}^P \tilde{\mathbf{X}}_p^T(m) \mathbf{w}_{pq}, \quad (21)$$

em que $\tilde{\mathbf{X}}_p(m)$ pode ser escrito da seguinte forma compacta:

$$\tilde{\mathbf{X}}_p(m) = \begin{bmatrix} \mathbf{x}_p(mS) & \mathbf{x}_p(mS - 1) & \cdots & \mathbf{x}_p(mS + N - 1) \end{bmatrix}, \quad (22)$$

ou ainda, como uma matriz expandida de dimensão $S \times N$:

$$\tilde{\mathbf{X}}_p^T(m) = \begin{bmatrix} x_p(mS) & \cdots & x_p(mS - S + 1) \\ x_p(mS - 1) & \ddots & x_p(mS - S + 2) \\ \vdots & \ddots & \vdots \\ x_p(mS + N - 1) & \cdots & x_p(mS - S + N) \end{bmatrix}. \quad (23)$$

Analisando a Equação (23) podemos observar que a matriz $\tilde{\mathbf{X}}_p(m)$ possui uma estrutura denominada de matriz de *Toeplitz*⁴ para todo p .

Cabe ressaltar que a escolha de particionamento do sinal em blocos, apresentado nesta dissertação (proposto em [14]) ocorre por meio de sobreposição. Tal sobreposição equivale a substituir o índice de bloco mS por $m(S/\alpha)$ em que α é um fator de sobreposição que pode variar de $1 \leq \alpha \leq S$. Não obstante, para facilitar a formulação matricial, omite-se este fator. A Figura 10, adaptada de [40], apresenta um sinal dividido em blocos com uma sobreposição de sinal de 50 %.

A Equação (24) apresenta uma matriz $\mathbf{Y}_q(m)$ de dimensão $N \times D$, como uma extensão de $\mathbf{y}_q(m)$, para contemplar os D atrasos de tempo presentes na matriz de correlação que por sua vez, será empregada na função de custo escolhida.

⁴Uma matriz de *Toeplitz* apresenta o valor dos elementos de cada diagonal descendente da esquerda para a direita constante, em outras palavras, cada elemento deve obedecer a igualdade $\mathbf{A}_{i,j} = \mathbf{A}_{i-1,j-1}$.

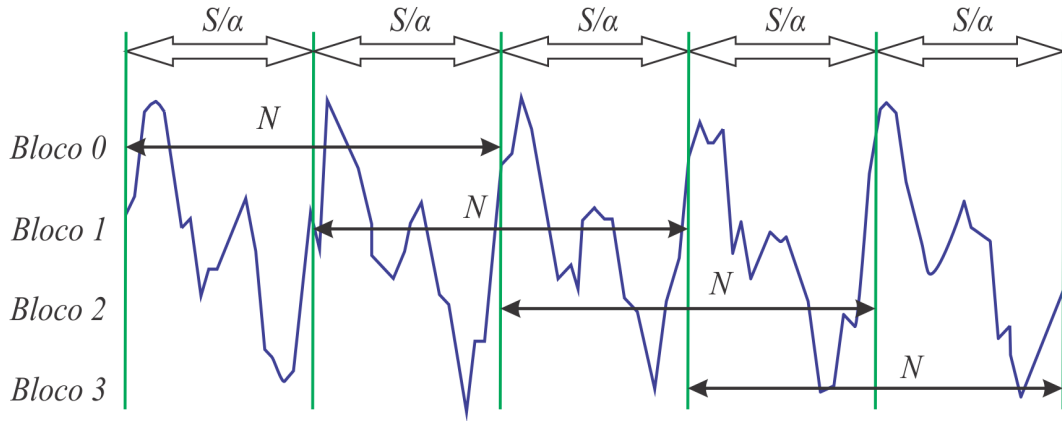


Figura 10 - Divisão do sinal em blocos sobrepostos [40].

$$\mathbf{Y}_q(m) = \begin{bmatrix} y_q(mS) & \cdots & y_q(mS - D + 1) \\ y_q(mS - 1) & \ddots & y_q(mS - D + 2) \\ \vdots & \ddots & \vdots \\ y_q(mS + N - 1) & \cdots & y_q(mS - D + N) \end{bmatrix}. \quad (24)$$

Analogamente a Equação (21) podemos expressar $\mathbf{Y}_q(m)$ como:

$$\mathbf{Y}_q(m) = \sum_{p=1}^P \mathbf{X}_p(m) \mathbf{W}_{pq}, \quad (25)$$

sendo $\mathbf{X}_p(m)$ uma matriz obtida a partir da concatenação de duas matrizes $\tilde{\mathbf{X}}_p(m)$, logo apresenta dimensão $N \times 2S$ e podemos expressá-la como:

$$\mathbf{X}_p(m) = \begin{bmatrix} \tilde{\mathbf{X}}_p^T(m) & \tilde{\mathbf{X}}_p^T(m-1) \end{bmatrix}, \quad (26)$$

e \mathbf{W}_{pq} uma matriz de dimensão $2S \times D$, conhecida como matriz *Sylvester*⁵ que podemos expressar por:

⁵Uma matriz *Sylvester* é uma matriz que exhibe um formato especial de estrutura *Toeplitz*, e permite representar a operação de convolução como o produto entre uma matriz e um vetor.

$$\mathbf{W}_{pq} = \begin{bmatrix} w_{pq}(0) & 0 & \cdots & 0 \\ w_{pq}(1) & w_{pq}(0) & \ddots & \vdots \\ \vdots & w_{pq}(1) & \ddots & 0 \\ w_{pq}(S-1) & \vdots & \ddots & w_{pq}(0) \\ 0 & w_{pq}(S-1) & \ddots & w_{pq}(1) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & w_{pq}(S-1) \\ 0 & \cdots & 0 & 0 \\ \vdots & \cdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \end{bmatrix}. \quad (27)$$

Uma vez que $1 \leq D \leq S$, as últimas $S - D + 1$ linhas da matriz \mathbf{W}_{pq} na Equação (27) precisam ser preenchidas com zeros para garantir uma dimensão compatível no produto matricial com $\mathbf{X}_p(m)$. Em [14] temos que para $D = S$ é uma escolha ótima para o algoritmo no domínio do tempo, e que a condição de $D > 1$, faz-se necessária para contemplar ambas as propriedades de não estacionariedade e não branqueamento.

Para tornar a notação mais conveniente, podemos combinar todas os canais, permitindo então expressar na Equação (28) um formato genérico e compacto:

$$\mathbf{Y}(m) = \mathbf{X}(m)\mathbf{W}, \quad (28)$$

sendo $\mathbf{Y}(m)$ uma matriz global de dimensão $N \times PD$, que para o caso determinado, ou seja, $P = Q$, pode ser representada por:

$$\mathbf{Y}(m) = \begin{bmatrix} \mathbf{Y}_1(m) & \cdots & \mathbf{Y}_P(m) \end{bmatrix}, \quad (29)$$

analogamente, a matriz global $\mathbf{X}(m)$ de dimensão $N \times 2PS$ e \mathbf{W} de dimensão $2PS \times PD$ podem ser expressas respectivamente da seguinte maneira:

$$\mathbf{X}(m) = \begin{bmatrix} \mathbf{X}_1(m) & \cdots & \mathbf{X}_P(m) \end{bmatrix}, \quad (30)$$

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \cdots & \mathbf{W}_{1P} \\ \vdots & \ddots & \vdots \\ \mathbf{W}_{P1} & \cdots & \mathbf{W}_{PP} \end{bmatrix}. \quad (31)$$

2.3 Matriz de Correlação

Após apresentarmos uma formulação compacta para modelar os filtros do sistema de separação, é preciso avaliar o critério de otimização proposto em [14] que explora simultaneamente as propriedades de não branqueamento e não estacionariedade dos sinais originais. Para isso, precisamos primeiramente definir as matrizes de correlação de tempo curto, que para o m -ésimo bloco podem ser expressas como:

$$\mathbf{R}_{\mathbf{xx}}(m) = \mathbf{X}^H(m) \cdot \mathbf{X}(m), \quad (32)$$

$$\mathbf{R}_{\mathbf{yy}}(m) = \mathbf{Y}^H(m) \cdot \mathbf{Y}(m), \quad (33)$$

sendo $(\cdot)^H$ o operador hermitiano (transposto conjugado da matriz) e cujas dimensões são respectivamente $2PL \times 2PL$ e $PD \times PD$.

Cumpramos notar que os sinais acústicos, empregados neste trabalho, possuem valores reais, desta forma, o operador hermitiano empregado nas equações possuem o mesmo efeito do operador $(\cdot)^T$ [40].

Além disso, cabe ressaltar que a matriz $\mathbf{R}_{\mathbf{yy}}(m)$ devido às considerações realizadas anteriormente, já contempla implicitamente, as correlações em diferentes atrasos (intervalos) de tempo e posto completo, uma vez que a condição $N \geq PD$ é satisfeita e que pela Equação (32) a matriz $\mathbf{R}_{\mathbf{yy}}(m)$ é obtida através do produto de uma matriz de dimensão $PD \times N$ por outra matriz de dimensão $N \times PD$ [23].

Podemos então representar a matriz $\mathbf{R}_{\mathbf{yy}}(m)$ por:

$$\mathbf{R}_{\mathbf{yy}}(m) = \begin{bmatrix} \mathbf{Y}_1^H(m) \\ \mathbf{Y}_2^H(m) \\ \vdots \\ \mathbf{Y}_Q^H(m) \end{bmatrix} \begin{bmatrix} \mathbf{Y}_1(m) & \mathbf{Y}_2(m) & \cdots & \mathbf{Y}_Q(m) \end{bmatrix}, \quad (34)$$

obtendo-se assim, a seguinte expressão na forma expandida:

$$\mathbf{R}_{\mathbf{yy}}(m) = \begin{bmatrix} \mathbf{Y}_1^H(m)\mathbf{Y}_1(m) & \mathbf{Y}_1^H(m)\mathbf{Y}_2(m) & \cdots & \mathbf{Y}_1^H(m)\mathbf{Y}_Q(m) \\ \mathbf{Y}_2^H(m)\mathbf{Y}_1(m) & \mathbf{Y}_2^H(m)\mathbf{Y}_2(m) & \cdots & \mathbf{Y}_2^H(m)\mathbf{Y}_Q(m) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Y}_Q^H(m)\mathbf{Y}_1(m) & \mathbf{Y}_Q^H(m)\mathbf{Y}_2(m) & \cdots & \mathbf{Y}_Q^H(m)\mathbf{Y}_Q(m) \end{bmatrix}, \quad (35)$$

em que cada termo $\mathbf{Y}_q^H(m)\mathbf{Y}_q(m)$ para $q = 1, \dots, Q$, representa a matriz de autocorrelação da q -ésima saída, que por sua vez representamos por $\mathbf{R}_{\mathbf{y}_q\mathbf{y}_q}(m)$. Por outro lado, representamos cada termo $\mathbf{Y}_p^H(m)\mathbf{Y}_q(m)$, para $p \neq q$, como a matriz de correlação cruzada entre a p -ésima e a q -ésima saída, também representado por $\mathbf{R}_{\mathbf{y}_p\mathbf{y}_q}(m)$.

2.4 Função de Custo

A função de custo apresentada em [14] e [44] adotada neste trabalho é:

$$\Phi(m) = \sum_{i=0}^{\infty} \beta(i, m) \{ \log \det \text{bdiag}[\mathbf{Y}^H(i) \cdot \mathbf{Y}(i)] - \log \det[\mathbf{Y}^H(i) \cdot \mathbf{Y}(i)] \}, \quad (36)$$

sendo $\beta(i, m) \geq 0$ um ponderador em que:

$$\sum_{i=0}^{\infty} \beta(i, m) = 1, \quad (37)$$

além disso, a Equação (36), dita não negativa conforme apresentada em [23] e [45], introduz o operador *block diagonal* ($\text{bdiag}(\cdot)$), que interpreta uma matriz como uma associação de submatrizes (partições), em que cada submatriz possui origem distinta, e a partir desta subdivisão em partições, faz com que todas as submatrizes que não pertencem a diagonal principal da matriz de partições sejam matrizes nulas com dimensões similares às partições correspondentes. Podemos então, exemplificar o operador mencionado para um sistema de duas fontes e duas misturas ($Q = P = 2$), também denominado um sistema TITO (do inglês *Two Input Two Output*), através da seguinte matriz de correlação:

$$\mathbf{R}_{\mathbf{y}\mathbf{y}}(m) = \begin{bmatrix} \mathbf{Y}_1^H(m)\mathbf{Y}_1(m) & \mathbf{Y}_1^H(m)\mathbf{Y}_2(m) \\ \mathbf{Y}_2^H(m)\mathbf{Y}_1(m) & \mathbf{Y}_2^H(m)\mathbf{Y}_2(m) \end{bmatrix}, \quad (38)$$

em que aplicando-se o operador $\text{bdiag}(\cdot)$, temos:

$$\mathbf{R}_{\mathbf{y}\mathbf{y}}(m) = \begin{bmatrix} \mathbf{Y}_1^H(m)\mathbf{Y}_1(m) & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_2^H(m)\mathbf{Y}_2(m) \end{bmatrix}, \quad (39)$$

sendo $\mathbf{0}$ uma matriz nula com as mesmas dimensões de $\mathbf{R}_{\mathbf{y}_p\mathbf{y}_q}(m)$. A Figura 11, adaptação de [34], representa uma ilustração para o exemplo descrito.

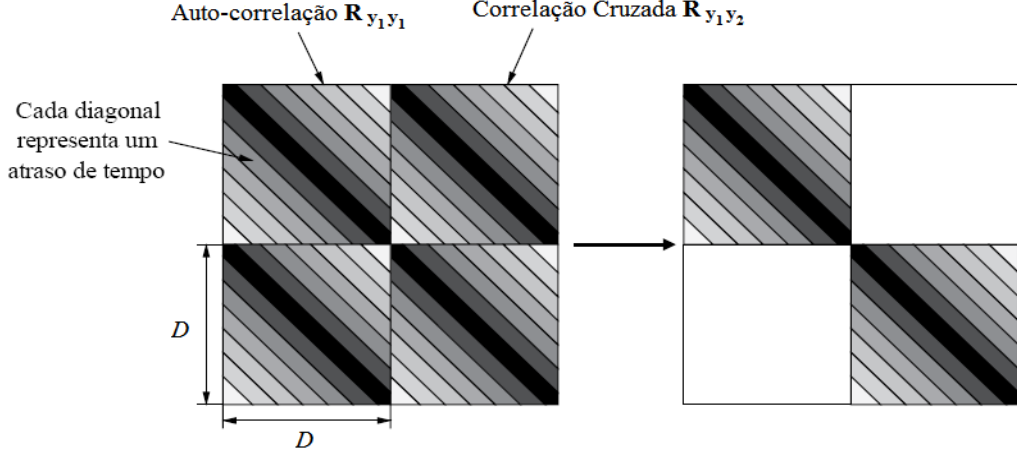


Figura 11 - Ilustração de correlação para o caso para $Q = P = 2$ [34].

2.4.1 Otimização da Função de Custo

O gradiente de primeira ordem é a ferramenta matemática empregada na minimização da função de custo apresentada na Equação (36).

A otimização da função de custo nos permite expressar as equações referentes as atualizações dos filtros de separação \mathbf{w}_{pq} (Equação (19)), presentes implicitamente na matriz \mathbf{W} (Equação (29)). Estas atualizações dos filtros ocorrem via gradiente descendente e podem ser mensuradas através do gradiente da função de custo (considerada uma função real) em relação a matriz \mathbf{W} (considerada a priori uma matriz complexa), conforme apresentado na Equação (40):

$$\nabla_{\mathbf{w}} \Phi(m) = 2 \cdot \frac{\partial \Phi(m)}{\partial \mathbf{W}^*}, \quad (40)$$

em que utilizamos a forma complexa, visto que por definição o gradiente de uma função real de uma matriz complexa, conforme as considerações realizadas é dado mediante exposto na Equação (40) [46]. Para facilitar o desenvolvimento matemático, omitiremos o índice “ m ” referente ao m -ésimo bloco. E como resultado da aplicação deste gradiente à função de custo da Equação (36), após algumas manipulações matemáticas apresentada em [23] e em destaque no Apêndice A, temos:

$$\begin{aligned} \nabla_{\mathbf{w}} \Phi(m) &= 2 \frac{\partial \left\{ \sum_{i=0}^{\infty} [\beta(i, m) \log \det \text{bdiag}(\mathbf{Y}^H \cdot \mathbf{Y}) - \log \det(\mathbf{Y}^H \cdot \mathbf{Y})] \right\}}{\partial \mathbf{W}^*} \\ &= 2 \sum_{i=0}^{\infty} \beta(i, m) \mathbf{R}_{\mathbf{x}\mathbf{y}} \mathbf{R}_{\mathbf{y}\mathbf{y}}^{-1} (\mathbf{R}_{\mathbf{y}\mathbf{y}} - \text{bdiag} \mathbf{R}_{\mathbf{y}\mathbf{y}}) \text{bdiag}^{-1} \mathbf{R}_{\mathbf{y}\mathbf{y}}. \end{aligned} \quad (41)$$

Analisando a Equação (41) verificamos que a mesma torna-se zero se e somente se as partições não pertencentes a diagonal principal de $\mathbf{Y}^H(m)\mathbf{Y}(m)$, que correspondem as correlações cruzadas entre os sinais de saída do sistema de separação, forem iguais a zero para todos os atrasos de tempo considerados, explorando assim, ambas as propriedades: não-estacionariedade e não brancura (não branqueamento).

2.4.2 Gradiente Natural

Devido ao fato do método do gradiente descendente apresentar uma convergência lenta em diversas aplicações práticas, uma vez que depende dos dados a serem processados [46], nesta dissertação, emprega-se o gradiente natural, que é uma versão mais robusta, como fator de atualização dos filtros de separação [14], dado a seguir:

$$\nabla_{\mathbf{W}}^{NG} \Phi(m) = \mathbf{W}\mathbf{W}^H \nabla_{\mathbf{W}} \Phi(m). \quad (42)$$

Aplicando-se o gradiente natural na Equação (41) temos:

$$\begin{aligned} \nabla_{\mathbf{W}}^{NG} \Phi(m) &= 2 \sum_{i=0}^{\infty} \beta(i, m) \left[\mathbf{W}\mathbf{W}^H \mathbf{R}_{\mathbf{x}\mathbf{y}} \mathbf{R}_{\mathbf{y}\mathbf{y}}^{-1} (\mathbf{R}_{\mathbf{y}\mathbf{y}} - \text{bdiag} \mathbf{R}_{\mathbf{y}\mathbf{y}}) \text{bdiag}^{-1} \mathbf{R}_{\mathbf{y}\mathbf{y}} \right] \\ &= 2 \sum_{i=0}^{\infty} \beta(i, m) \left[\mathbf{W}\mathbf{W}^H \mathbf{X}^H \mathbf{Y} \mathbf{R}_{\mathbf{y}\mathbf{y}}^{-1} (\mathbf{R}_{\mathbf{y}\mathbf{y}} - \text{bdiag} \mathbf{R}_{\mathbf{y}\mathbf{y}}) \text{bdiag}^{-1} \mathbf{R}_{\mathbf{y}\mathbf{y}} \right] \\ &= 2 \sum_{i=0}^{\infty} \beta(i, m) \left[\mathbf{W} (\mathbf{X}\mathbf{W})^H \mathbf{Y} \mathbf{R}_{\mathbf{y}\mathbf{y}}^{-1} (\mathbf{R}_{\mathbf{y}\mathbf{y}} - \text{bdiag} \mathbf{R}_{\mathbf{y}\mathbf{y}}) \text{bdiag}^{-1} \mathbf{R}_{\mathbf{y}\mathbf{y}} \right] \\ &= 2 \sum_{i=0}^{\infty} \beta(i, m) \left[\mathbf{W} \mathbf{Y}^H \mathbf{Y} \mathbf{R}_{\mathbf{y}\mathbf{y}}^{-1} (\mathbf{R}_{\mathbf{y}\mathbf{y}} - \text{bdiag} \mathbf{R}_{\mathbf{y}\mathbf{y}}) \text{bdiag}^{-1} \mathbf{R}_{\mathbf{y}\mathbf{y}} \right] \\ &= 2 \sum_{i=0}^{\infty} \beta(i, m) \left[\mathbf{W} \mathbf{R}_{\mathbf{y}\mathbf{y}} \mathbf{R}_{\mathbf{y}\mathbf{y}}^{-1} (\mathbf{R}_{\mathbf{y}\mathbf{y}} - \text{bdiag} \mathbf{R}_{\mathbf{y}\mathbf{y}}) \text{bdiag}^{-1} \mathbf{R}_{\mathbf{y}\mathbf{y}} \right] \\ &= 2 \sum_{i=0}^{\infty} \beta(i, m) \left[\mathbf{W} (\mathbf{R}_{\mathbf{y}\mathbf{y}} - \text{bdiag} \mathbf{R}_{\mathbf{y}\mathbf{y}}) \text{bdiag}^{-1} \mathbf{R}_{\mathbf{y}\mathbf{y}} \right]. \end{aligned} \quad (43)$$

A modificação realizada propiciou uma significativa redução computacional em relação a Equação (41), visto que o cálculo da inversa da matriz $\mathbf{R}_{\mathbf{y}\mathbf{y}}(m)$ de dimensão $PD \times PD$, não é mais necessária. Por outro lado, o cálculo do termo $\text{bdiag}^{-1}(\mathbf{R}_{\mathbf{y}\mathbf{y}}(m))$ também presente na Equação (43), necessita ser calculado, não obstante cabe ressaltar que devido ao operador $\text{bdiag}(\cdot)$, a inversa do termo em questão resulta apenas da inversão de P submatrizes (partições) de $\mathbf{R}_{\mathbf{y}\mathbf{y}}(m)$, de dimensão $D \times D$. Em outras palavras, o cálculo do termo

$\text{bdiag}^{-1}(\mathbf{R}_{\mathbf{y}\mathbf{y}}(m))$ para um sistema TITO é apresentado a seguir:

$$\text{bdiag}^{-1}(\mathbf{R}_{\mathbf{y}\mathbf{y}}(m)) = \begin{bmatrix} \mathbf{R}_{\mathbf{y}_1\mathbf{y}_1}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{\mathbf{y}_2\mathbf{y}_2}^{-1} \end{bmatrix}. \quad (44)$$

Finalmente podemos expressar a atualização da matriz de separação \mathbf{W} , através de um procedimento iterativo e recursivo, conforme apresentado a seguir:

$$\mathbf{W}_j = \mathbf{W}_{j-1} - \mu (\nabla_{\mathbf{W}}^{NG} \Phi(m)), \quad (45)$$

em que μ é o fator (passo) de aprendizagem e j representa a j -ésima iteração.

Podemos então, representar a atualização da matriz de separação \mathbf{W} de um sistema TITO como:

$$\mathbf{W}_j = \mathbf{W}_{j-1} - 2\mu \sum_{i=0}^{\infty} \beta(i, m) \begin{bmatrix} \mathbf{W}_{12} \mathbf{R}_{\mathbf{y}_2\mathbf{y}_1} \mathbf{R}_{\mathbf{y}_1\mathbf{y}_1}^{-1} & \mathbf{W}_{11} \mathbf{R}_{\mathbf{y}_1\mathbf{y}_2} \mathbf{R}_{\mathbf{y}_2\mathbf{y}_2}^{-1} \\ \mathbf{W}_{22} \mathbf{R}_{\mathbf{y}_2\mathbf{y}_1} \mathbf{R}_{\mathbf{y}_1\mathbf{y}_1}^{-1} & \mathbf{W}_{21} \mathbf{R}_{\mathbf{y}_1\mathbf{y}_2} \mathbf{R}_{\mathbf{y}_2\mathbf{y}_2}^{-1} \end{bmatrix}. \quad (46)$$

Com o propósito de reduzir a complexidade computacional do algoritmo, em [47], mostra-se que podemos simplificar o cálculo do fator $\mathbf{R}_{qq}^{-1}(m)$ utilizando-se a seguinte normalização:

$$\mathbf{R}_{qq}(m) \approx \mathbf{y}_q^T(m) \mathbf{y}_q(m) \mathbf{I} \quad (47)$$

em que a inversa da matriz $\mathbf{R}_{qq}(m)$ é reduzida ao inverso da potência de um único bloco de sinal de saída, visto que resulta do produto de uma matrizes de dimensão $1 \times N$ por outra de dimensão $N \times 1$ (Equação (20)), e \mathbf{I} é uma matriz identidade ($D \times D$) [15].

Além disso, podemos utilizar apenas os S primeiros elementos da primeira coluna de \mathbf{W}_{pq} na Equação (27), para reduzir o custo computacional [14], conforme apresentado na Figura 12, adaptação de [34]:

2.4.3 Implementações do algoritmo: *offline*, *online* e *block online*

O parâmetro $\beta(i, m)$ na Equação (43) proposto em [47] permite adaptar o algoritmo em três implementações: *online*, *block online* e *offline*. Cabe ressaltar que a escolha do parâmetro $\beta(i, m)$ influencia apenas a velocidade de convergência do algoritmo.

Para a implementação do algoritmo *offline*, ou também chamado algoritmo tipo batelada (*batch*), a atualização dos filtros de separação ocorre utilizando-se todos os blocos

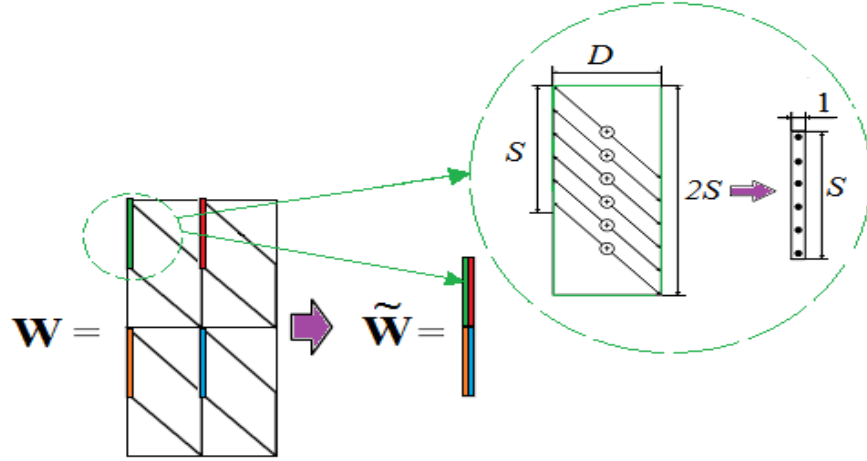


Figura 12 - Ilustração da redução computacional realizada, empregando-se os S primeiros elementos da primeira coluna de \mathbf{W}_{pq} para um sistema TITO [34].

do sinal. O parâmetro $\beta(i, m)$ nesta implementação corresponde a uma janela retangular, conforme apresentado na Figura 13 [47], que podemos descrever da seguinte maneira:

$$\beta(i, m) = \frac{1}{b} \epsilon_{0, (b-1)}(i) \quad (48)$$

sendo que b é o número de blocos empregados no algoritmo e $\epsilon_{0, (b-1)}(i) = 1$ para o intervalo $0 \leq i \leq (b-1)$ e caso contrário, $\epsilon_{0, (b-1)}(i) = 0$.

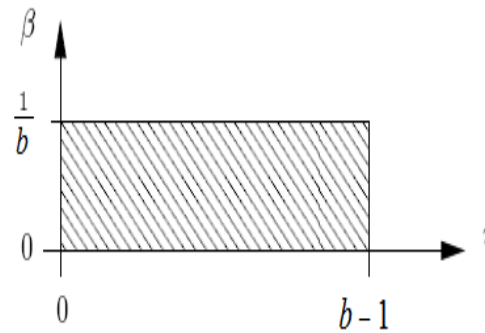


Figura 13 - O parâmetro $\beta(i, m)$ numa implementação *offline* [47].

Na implementação do algoritmo *online* a atualização dos filtros de separação ocorre à medida que temos acesso a um novo bloco de dados, a função de ponderação, $\beta(i, m)$ corresponde a uma função exponencial com fator de esquecimento λ , apresentada na Figura 14 [47], que podemos expressar da seguinte maneira:

$$\beta(i, m) = (1 - \lambda) \lambda^{m-i} \epsilon_{0, m}(i) \quad (49)$$

em que $0 \leq \lambda \leq 1$ e a função $\epsilon_{0,m}(i) = 1$ para o intervalo $0 \leq i \leq m$ e caso contrário, $\epsilon_{0,m}(i) = 0$.

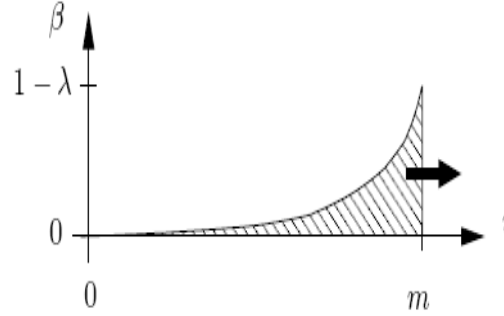


Figura 14 - O parâmetro $\beta(i, m)$ numa implementação *online* [47].

Uma implementação do algoritmo *block online* é considerada híbrida uma vez que corresponde a uma combinação entre as implementações *offline* e *online*, mencionadas anteriormente, cuja função $\beta(i, m)$ empregada é apresentada na Figura 15 [47]. No algoritmo *block online* a atualização dos filtros de separação, ocorre executando a implementação *offline*, a cada K blocos de comprimento N , por um determinado número de iterações.

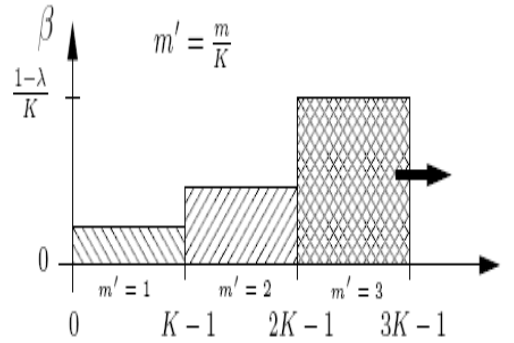


Figura 15 - O parâmetro $\beta(i, m)$ numa implementação *block online* [47].

Pelo fato de empregarmos neste trabalho apenas a implementação *offline*, podemos reescrever a Equação (45) substituindo o termo $\beta(i, m)$ e expressando a atualização da matriz \mathbf{W} da seguinte maneira:

$$\mathbf{W}_j = \mathbf{W}_{j-1} - \frac{2\mu}{b} \sum_{i=0}^{b-1} [\mathbf{W} (\mathbf{R}_{yy} - b \text{diag} \mathbf{R}_{yy}) b \text{diag}^{-1} \mathbf{R}_{yy}]. \quad (50)$$

A implementação *offline* permite o algoritmo acessar os dados do sinal repetitivamente a cada iteração j , para atualizar os filtros de separação, desta forma este tipo de

implementação possui desempenho superior quando comparado às implementações *online* e *block online*.

2.4.4 Fator de aprendizagem adaptativo

O algoritmo apresentado na Equação (50) adota um fator de aprendizagem fixo (μ), tal consideração torna o algoritmo dependente de uma escolha apropriada deste fator, o que em geral, pode resultar em uma taxa de convergência lenta, caso a escolha seja um valor pequeno ou em instabilidade do algoritmo, caso a escolha seja um valor muito grande.

Nesta dissertação empregamos um fator de aprendizagem adaptativo, $\mu(j)$ para acelerar a convergência do algoritmo. Este fator de aprendizagem adaptativo pode ser inserido na Equação (50) de atualização da matriz \mathbf{W} da seguinte maneira:

$$\mathbf{W}_j = \mathbf{W}_{j-1} - \frac{2\mu(j)}{b} \sum_{i=0}^{b-1} [\mathbf{W} (\mathbf{R}_{yy} - b \text{diag} \mathbf{R}_{yy}) b \text{diag}^{-1} \mathbf{R}_{yy}] \quad (51)$$

Os valores do passo de adaptação inserido na Equação (51) podem variar (adaptar-se) dependendo da convergência dos coeficientes do sistema de separação a cada iteração. Caso mais do que 70% dos coeficientes dos filtros de separação apresentem em duas iterações consecutivas alterações na mesma direção (indicando a convergência do algoritmo), na iteração seguinte o valor de μ é escalado por um fator $\alpha = 1,025$ ($\mu_j = \alpha \cdot \mu_{(j-1)}$), caso contrário (ou seja, em caso de divergência do algoritmo), se o passo for superior ao μ inicial, o fator de aprendizagem, na iteração seguinte, retorna ao seu valor inicial ($\mu_j = \mu_{inicial}$); e por fim caso as duas condições não sejam verdadeiras, o valor de μ da iteração seguinte é reduzido (dividido) pela mesma constante α ($\mu_j = \mu_{(j-1)}/\alpha$) [15].

2.4.5 Inicialização dos filtros de separação

Neste trabalho consideramos o cenário acústico para duas fontes e dois sensores, conforme apresentado na Figura 16 [40], em que cada fonte encontra-se na metade do plano formado pelos sensores, por conseguinte para se obter a separação dos sinais, apenas filtros FIR causais são necessários [40].

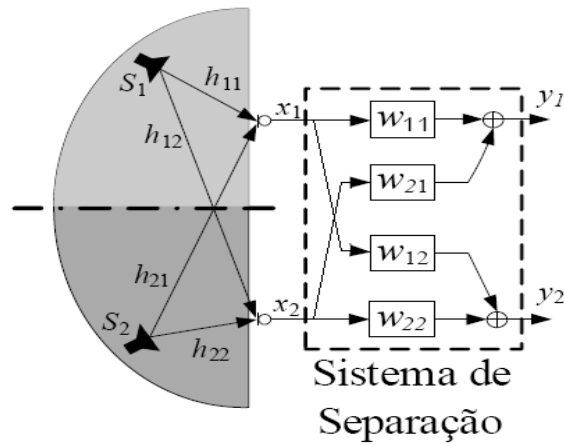


Figura 16 - Ilustração de um cenário acústico para um sistema TITO [40].

Para este cenário acústico, conforme proposto em [14], o primeiro coeficiente dos filtros de separação \mathbf{w}_{pp} , para $p = 1, \dots, P$, na Equação (19), podem ser inicializados com um impulso unitário, isto é, $w_{pp}(0) = 1$, enquanto os demais coeficientes do filtro são inicializados em zero. Todos os coeficientes de \mathbf{w}_{pq} , para $p \neq q$ também são inicializados em zero.

3 DETECTORES DE ATIVIDADE DE VOZ

3.1 Introdução

Este capítulo apresenta uma breve abordagem sobre detectores de voz, destacando os principais detectores de voz encontrados na literatura, enfatizando os motivos da adoção do detector de voz empregado nesta dissertação.

Os detectores de atividade de voz (VADs, do inglês *Voice Activity Detectors*) almejam a determinação dos trechos de atividade de fala existentes em um sinal de voz. De modo geral, são algoritmos encontrados na literatura, capazes de identificar a existência de trechos que contenham voz em sinais possivelmente ruidosos compostos, alternadamente, por trechos de fala e de silêncio [48]. Detectores de voz são amplamente explorados na literatura em aplicações como: cancelamento de eco [49], filtragem adaptativa [49] e supressão de ruído [50].

A forma na qual um detector de voz consegue avaliar e decidir, a qual classe pertence um trecho de sinal de voz, é baseada na comparação entre um limiar, que em geral é adaptativo, e um vetor de observação (quantidade em análise). Embora existam diferentes vetores de observações, empregados em diversas técnicas encontradas na literatura (tais como energia do sinal, entropia espectral, *zero-crossing rate*, coeficientes da transformada de *Wavelet*, entre outros) a identificação precisa, e conseqüentemente a classificação correta, de trechos com presença ou ausência de fala, é geralmente difícil, principalmente quando o sinal de voz é corrompido por ruído de fundo ou interferência indesejada, no entanto, algumas técnicas, como por exemplo, as que utilizam como vetor de observação, a energia do sinal, apresentam bons resultados [18] e [51].

Visando discernir se o trecho de sinal de voz analisado contém fala ou não, a resposta do detector de voz pode ser representada por um bit, caso o vetor de observação esteja abaixo de um limiar, atribui-se “0”, para representar que o intervalo do sinal de voz em análise contém apenas o ruído do ambiente, caso contrário, atribui-se “1”, para representar todo intervalo do sinal de voz no qual espera-se encontrar não somente o ruído do ambiente, mas também a fala, em outras palavras:

$$Decisão \begin{cases} 0, & \text{se } \mathbf{x}=\mathbf{n} \\ 1, & \text{se } \mathbf{x}=\mathbf{n}+\mathbf{v}, \end{cases} \quad (52)$$

em que \mathbf{x} representa o trecho do sinal da mistura, \mathbf{n} representa a presença de ruído ambiente,

pertinente aos trechos de silêncio e \mathbf{v} representa a presença de voz e ruído ambiente, existentes em trechos onde está contida a fala. Assim, espera-se encontrar na saída do detector de voz, pulsos retangulares enquadrando os trechos com atividade vocal.

Alguns tipos de detectores existentes na literatura, foram avaliados e implementados, tais como: o Detector Linear e Adaptativo de Energia (ALED, do inglês *Adaptive Linear Energy-Based Detector*) [49], o Detector Linear de Energia em Sub-bandas (LSED, do inglês *Linear Sub-Band Energy Detector*) [49] e o Modelo Estatístico de Detecção de Voz (SMBVAD, do inglês *Statistical Model-Based Voice Activity Detector*) [52]. Todavia conjugando eficiência e baixo custo computacional, utiliza-se neste trabalho o ALED.

A seguir serão apresentados três algoritmos de detectores de atividade de voz encontrados na literatura e analisados neste trabalho.

3.2 Detector Linear e Adaptativo de Energia (ALED)

O ALED, empregado nesta dissertação, assim como muitos outros algoritmos detectores de voz, é baseado em um vetor de observação de energia, por conseguinte, é capaz de discernir a presença ou ausência de voz em um determinado intervalo, embasado nas variações de energia do sinal de voz em análise. De um modo geral, este tipo de detector é bem atrativo, visto que possui um desempenho satisfatório e baixo custo computacional quando comparado aos demais detectores. Além disso, o ALED costuma ser mais eficiente em avaliar e classificar a energia de um trecho de sinal de voz, quando comparado a um detector de limiar fixo, uma vez que o limiar do ALED pode adaptar-se à energia do ruído presente no ambiente, que não é considerada constante, enquanto o mesmo não ocorre em um detector de limiar fixo, evitando assim que trechos de silêncio, pelo fato da energia do ruído ser variável, sejam considerados trechos de fala, por apresentarem energia superior a algum limiar fixo pré-estabelecido, o que é indesejável [49].

Por outro lado, para fins de separação cega de sinais, alvo desta dissertação, é preferível que um trecho de silêncio seja considerado um trecho de sinal de fala do que o contrário, ressaltando que nas proximidades de um intervalo contendo a voz, optou-se em incluir alguns possíveis trechos de silêncio, não somente devido à reverberação do sinal no ambiente, mas também para garantir a eficiência do detector, visto que o desempenho do ALED conforme mencionado, pode ser comprometido a partir de variações de energia do ruído presente no sinal desejado ou observado (misturas).

Para a implementação do ALED, seja $x(n)$ o sinal em que se deseja efetuar a detecção

de fala, e que corresponde ao sinal proveniente do sistema de mistura, é preciso dividir este sinal em quadros (*frames*) de tamanhos pré-estabelecidos. Assim, temos que:

$$q_j = x(i)|_{i=(j-1)K+1}^{jK}, \quad (53)$$

em que $x(i)$ corresponde i -ésima amostra do sinal $x(n)$, e q_j representa o j -ésimo quadro q de tamanho K .

Podemos agora expressar a energia do j -ésimo quadro por:

$$E_j = \frac{1}{K} \sum_{i=(j-1)K+1}^{jK} x^2(i), \quad (54)$$

sendo E_j a energia de q_j .

A energia de cada quadro mensurada conforme a Equação (54) pode então ser comparada com o limiar, que ao contrário do detector de voz de limiar fixo, não é constante. Logo é preciso estabelecer a forma como ocorre esta atualização. Em [53] adota-se o seguinte critério para a atualização do limiar de detecção:

$$L_{novo} = (1 - \varphi)L_{anterior} + \varphi E_{silêncio}, \quad (55)$$

em que L_{novo} representa o novo limiar a ser mensurado, $L_{anterior}$ o último limiar calculado utilizado, $E_{silêncio}$ a última energia calculada para o último quadro em que não houve detecção de voz e φ é um parâmetro que pode variar de $0 < \varphi < 1$ [53]. A escolha ideal de φ é imprescindível para obter-se um melhor desempenho do detector.

Com o intuito de elevar a robustez do detector em relação às estatísticas de ruído de fundo, utilizou-se para a atualização de φ , a informação estatística de segunda ordem, do conjunto das energias calculadas para os m quadros mais recentes onde não houve detecção de voz, conforme proposto em [15] e [53]. Para isso, fez-se necessário criar uma memória denominada M_n , de m componentes, preenchida com as energias dos m últimos quadros de trechos de silêncio, cuja atualização da memória supracitada ocorre no instante em que um novo quadro de ausência de fala é detectado, o registro armazenado da componente de energia mais antigo é descartado, e substituído pela componente de energia atual. Face ao exposto, podemos expressar:

$$\sigma^2 = var(M_n), \quad (56)$$

em que $var(\cdot)$ é a variância dos m componentes da memória M_n [49].

Desta forma, a cada quadro desprovido de fala detectado, uma nova memória é formada, via substituição da componente mais antiga pela recente, e por conseguinte, uma nova variância da memória, denominada σ_{novo}^2 , precisa ser calculada. Em vista disso, a escolha de φ pode ser orientada a partir de ξ , que é um parâmetro definido por:

$$\xi = \frac{\sigma_{novo}^2}{\sigma_{antigo}^2}, \quad (57)$$

em que σ_{antigo}^2 , representa a variância da memória no instante antes da atualização. A relação entre φ e ξ é apresentada na Tabela 1 conforme descrito em [53].

Tabela 1 - Critério de atualização de p .

$\xi = \frac{\sigma_{novo}^2}{\sigma_{antigo}^2}$	φ
$\xi \geq 1,25$	0,25
$1,10 \leq \xi < 1,25$	0,20
$1,00 \leq \xi < 1,10$	0,15
$\xi < 1,00$	0,10

Cabe ressaltar que embora seja possível reduzir ou elevar o limiar de detecção, a partir do monitoramento das variações de energia do ruído do ambiente, a escolha dos m componentes de M_n , para um bom desempenho do ALED, é essencial.

Por simplicidade, pode-se considerar [53] que nos m primeiros quadros temos ausência de fala, preenchendo assim a memória inicial necessária para o algoritmo, e também para estimar o limiar inicial, da seguinte maneira:

$$L_{inicial} = \frac{k}{m} \sum_{i=1}^m E_i, \quad (58)$$

em que k é conhecido como margem de erro.

Por outro lado, como não podemos afirmar que os m primeiros quadros disponíveis, nem mesmo se os demais quadros do sinal de voz apresentarão ausência de fala, um limiar inicial foi arbitrado embasado em um quadro de silêncio de uma das mistura em análise. Tal procedimento é crucial no desempenho da separação cega, pois não compromete o desempenho do ALED, uma vez que caso o detector identifique trechos de silêncio, este limiar inicial é imediatamente atualizado, caso contrário, ou seja, caso não haja intervalo de silêncio no sinal em análise, o limiar não necessita ser alterado.

Outra alteração de suma importância realizada no ALED é a incorporação de alguns quadros desprovidos de fala a quadros com fala detectada, o motivo deve-se ao ganho de desempenho do algoritmo de separação cega de sinais, apresentado no capítulo 2, pois simul-

taneamente conseguimos com esta ação, agregar alguns trechos referente a reverberação do ambiente e maior estabilidade na resposta do detector, resultando em aumento de desempenho.

Além disso, uma sutil diferença na Tabela 1 foi necessária em relação ao apresentado em [53], para elevar o desempenho do detector ALED, tal alteração é apresentada na Tabela 2.

Tabela 2 - Critério de atualização de p .

$\xi = \frac{\sigma_{novo}^2}{\sigma_{antigo}^2}$	φ
$\xi \geq 1,25$	0,30
$1,10 \leq \xi < 1,25$	0,25
$1,00 \leq \xi < 1,10$	0,15
$\xi < 1,00$	0,10

O ALED, empregado nesta dissertação, é uma etapa de pré-processamento que visa detectar os trechos de fala no sinal de voz, de modo a desativar o algoritmo de separação cega em trechos nos quais ocorre ausência de voz.

As figuras 17 e 18 apresentam a resposta do detector de voz supracitado, em vermelho, quando aplicado a uma mistura, que é desprovida de trechos de silêncio, e outra que contempla intervalos de silêncio.

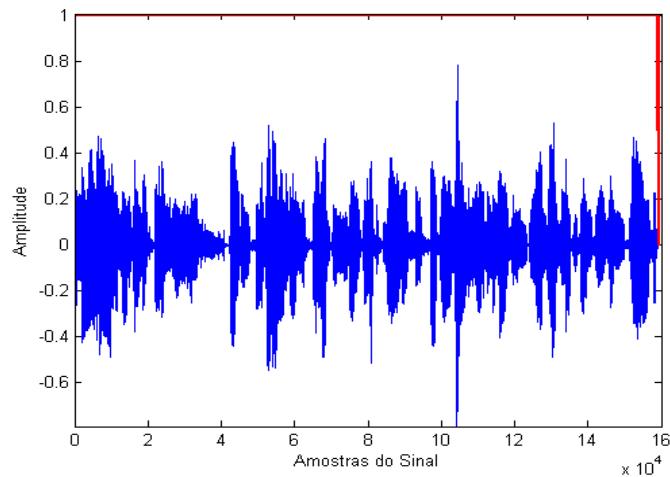


Figura 17 - Desempenho do ALED empregado em mistura com ausência de trechos de silêncio.

Na implementação do ALED, empregado nesta dissertação, utilizamos os parâmetros apresentados na Tabela 3.

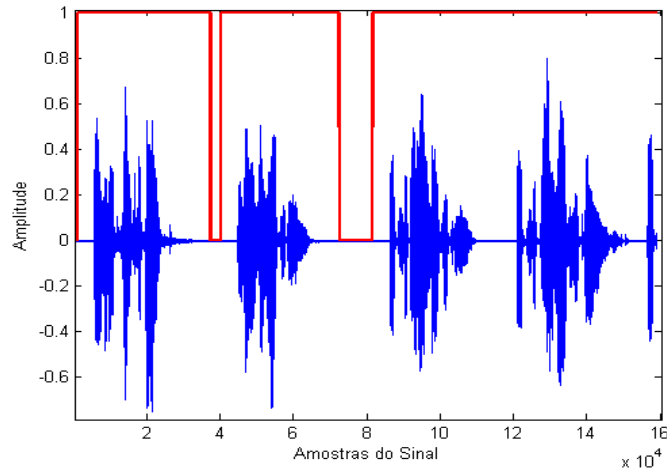


Figura 18 - Desempenho do ALED empregado em mistura com presença de trechos de silêncio.

Tabela 3 - Outros parâmetros empregados no ALED.

Tamanho do Quadro	32
Tamanho da Memória (m)	64
Margem de Erro (k)	4

3.3 Detector Linear de Energia em Sub-bandas (LSED)

O LSED é um detector de voz, cujo algoritmo é de um modo geral, bem similar ao algoritmo do ALED, com algumas peculiaridades. Neste âmbito, podemos destacar algumas características como divisão do sinal em quadros, limiar de detecção adaptativo, também é embasado em um vetor de observação de energia, por outro lado, ao contrário do ALED, a detecção da energia dos quadros ocorre em sub-bandas.

Para a implementação do algoritmo LSED assim como o ALED, faz-se necessário dividir o sinal de voz em quadros similar a Equação (53). Em seguida, visando adaptar o algoritmo para o domínio da frequência, é necessário calcular a Transformada de Cosseno Discreta (DCT, do inglês *Discrete Cosine Transform*) de cada quadro, em outras palavras:

$$F(q_j) = DCT(q_j), \quad (59)$$

em que $F(q_j)$ apresenta o resultado da DCT do j -ésimo quadro.

Face ao exposto, adotou-se a largura de $1kHz$ para cada sub-banda, conforme apresentado em [53], todavia tendo em vista que os sinais de voz empregados neste trabalho, contemplam frequência de amostragem de $16kHz$, ao contrário de [53], cujos sinais em análise apresentam frequência de amostragem de $8kHz$, foi necessário adaptar o modelo proposto

em [53], apresentado na Figura 19, com o intuito de tornar o detector de voz eficaz.

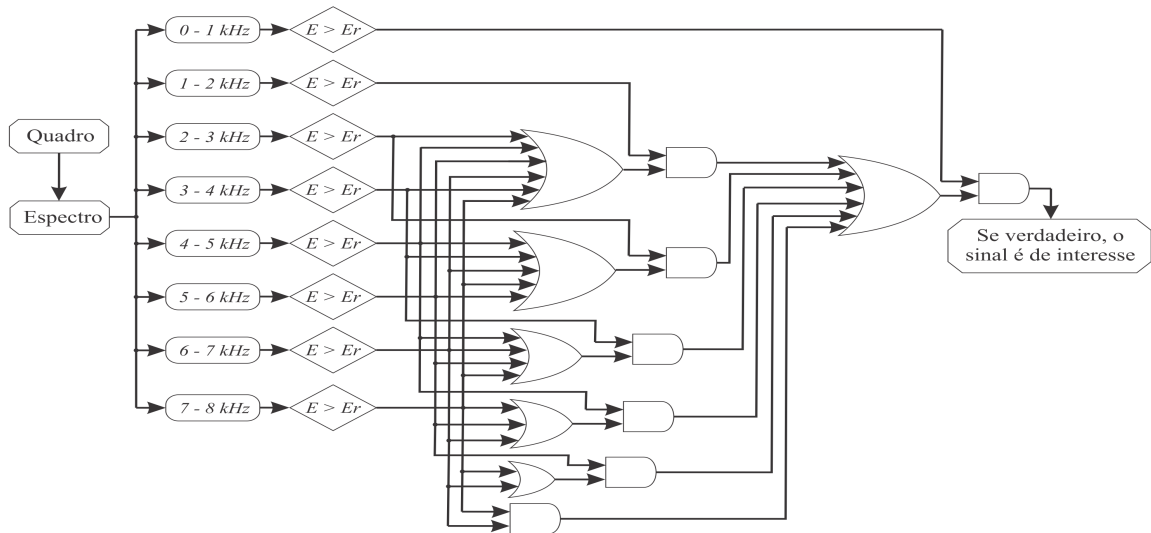


Figura 19 - Modelo lógico do LSED.

A energia de cada sub-banda (ou faixa de frequência) em cada quadro pode ser expressa como:

$$E_{b,j} = \sum_n |F_b(q_j)|^2, \quad (60)$$

sendo $E_{b,j}$ a energia das n raias da b -ésima sub-banda do j -ésimo quadro e $F_b(q_j)$ é a b -ésima partição de $F(q_j)$ contendo n raias da aludida sub-banda.

A classificação do trecho de sinal de voz analisado para o LSED, que segue a lógica apresentada na Figura 19, consiste em realizar uma análise similar à que ocorria no ALED, entretanto, no LSED o critério de decisão, bem como a atualização do limiar, ocorrem em cada uma das sub-bandas. Por conseguinte, para esta dissertação cujos sinais de voz apresentam uma frequência de amostragem de $16kHz$, faz-se necessário não somente calcular a energia a cada quadro conforme exposto para o ALED, mas também, a DCT (que representa um custo computacional adicional) e exige do algoritmo do detector o monitoramento em cada sub-banda para efetuar possíveis atualizações de limiares, o elevando o custo computacional do detector.

Cabe ressaltar que pelo fato de grande parte da energia da voz concentrar-se nas bandas inferiores, principalmente na SB_0 , vide Tabela 3, ao dividir o sinal em sub-bandas, tornamos o sistema de detecção de fala do LSED mais robusto em comparação ao ALED, visto que a energia referente ao ruído do ambiente presente no sinal é reduzida, pois a mesma

é distribuída em todas as bandas [49]. No entanto, conforme mencionado, embora haja uma confiabilidade maior em relação ao ALED, o LSED não foi utilizado, em virtude da complexidade computacional e pelo fato do ALED, para os sinais analisados, apresentar um desempenho satisfatório, para a separação cega de sinais, foco principal deste trabalho.

Tabela 4 - Relação entre a sub-banda e a respectiva faixa de frequência.

Sub-banda	Faixa de Frequência
SB_0	$0 - 1kHz$
SB_1	$1 - 2kHz$
SB_2	$2 - 3kHz$
SB_3	$3 - 4kHz$
SB_4	$4 - 5kHz$
SB_5	$5 - 6kHz$
SB_6	$6 - 7kHz$
SB_7	$7 - 8kHz$

Nesta nova abordagem, utilizamos os dados apresentados na Tabela 5.

Tabela 5 - Parâmetros empregados no LSED.

Tamanho do Quadro	400
Tamanho da Memória (m)	32
Margem de Erro (k)	2

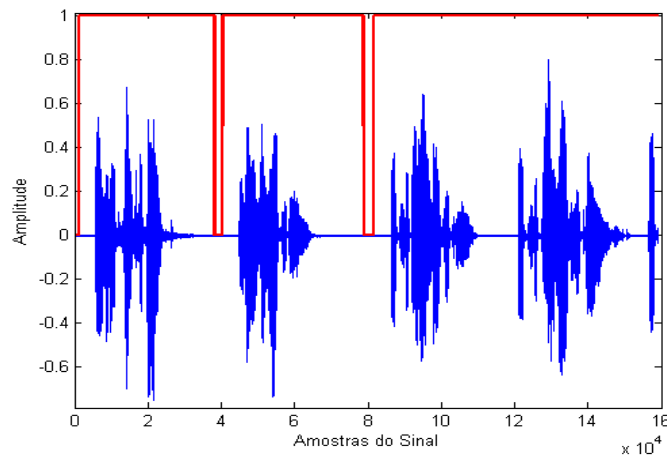


Figura 20 - Desempenho do LSED empregado em mistura com presença de trechos de silêncio.

A Figura 20 apresenta o resultado do LSED para a mesma mistura analisada na Figura 18 pelo ALED, com trechos de silêncio, em que observa-se uma ligeira diferença de detecção em relação à resposta do ALED, no entanto, embora tal diferença exista, o desempenho do algoritmo de separação cega de sinais, posterior ao detector, não é comprometido, reforçando

assim, a escolha do ALED, uma vez que para o sinal analisado, apesar de poucos trechos detectados de silêncio, conseguiu-se retirar um maior número de quadros em relação ao LSED, sem afetar o desempenho do algoritmo de separação cega de sinais, enquanto que para misturas desprovidas de intervalos de silêncio, este detector possui desempenho análogo ao ALED, ou seja, resultado idêntico à Figura 17.

3.4 Modelo Estatístico de Detecção de Voz (SMBVAD)

Visando contornar algumas dificuldades encontradas em alguns tipos de detectores, em [54] foi proposto um modelo de detector de voz, posteriormente adaptado em [52], que adota técnicas estatísticas robustas, em relação aos demais detectores apresentados, para identificar com acurácia [54] os intervalos em que a fala ocorre.

Este detector de voz, baseia-se no pressuposto de menor variabilidade das estatísticas de ruído de fundo, em relação às estatísticas do sinal desejado no qual a fala existe, e que é possível mensurar as estatísticas de ruído de fundo através de um estimador estatístico. Destarte, podemos empregar uma métrica estatística para determinar um limiar de decisão, capaz de discernir os trechos em que a fala existe dos trechos nos quais apenas o ruído de fundo ocorre, utilizando para a estimativa estatística um critério de máxima verossimilhança [48] e [54].

O detector de voz SMBVAD, assume também que a cada quadro tanto o sinal de voz em análise quanto o ruído podem ser classificados via modelos estatísticos gaussianos e independentes entre si.

Face ao exposto, as considerações apresentadas para o aludido detector, a partir da adaptação da Equação (52), nos permite expressar:

$$\begin{cases} H_0 : \text{ausência de voz} \rightarrow \bar{\mathbf{X}} = \bar{\mathbf{N}} \\ H_1 : \text{presença de voz} \rightarrow \bar{\mathbf{X}} = \bar{\mathbf{N}} + \bar{\mathbf{V}} \end{cases}, \quad (61)$$

em que $\bar{\mathbf{X}}$, $\bar{\mathbf{N}}$ e $\bar{\mathbf{V}}$ são as Transformadas Discretas de Fourier (DFT, do inglês *Discrete Fourier Transform*) de dimensão L_q , referente aos sinais de mistura, de ruído e de fala, cujos k -ésimos elementos são representados por $\bar{\mathbf{X}}_k$, $\bar{\mathbf{N}}_k$ e $\bar{\mathbf{V}}_k$ [54].

Além disso, tal consideração permite expressar as PDFs condicionadas por H_0 e H_1 por:

$$\begin{aligned}
p(\bar{\mathbf{X}}_k|H_{0,k}) &= \frac{1}{\pi\lambda_{N,k}} e^{-\frac{|\bar{\mathbf{X}}_k|^2}{\lambda_{N,k}}}, \\
p(\bar{\mathbf{X}}_k|H_{1,k}) &= \frac{1}{\pi(\lambda_{N,k} + \lambda_{V,k})} e^{-\frac{|\bar{\mathbf{X}}_k|^2}{\lambda_{N,k} + \lambda_{V,k}}},
\end{aligned} \tag{62}$$

sendo λ_N e λ_V , representam respectivamente as variâncias de N_k e V_k [54].

O estimador estatístico de máxima verossimilhança supracitado, responsável por identificar a presença ou ausência de fala dos quadros do sinal analisado, pode ser definido a partir da razão entre as PDFs condicionadas calculadas na Equação (62), ou seja:

$$\begin{aligned}
\Lambda_k &= \frac{p(\bar{\mathbf{X}}_k|H_{1,k})}{p(\bar{\mathbf{X}}_k|H_{0,k})}, \\
&= \frac{1}{\pi(\lambda_{N,k} + \lambda_{V,k})} \pi\lambda_{N,k} \exp\left\{-\frac{|\bar{\mathbf{X}}_k|^2}{\lambda_{N,k} + \lambda_{V,k}} + \frac{|\bar{\mathbf{X}}_k|^2}{\lambda_{N,k}}\right\}, \\
&= \frac{\lambda_{N,k}}{(\lambda_{N,k} + \lambda_{V,k})} \exp\left\{\frac{-\lambda_{N,k}|\bar{\mathbf{X}}_k|^2 + (\lambda_{N,k} + \lambda_{V,k})|\bar{\mathbf{X}}_k|^2}{(\lambda_{N,k} + \lambda_{V,k})\lambda_{N,k}}\right\}, \\
&= \frac{\lambda_{N,k}}{(\lambda_{N,k} + \lambda_{V,k})} \exp\left\{\frac{\lambda_{V,k}|\bar{\mathbf{X}}_k|^2}{(\lambda_{N,k} + \lambda_{V,k})\lambda_{N,k}}\right\}, \\
&= \frac{1}{\left(1 + \frac{\lambda_{V,k}}{\lambda_{N,k}}\right)} \exp\left\{\frac{|\bar{\mathbf{X}}_k|^2}{\lambda_{N,k}} \frac{1}{\left(1 + \frac{\lambda_{V,k}}{\lambda_{N,k}}\right)} \frac{\lambda_{V,k}}{\lambda_{N,k}}\right\}, \\
&= \frac{1}{1 + \xi_k} \exp\left\{\frac{(1 + \gamma_k)\xi_k}{1 + \xi_k}\right\},
\end{aligned} \tag{63}$$

em que Λ_k é o k -ésimo estimador estatístico de máxima verossimilhança, ξ_k e γ_k são interpretadas respectivamente como a Relação Sinal Ruído (SNR, do inglês *Signal-to-Noise Ratio*) a priori e a posteriori [54], que podem ser mensuradas conforme proposto em [52] a partir de:

$$\xi_k = \frac{\lambda_{V,k}}{\lambda_{N,k}}, \tag{64}$$

$$\gamma_k = \frac{|\bar{\mathbf{X}}_k|^2}{\lambda_{N,k}} - 1. \tag{65}$$

Não obstante, ainda permanecemos com algumas variáveis desconhecidas, tais como as variâncias, λ_N e λ_V . Em vista disso, investigou-se métodos na literatura com o intuito de contornar esta dificuldade. A seguir, descreveremos um método para estimar a variância

referente ao ruído de fundo (λ_N), proposto em [52], que parte do pressuposto de que a probabilidade de ausência de fala em um trecho de sinal, com auxílio do teorema de Bayes, pode ser expresso por:

$$p(H_{0,k}|\bar{\mathbf{X}}_k) = \frac{1}{1 + \frac{[1-p(H_{0,k})]\Psi_k}{p(H_{0,k})}}, \quad (66)$$

em que $p(H_{0,k})$ corresponde a probabilidade a priori, definida posteriormente e Ψ_k representa um parâmetro de memória do sistema, que por sua vez, para a i -ésima iteração pode ser expresso por:

$$\Psi_k^i = e^{[\Upsilon \log(\Psi_k^{i-1})]} + (1 - \Upsilon) \log(\Lambda_k^i), \quad (67)$$

em que Ψ_k^{i-1} representa o parâmetro de amortecimento do sistema na iteração anterior e Υ representa o fator de amortecimento atual.

É preciso definir agora a probabilidade a priori, apresentada na Equação (66), que conforme apresentada em [55] e adaptada em [49], sugere que podemos expressar a probabilidade a priori por:

$$p(H_{0,k}) = \frac{1}{1 + e^{-\xi_k} I_0(2\sqrt{\gamma_k \xi_k})}, \quad (68)$$

onde I_0 representa a função de Bessel modificada de ordem zero.

Neste âmbito, podemos expressar a potência do ruído de fundo para a i -ésima iteração por:

$$E\left(|\bar{\mathbf{N}}_k^i|^2 \mid \bar{\mathbf{X}}_k^i\right) = p(H_{0,k}|\bar{\mathbf{X}}_k^i) |\bar{\mathbf{X}}_k^i|^2 + \left[1 - p(H_{0,k}|\bar{\mathbf{X}}_k^i)\right] \lambda_{N,k}^{i-1}. \quad (69)$$

Através da potência do ruído de fundo, podemos então finalmente expressar a variância de ruído de fundo, na i -ésima iteração por:

$$\lambda_{N,k}^i = \eta \lambda_{N,k}^{i-1} + (1 - \eta) E\left(|\bar{\mathbf{N}}_k^i|^2 \mid \bar{\mathbf{X}}_k^i\right), \quad (70)$$

em que η representa o fator de controle de aprendizagem ou de adaptação.

Entretanto, estimar a variância referente ao espectro fala (λ_V , Equação (64)), é uma tarefa mais complexa. Em vista disso, em [52] é proposto que ξ_k seja calculado sem o conhecimento direto de λ_V , da seguinte maneira:

$$\xi_k^i = \frac{\alpha |\widehat{\mathbf{V}}_k^{i-1}|^2}{\lambda_{N,k}^{i-1}} + (1 - \alpha)\max(\gamma_k, 0), \quad (71)$$

em que α representa a uma constante de ponderação e $|\widehat{\mathbf{V}}_k|$ é a amplitude espectral da parcela de fala do sinal desejado, estimada segundo [56], empregando a técnica de minimização do erro quadrático médio, representado por:

$$\widehat{\mathbf{V}}_k = \frac{\xi_k}{1 + \xi_k} \overline{\mathbf{X}}_k. \quad (72)$$

Uma vez definido os procedimentos para possibilitar os cálculos de ξ_k e γ_k , que por sua vez, permitem o cálculo de Λ_k para todas as raias em cada quadro do sinal contemplado, o critério de decisão de atividade de voz é realizado a cada quadro, e não a cada raia da DFT, em virtude disso, defini-se em [49] e [57] uma forma com o intuito de se avaliar a existência de fala a cada quadro da seguinte forma:

$$\Psi = \left(\prod_{k=1}^K \Psi_k \right)^{\frac{1}{K}} \geq L_{EST}, \quad (73)$$

em que L_{EST} é o limiar estatístico segundo o qual o detector de voz SMBVAD realiza a classificação do quadro analisado e Ψ representa a média geométrica de Ψ_k para todas as K raias.

Visando obter um desempenho satisfatório ao algoritmo de BSS foco deste trabalho, utilizamos os valores, apresentados na Tabela 6 para o SMBVAD.

Tabela 6 - Parâmetros empregados no SMBVAD.

Tamanho do Quadro	64
Limiar Estatístico (L_{EST})	1,001
Fator de Amortecimento (Υ)	0,97
Constante de Ponderação (α)	0,96
Fator de Controle de Aprendizagem (η)	0,80

A Figura 21 apresenta o resultado do SMBVAD para a mesma mistura analisada na Figura 18 pelo ALED, com intervalos desprovidos de fala. Observa-se um desempenho similar tanto ao ALED quanto ao LSED, entretanto o custo computacional é superior, resultando inclusive em um tempo de resposta maior quando comparado aos demais detectores. Visando eficiência e baixo custo computacional escolhemos para este trabalho o ALED.

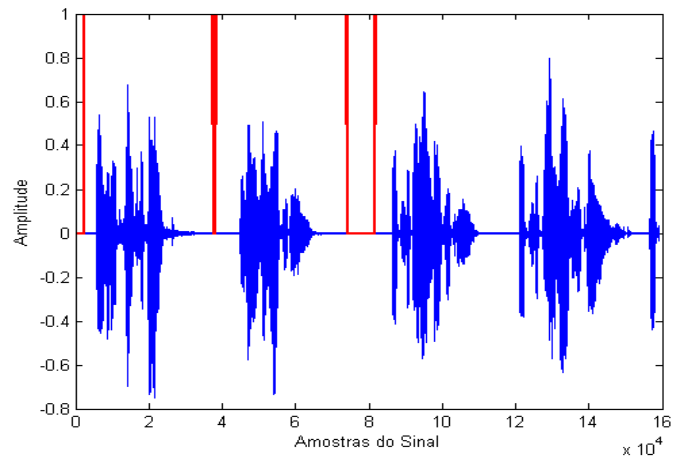


Figura 21 - Desempenho do SMBVAD empregado em mistura com presença de trechos de silêncio.

4 BANCO DE FILTROS E BSS EM SUB-BANDAS

4.1 Introdução

Este capítulo introduz os conceitos básicos de banco de filtros, obtidos a partir de modulação por cosseno de um filtro protótipo e o emprego de um sistema multitaxa, com o objetivo de dividir o sinal em bandas distintas, com diferentes taxas de amostragem, aplicando em cada sub-banda a técnica de BSS apresentada no Capítulo 2, que é o foco deste trabalho. Serão investigadas neste trabalho duas estruturas de banco de filtros: a estrutura uniforme e a estrutura não uniforme.

4.2 Sistemas Multitaxas

Os sistemas multitaxas realizam alterações na taxa de amostragem do sinal em análise, a partir da inserção ou exclusão de amostras do sinal. Tal processamento é realizado através do emprego de determinados operadores multitaxas, em geral, decimadores e expansores, em conjunto com filtros digitais [58].

4.2.1 Decimador

O decimador, apresentado na Figura 22, também denominado em algumas literaturas como “*downsampler*” ou “compressor” [20] é um operador multitaxa responsável por decimar um sinal de entrada por um fator FD , reduzindo assim em FD vezes a taxa de amostragem do sinal. Em outras palavras, na saída do aludido operador, cuja entrada é um sinal $x(n)$, encontram-se apenas as amostras de $x(n)$ em que n é múltiplo de FD , tal comportamento é ilustrado na Figura 22 (adaptada de [20]), podemos então expressar a resposta do decimador por:

$$y_D(n) = x(nFD), \quad (74)$$

sendo FD um número inteiro, denominado Fator de Decimação.

4.2.2 Expansor

O expansor, apresentado na Figura 23, que em alguns trabalhos é denominado “*upsampler*” ou “interpolador” [20] é outro importante operador multitaxa responsável por expandir

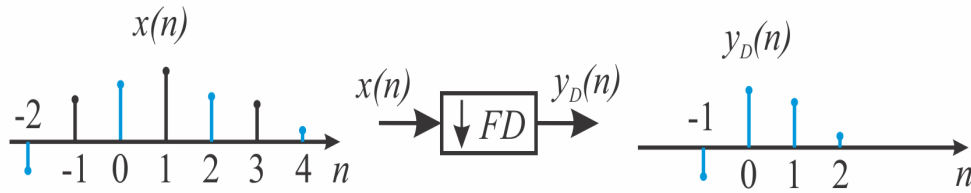


Figura 22 - Ilustração de funcionamento do decimador para $FD = 2$.

o sinal de entrada por um fator FD , via inserção de $FD - 1$ zeros entre cada amostra do sinal, destarte a taxa de amostragem do sinal é elevada em FD vezes. Analogamente ao decimador, podemos expressar a resposta do expansor, cujo comportamento é ilustrado na Figura 23 (adaptada de [20]), para um sinal de entrada $x(n)$ por:

$$y_E(n) = \begin{cases} x(n/FD), & \text{se } n \text{ é um inteiro múltiplo de } FD; \\ 0, & \text{outros valores.} \end{cases} \quad (75)$$

Cabe ressaltar que no expansor não temos perda de informação, conforme ocorre no decimador.

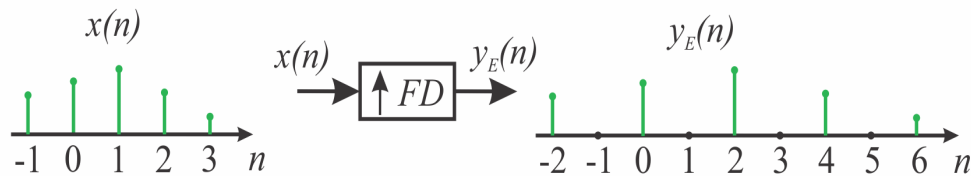


Figura 23 - Ilustração de funcionamento do expansor para $FD = 2$.

4.2.3 Banco de filtros

Um banco de filtros é uma estrutura composta por um conjunto de filtros⁶, que tem como característica, a entrada ou a saída comum, conforme apresentado na Figura 24. Em geral, a estrutura de um banco de filtros de M bandas ou canais, costuma apresentar um filtro passa-baixa, $M - 2$ filtros passa-banda e um passa-alta, todavia a largura dos filtros

⁶Por comodidade, costuma-se representar os sistemas discretos no tempo via Transformada z , que é uma generalização da Transformada de Fourier.

dependerá da estrutura do banco de filtros, a qual deseja-se trabalhar, caso todos os filtros que compõem o banco possuam a mesma largura, temos uma estrutura conhecida como banco de filtros em estrutura uniforme, caso contrário temos um banco de filtros em estrutura não uniforme.

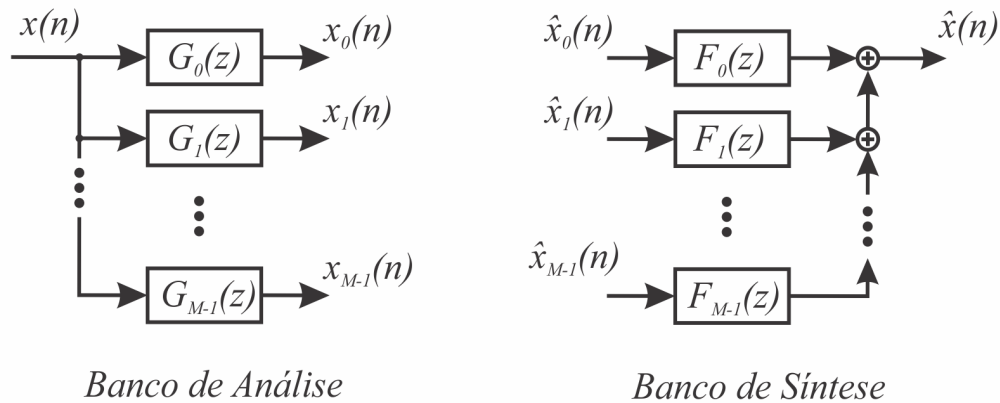


Figura 24 - Ilustração de bancos de filtros de análise e síntese respectivamente.

A finalidade de um banco de filtros que possui a entrada comum, denominado banco de análise, apresentado na Figura 24, é decompor um sinal de entrada $x(n)$ em várias sub-bandas de frequência, isto é, significa dividir o espectro de frequência do sinal $x(n)$ em M sub-bandas distintas. Por outro lado, a finalidade de um banco de filtros que possui a saída comum, denominado banco de síntese, também ilustrado na Figura 24, é compor um sinal $\hat{x}(n)$ através de várias sub-bandas, ou seja, significa recompor (recuperar) o sinal $x(n)$, a partir das M sub-bandas do sinal decomposto.

4.3 Banco de Filtros em Estrutura Uniforme

Um sistema multitaxa completo denominado banco de filtros em estrutura uniforme, empregado neste trabalho e apresentado na Figura 25 (adaptação de [19]), integra os componentes multitaxas supracitados, cujos filtros, $\{G_0(z), G_1(z), \dots, G_{M-1}(z)\}$ e $\{F_0(z), F_1(z), \dots, F_{M-1}(z)\}$ denominados respectivamente banco de análise e síntese, possuem a mesma largura de faixa de passagem.

No sistema apresentado na Figura 25, um sinal de entrada $x(n)$, é decomposto em M sub-bandas através do banco de análise, cuja saída resulta em sinais $x_k(n)$, para $k = 0, 1, \dots, M - 1$, que são decimados em uma etapa seguinte por um respectivo fator inteiro FD_k , que por sua vez resultam em sinais denominados $x_{D,k}(n)$. A decomposição em sub-

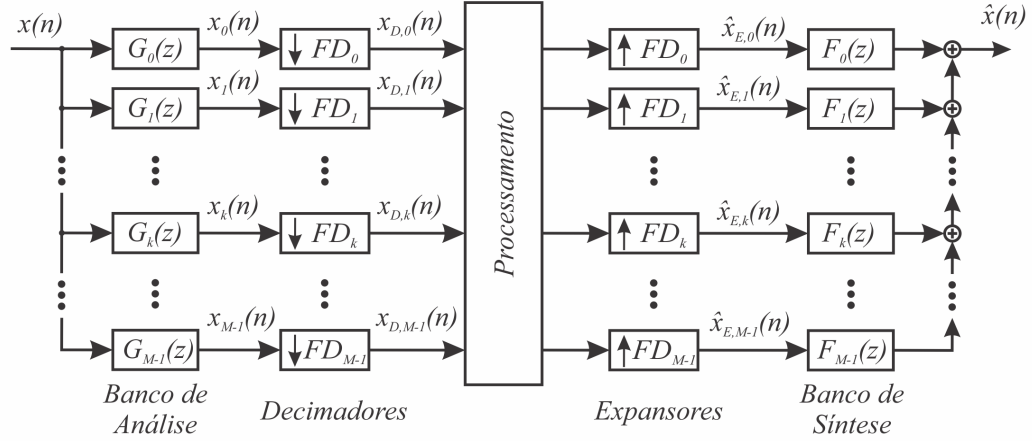


Figura 25 - Ilustração de um sistema multitaxa completo em estrutura uniforme.

banda possibilita o processamento dos sinais $x_{D,k}(n)$ em M bandas distintas do sinal $x(n)$, que podem estar a uma taxa de amostragem de até $FD_k = M$ menor que a taxa de amostragem do sinal de entrada. Após o processamento, os sinais em sub-bandas são expandidos pelo mesmo fator FD_k originando sinais $x_{E,k}(n)$ que posteriormente são recombinaados pelo banco de síntese, resultando no sinal $\hat{x}(n)$, cuja taxa de amostragem é a mesma do sinal de entrada $x(n)$ [19].

Uma outra classificação pertinente a sistemas multitaxas é quanto à estrutura ser maximamente decimada e não maximamente decimada. Considera-se que uma estrutura é maximamente ou criticamente decimada quando a seguinte condição é satisfeita:

$$\sum_{k=0}^{M-1} \frac{1}{FD_k} = 1, \quad (76)$$

sendo FD_k apresentado na Figura 25, o Fator de Decimação da k -ésima sub-banda. Caso contrário, temos uma estrutura não maximamente decimada.

Cumprer notar que mesmo sem a etapa de processamento dos sinais em sub-bandas, a estrutura multitaxa apresenta um erro de reconstrução dos próprios bancos de filtros, que pode ser provocado por um conjunto de fatores, dentre eles podemos citar: a distorção de amplitude, a distorção de fase e o *aliasing*, que é a sobreposição do espectro entre as respostas típicas dos filtros de análise e síntese, tal sobreposição pode ser verificada na Figura 26 [19] na qual apresentamos a resposta em frequência dos filtros $G_k(z)$, cuja resposta dos filtros $F_k(z)$ é análoga aos filtros $G_k(z)$ [19] e [20]. Não obstante, os impactos destes fatores podem ser minimizados durante a etapa de projeto do banco de filtros, propiciando uma reconstrução perfeita do sinal de entrada [20].

Destarte podemos expressar o sinal reconstruído, $\hat{x}(n)$, uma versão atrasada do sinal de entrada, por:

$$\hat{x}(n) = \frac{1}{FD}x(n - \Theta), \quad (77)$$

em que o fator FD no denominador é uma atenuação oriunda da etapa de decimação passível de ser compensada na construção dos bancos de análise e síntese e Θ representa o atraso inserido pelo banco de filtros.

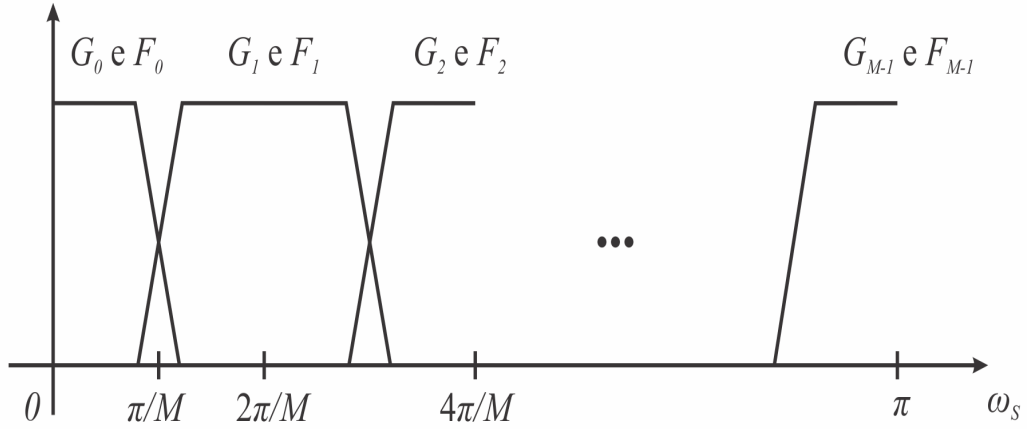


Figura 26 - Resposta em frequência dos filtros $G_k(z)$.

É imprescindível destacar que no sistema multitaxa, o custo computacional do processamento será elevado caso não ocorra a decimação dos sinais $x_k(n)$, ou seja, caso FD_k seja unitário, para todo k , uma vez que o conjunto de todas as amostras oriundas do banco de filtros será M vezes maior do que o total do sinal original [59]. Por outro lado uma estrutura uniforme não maximamente decimada ($FD_k < M$) implica em redução da sobreposição resultante da resposta em frequência na saída do banco de análise e do decimador (reduzindo assim o *aliasing*) além de possivelmente proporcionar uma suavização nas transições das faixas de passagem e rejeição nos filtros do banco [19] e [20].

4.4 Banco de Filtros em Estrutura Não Uniforme

Um sistema multitaxa completo denominado banco de filtros em estrutura não uniforme, é um sistema que também integra os componentes multitaxas, todavia os filtros, $\{G^{0,0}(z), G^{0,1}(z), \dots, G^{0,M-1}(z)\}$ e $\{F^{0,0}(z), F^{0,1}(z), \dots, F^{0,M-1}(z)\}$ ⁷ denominados banco de

⁷O primeiro índice sobrescrito representa a configuração do subfiltro empregado, caso seja 0 temos um subfiltro passabaixa e caso seja 1 temos um subfiltro passa-alta, já o segundo índice sobrescrito representa o nível da estrutura em árvore em observação.

análise e síntese respectivamente, não possuem, em geral, a mesma largura de faixa de passagem. Um sistema multitaxa em estrutura não uniforme amplamente explorado na literatura e contemplado nesta dissertação, é o banco de filtros em árvore, apresentado na Figura 26 (adaptação de [19]), obtido a partir de um arranjo entre um conjunto de bancos uniformes, cujas saídas são combinadas de tal modo que resulta em um banco não uniforme.

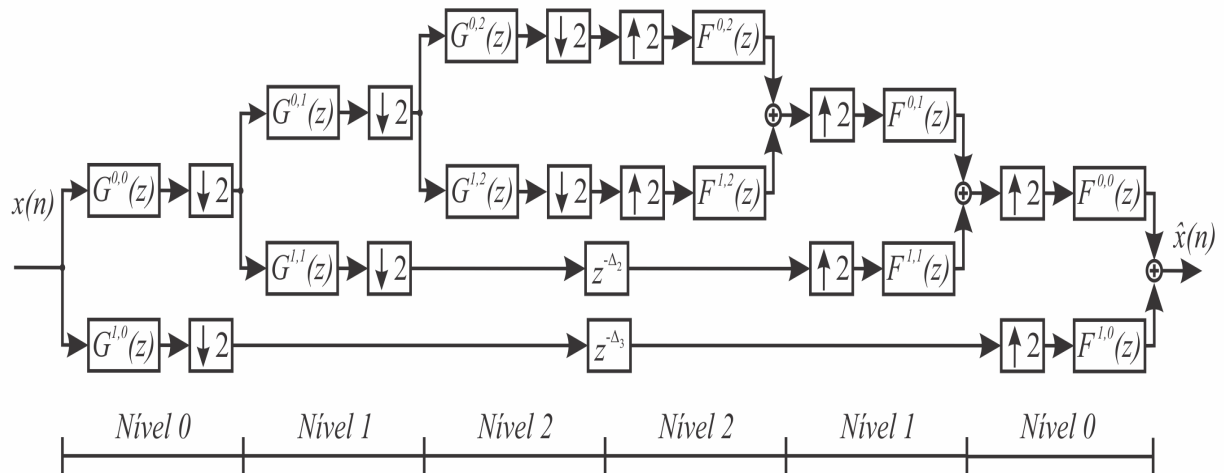


Figura 27 - Ilustração de um sistema multitaxa completo em estrutura não uniforme maximamente decimado para $M = 4$.

Uma estrutura não uniforme é dita maximamente decimada quando a condição na Equação (76) é satisfeita, vide Figura 27, caso contrário temos uma estrutura não maximamente decimada.

A estrutura em árvore contemplada na Figura 27 costuma ser analisada em níveis, sendo que em cada nível há uma característica comum que é a existência de um banco de filtros uniforme para $M = 2$, junto ao operador multitaxa correspondente (o decimador em conjunto com o banco de análise e o expensor com o banco de síntese).

Na estrutura não uniforme um sinal de entrada $x(n)$, é decomposto inicialmente em duas bandas por um banco de análise uniforme, que posteriormente são decimadas. Em um nível seguinte, uma destas sub-bandas é novamente decomposto por um banco de análise uniforme de duas bandas e em seguida, os sinais resultantes são devidamente decimados, tal ciclo repete-se para $M - 1$ níveis (sendo M o número de sub-bandas). O processo de recominação (reconstrução) do sinal para esta estrutura ocorre seguindo um raciocínio inverso ao procedimento supracitado, realizando é claro, algumas alterações como a substituição do banco de análise uniforme pelo banco de síntese uniforme e o decimador pelo expensor. Face ao exposto, o sinal $\hat{x}(n)$ obtido, contempla a mesma taxa de amostragem do sinal de entrada

$x(n)$ [19].

Para facilitar a compreensão, podemos fazer uso das identidades nobres [20] para modelar uma representação equivalente para uma estrutura em árvore, com este intuito, na Figura 28, apresentamos uma estrutura equivalente a estrutura ilustrada na Figura 27.

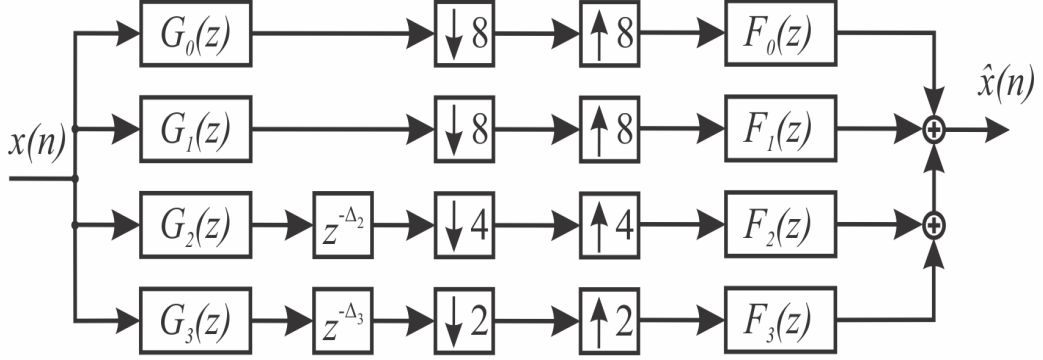


Figura 28 - Representação equivalente de um sistema multitaxa completo em estrutura não uniforme maximamente decimado para $M = 4$.

Podemos expressar os M filtros de análise, pertinentes a representação equivalente por:

$$G_0(z) = \prod_{j=0}^{M-2} G^{0,j}(z^{2^j}),$$

$$G_k(z) = G^{(1,M-1-k)} z^{2^{(M-1-k)}} \prod_{j=0}^{M-k-2} G^{0,j}(z^{2^j}), \text{ para } k = 1, 2, \dots, M-1, \quad (78)$$

por analogia temos que os M filtros de síntese podem ser representados por:

$$F_0(z) = \prod_{j=0}^{M-2} F^{0,j}(z^{2^j}),$$

$$F_k(z) = F^{(1,M-1-k)} z^{2^{(M-1-k)}} \prod_{j=0}^{M-k-2} F^{0,j}(z^{2^j}), \text{ para } k = 1, 2, \dots, M-1. \quad (79)$$

Podemos também representar o Fator de Decimação para esta representação equivalente da seguinte maneira:

$$\begin{cases} FD_0 = 2^{M-1}, \\ FD_k = 2^{M-k}, \end{cases} \quad (80)$$

sendo $1 \leq k < M - 1$ e FD_k equivale ao fator de decimação resultante para a k -ésima sub-banda para esta estrutura equivalente.

A resposta em frequência dos filtros $G_k(z)$, que é similar a resposta dos filtros $F_k(z)$, nesta estrutura não uniforme em árvore é apresentada na Figura 29 [19] e [20]. Cada nível incrementado nesta estrutura, resulta em reduzir à metade a largura da banda de passagem do subfiltro de menor índice, ou seja, $G_0(z)$ e $F_0(z)$, por este motivo esta estrutura também é conhecida na literatura como um banco de filtros não uniforme com decomposição de frequências em oitavas.

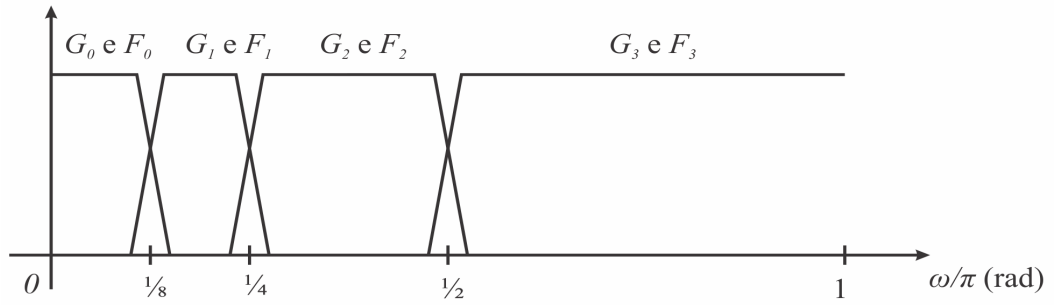


Figura 29 - Resposta em frequência dos filtros $G_k(z)$ em estrutura não uniforme em árvore para $M = 4$.

As figura 27 e 28 também contemplam os atrasos Δ_k , que são essenciais para compensar a diferença de comprimento dos filtros dos bancos de análise e síntese, consequência da estrutura não uniforme, bem como a influência dos decimadores e expansores nas sub-bandas. Em vista disso, faz-se necessário estimar os comprimentos dos filtros em todas as sub-bandas, destarte seja N_{G_k} o comprimento do filtro G_k , que possui o mesmo comprimento do filtro F_k , na k -ésima sub-banda na representação equivalente, temos que:

$$\begin{aligned} N_{G_0} &= \sum_{j=0}^{M-2} 2^j N_{G^0,j}, \\ N_{G_k} &= \sum_{j=0}^{M-k-2} 2^j N_{G^0,j} + 2^{M-k-1} N_{G^1,j}, \end{aligned} \quad (81)$$

em que $N_G^{0,j}$ e $N_G^{1,j}$ são os comprimentos dos filtros $G^{0,j}(z)$ e $G^{1,j}(z)$ respectivamente no j -ésimo nível. E assim, podemos expressar os atrasos Δ_k por:

$$\Delta_k = N_{G_0} - N_{G_k}. \quad (82)$$

Vale lembrar que para duas sub-bandas, ou seja, para $M = 2$, a estrutura não uniforme equivale a estrutura uniforme, isso porque necessitamos apenas de um nível para obter tal resultado, por conseguinte, resulta em empregar apenas um banco uniforme para $M = 2$, que conforme mencionado anteriormente, possuem filtros de mesmo tamanho. Por este motivo, os resultados (apresentados no Capítulo 5) obtidos para $M = 2$ são iguais nestas duas estruturas.

4.5 Bancos de Filtros Modulados por Cosseno

Almejando a reconstrução perfeita (PR, do inglês *Perfect Reconstruction*) do sinal, que é um grande desafio no sistema multitaxa e alvo de diversos trabalhos na literatura, optamos em construir os bancos de filtros de análise e de síntese, empregando um banco de filtros de modulação por cosseno (CMFB, do inglês *Cosine Modulated Filter Bank*).

O emprego de um CMFB neste trabalho tanto para todos os M filtros do banco de análise quanto para os M filtros do banco de síntese, deve-se a sua eficiência tanto para o projeto quanto para a implementação [60], e visa explorar algumas vantagens deste tipo de implementação, dentre elas podemos citar:

- Cada filtro do banco de análise possui o mesmo comprimento do correspondente filtro do banco de síntese [61];
- A técnica para a estrutura multitaxa pode ser empregada para um número arbitrário M de sub-bandas [61];
- O custo computacional necessário para construir tanto de um banco de análise quanto de um banco de síntese é igual ao custo computacional de um único filtro e de uma DCT;
- O número de parâmetros a serem otimizados, durante a etapa de projeto, é muito pequena, visto que apenas o filtro protótipo precisa ser otimizado;
- Um CMFB possibilita que os filtros dos banco de análise e síntese tenham coeficientes reais, caso o filtro protótipo tenha coeficientes reais.

O filtro protótipo utilizado neste trabalho na construção dos bancos de filtros é um filtro FIR, cuja resposta de frequência, conforme observado na Figura 30, é simétrica em relação a $\omega = 0$, possui coeficientes reais. Além disso, o filtro protótipo projetado segundo [20] e [62] tem por objetivo realizar a reconstrução perfeita do sinal.

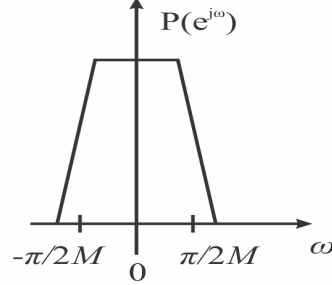


Figura 30 - Resposta em frequência do filtro protótipo.

Em geral, considera-se N_p o comprimento do filtro protótipo que costuma ser escolhido sendo um múltiplo par do número de sub-bandas, ou seja, $N_p = 2K_pM$, sendo K_p um número inteiro (que para este trabalho escolhemos $K_p = 8$, logo $N_p = 16M$) e $p(n)$ denota a resposta ao impulso do filtro protótipo. Destarte os M filtros dos bancos de análise e síntese podem ser obtidos da seguinte maneira:

$$\begin{aligned} g_k(n) &= 2p(n) \cos \left[\frac{\pi}{M}(k + 0,5) \left(n - \frac{\Omega}{2} \right) + \theta_k \right], \\ f_k(n) &= 2p(n) \cos \left[\frac{\pi}{M}(k + 0,5) \left(n - \frac{\Omega}{2} \right) - \theta_k \right], \end{aligned} \quad (83)$$

sendo $\Omega = N_p - 1$ e $\theta_k = (\pi/4)(-1)^k$, para $0 \leq k \leq M - 1$ e $0 \leq n \leq N_p - 1$.

4.6 Implementação de BSS em sub-bandas

Face ao exposto, podemos aplicar a técnica de BSS escolhida (ver Capítulo 2) em cada sub-banda, de forma independente, a partir da implementação dos sistemas multitaxas supracitados (tanto a estrutura uniforme quanto a não uniforme). Com o intuito de realizar a BSS em sub-bandas, apresentamos na Figura 31, um modelo completo⁸ de separação cega de fontes em sub-bandas para a configuração de um sistema TITO, considerando um CMFB de M canais.

⁸Por simplicidade os atrasos existentes nas estruturas não uniformes não foram contemplados neste modelo.

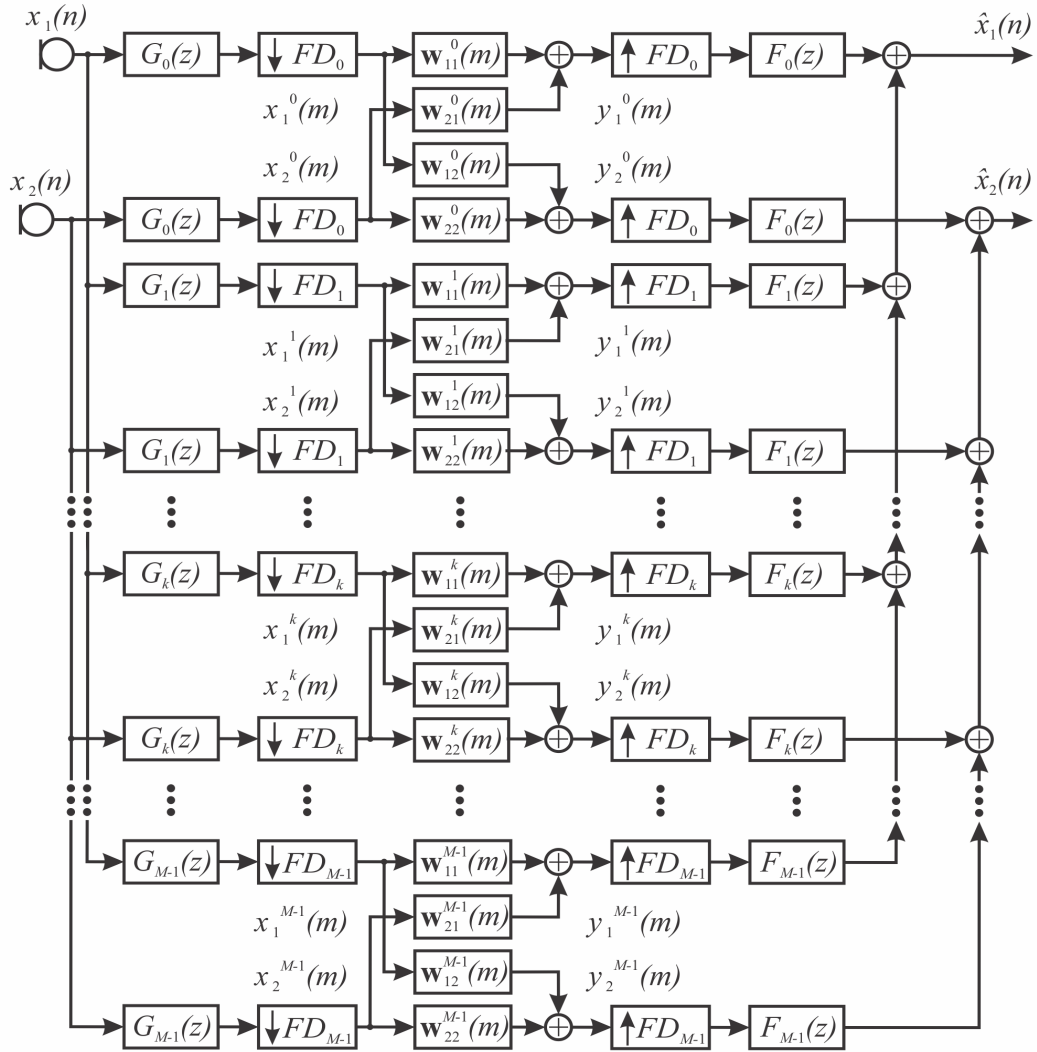


Figura 31 - Ilustração de BSS em sub-bandas para um sistema TITO.

Na aludida figura temos que o sinal de mistura $x_q(n)$ é decomposto pelos filtros de análise $G_k(z)$ originando M sub-bandas, que são devidamente decimadas pelo respectivo FD_k , e em seguida, cada sinal $x_q^k(n)$ resultante, é aplicado isoladamente a um conjunto de filtros de separação $w_{pq}^k(m)$. Nesta etapa ocorre o processamento do sinal pertinente à BSS e ao término deste processamento os sinais em cada sub-banda são expandidos pelo FD_k pertinente à sub-banda e posteriormente recombinados pelos filtros de síntese $F_k(n)$ para restaurar os sinais de saída em banda cheia (estimativa das fontes).

Por simplicidade expomos na Figura 32 [15], uma generalização para o modelo de BSS em sub-bandas para a k -ésima sub-banda.

Visto que a BSS ocorre independentemente em cada sub-banda, podemos então adequar para a k -ésima sub-banda a Equação (51), responsável pela atualização dos filtros no algoritmo de BSS escolhido no Capítulo 2 para uma implementação *offline*, da seguinte ma-

neira:

$$\mathbf{W}_j^k = \mathbf{W}_{j-1}^k - \frac{2\mu^k(j)}{b^k} \sum_{i=1}^{b-1} \left[\mathbf{W}^k \left(\mathbf{R}_{\mathbf{y}\mathbf{y}}^k(i) - \text{bdiag} \mathbf{R}_{\mathbf{y}\mathbf{y}}^k(i) \right) \text{bdiag}^{-1} \mathbf{R}_{\mathbf{y}\mathbf{y}}^k(i) \right], \quad (84)$$

em que o sobrescrito $(.)^k$ denota a sub-banda na qual será aplicada o processamento de BSS descrito no Capítulo 2, visto que a separação cega de fontes é realizada isoladamente em cada sub-banda.

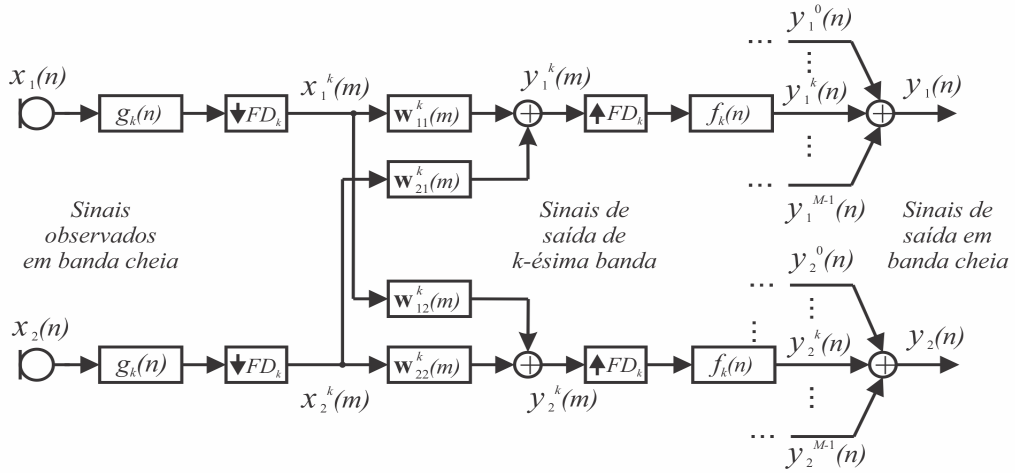


Figura 32 - Ilustração para o k -ésimo canal da configuração linear TITO para BSS em sub-bandas.

5 COMPLEXIDADE COMPUTACIONAL E RESULTADOS EXPERIMENTAIS

5.1 Introdução

Neste capítulo apresentaremos a dedução da complexidade computacional para o algoritmo de separação cega escolhido (ver Capítulo 2), descreveremos os experimentos realizados e apresentaremos os resultados obtidos, tanto para a implementação em banda cheia quanto em sub-banda para o .

5.2 Complexidade computacional

Nesta seção investigaremos a complexidade computacional dos algoritmos contemplados nesta proposta, não somente por blocos, mas também, por iteração. A complexidade computacional por iteração é fundamental para expor a importância do detector de voz no algoritmo de separação, uma vez que o emprego do mesmo, possibilita desativar o algoritmo de separação caso não haja a presença de voz no bloco em análise, proporcionando uma redução do número de blocos por iteração, que por sua vez resulta em redução da complexidade computacional, visto que o custo para estimar um bloco com ausência de fala não é mais necessário. É importante mencionar que a complexidade computacional envolvida para estimar o custo tanto do algoritmo do detector de voz empregado nos experimentos, o ALED, quanto do algoritmo responsável por dividir e reconstruir os sinais em sub-bandas, foram desprezados, visto que essas etapas ocorrem, quando necessárias, apenas uma única vez para todas as iterações.

Podemos inicialmente, estimar o número de multiplicações reais por bloco do sinal de entrada, necessárias para atualizar os coeficientes dos filtros FIR de separação, para a estrutura em banda cheia ($NMPB_{BC}$), considerando um sistema linear MIMO para o caso determinado, ou seja, $P = Q$ (Número de misturas iguais ao número de fontes) por:

$$NMPB_{BC} = P^2 NSD + PN + (P^2 - P) DN + P^2 \left[\left(\frac{S^2 + S}{2} \right) + S \right]. \quad (85)$$

A primeira parcela representa o custo para gerar todos os blocos atrasados no tempo de todas as saídas Equação (24), considerando que a matriz \mathbf{W}_{pq} (Equação (27)) possui apenas S elementos diferentes de zero por coluna. O segundo e terceiro termo correspondem

respectivamente ao custo necessário para se obter as submatrizes $\mathbf{R}_{\mathbf{y}_q \mathbf{y}_q}^{(m)}$ e $\mathbf{R}_{\mathbf{y}_p \mathbf{y}_q}^{(m)}$ da matriz de correlação, $\mathbf{R}_{\mathbf{y}\mathbf{y}}^{(m)}$ (Equação (35)) empregadas no cálculo do gradiente, considerando que apenas os S primeiros coeficientes da primeira coluna da matriz de adaptação são atualizados. Por fim a quarta parcela faz menção ao custo da adaptação contemplando apenas a atualização das S primeiras linhas da primeira coluna da matriz \mathbf{W}_{pq} (Equação (27)).

Cumpramos notar que consideramos em todas as expressões supracitadas, o fator de normalização simplificado da Equação (47), $D = S$ e $N = 2S$. Desta forma, podemos expressar o $NMPB_{BC}$ para a estrutura em banda cheia por:

$$NMPB_{BC} = \frac{P^2}{2} (4S^3 + 5S^2 + 3S) - 2P (S^2 - S). \quad (86)$$

É possível ainda simplificar a Equação (86) considerando que o comprimento dos filtros de separação seja elevado, ou seja, ($S \gg 1$), e assim, obtemos:

$$NMPB_{BC} \approx 2P^2 S^3. \quad (87)$$

Os cálculos efetuados para a complexidade computacional em banda cheia são os mesmos realizados em cada sub-banda, exceto pelo fato que em sub-bandas existe também o fator de decimação (FD). Assim, podemos generalizar a Equação (87) para M bandas, cada uma efetuando o processamento com o seu respectivo FD_k , por:

$$NMPB = \frac{P^2}{2} \sum_{k=0}^{M-1} \left(\frac{4S_k^3 + 5S_k^2 + 3S_k}{FD_k} \right) - 2P \sum_{k=0}^{M-1} \left(\frac{S_k^2 - S_k}{FD_k} \right), \quad (88)$$

em que o índice k , representa a k -ésima sub-banda, em que consideramos $D_k = S_k$ e $N_k = 2S_k$.

A Equação (88) possibilita estimar a complexidade computacional por bloco para o algoritmo de separação contemplado neste trabalho, tanto para banda cheia, na qual consideramos ($M = 1$ e $FD = 1$), quanto em sub-banda.

A partir da Equação (88) podemos mensurar a complexidade computacional para o número de multiplicações por iteração ($NMPI$) tanto em banda cheia, quanto em sub-banda por:

$$NMPI = \frac{P^2}{2} \sum_{k=0}^{M-1} b^k \left(\frac{4S_k^3 + 5S_k^2 + 3S_k}{FD_k} \right) - 2P \sum_{k=0}^{M-1} b^k \left(\frac{S_k^2 - S_k}{FD_k} \right), \quad (89)$$

em que b^k indica o número de blocos empregados no algoritmo de separação por iteração em cada sub-banda. Cabe ressaltar que sem o emprego do detector de voz, todas as sub-bandas

possuem o mesmo número de blocos, todavia utilizando o detector de voz é possível reduzir o número de blocos em cada sub-banda independentemente desde que o detector não reconheça a presença de fala em algum bloco.

A complexidade computacional apresentada na Equação (89) pode ser reduzida considerando uma implementação eficiente, utilizando-se FFTs (do inglês *Fast Fourier Transform*). Tal implementação é possível, devido ao fato de empregarmos estruturas matriciais conhecidas como *Toeplitz* (Equação (23)) e *Sylvester* (Equação (29)) na dedução dos sinais de saída e na atualização dos filtros de separação, que por sua vez, nos permite interpretar os produtos matriciais envolvidos nos cálculos, como convoluções lineares, e assim, podemos expressar por:

$$NMPI_{IE} = P^2 \sum_{k=0}^{M-1} b^k \left(\frac{S_k (13,75 + 6\log_2 S_k)}{FD_k} \right) - 2P \sum_{k=0}^{M-1} b^k \left(\frac{S_k (1 + \log_2 S_k)}{FD_k} \right), \quad (90)$$

em que $NMPI_{IE}$ corresponde ao número de multiplicações por iteração na implementação eficiente tanto em banda cheia, quanto em sub-banda.

5.3 Resultados experimentais

5.3.1 Considerações iniciais

Todos os experimentos foram realizados empregando sinais de fala com 10 segundos de duração e frequência de amostragem de $16kHz$. Além disso, todos os algoritmos foram executados em um processador Intel Core i5 2, $27GHz$.

Simulações computacionais comparam o desempenho do algoritmo de BSS *offline* implementado no domínio do tempo e adaptado para realizar a separação utilizando ou não detectores de voz, tanto em banda cheia ($M = 1$ e $FD = 1$) quanto em sub-bandas. Adotamos em todos os casos, os valores médios das estimativas das métricas SIR, SAR e SDR, definidas na Seção 1.9, para avaliar o desempenho do algoritmo, destarte, para a configuração TITO, podemos expressar as aludidas métricas em decibéis (dB) por:

$$SIR = 10\log_{10} \left(\frac{SIR_1 + SIR_2}{2} \right), \quad (91)$$

$$SAR = 10\log_{10} \left(\frac{SAR_1 + SAR_2}{2} \right), \quad (92)$$

$$SDR = 10 \log_{10} \left(\frac{SDR_1 + SDR_2}{2} \right). \quad (93)$$

Diversos sinais de voz foram empregados neste trabalho, com o intuito de ajustar, tanto os parâmetros do algoritmo de BSS, quanto dos detectores, todavia, por simplicidade, escolhemos quatro sinais de voz, para apresentar os resultados obtidos neste trabalho. Face ao exposto, os sinais escolhidos foram divididos em dois grupos, denominados:

- **Mistura 1:** Composto por duas vozes femininas (em inglês) desprovidas de intervalos de silêncio;
- **Mistura 2:** Composto por duas vozes femininas (em inglês) com a presença de intervalos de silêncio.

A sala acústica virtual de dimensões $4,45m \times 3,55m \times 2,5m$, apresentada na Figura 33, proposta em [63], foi empregada para obter de forma sintética as misturas utilizadas na maioria dos experimentos realizados, nos quais adotamos a configuração TITO ($P = Q = 2$), considerando os dois microfones/sensores espaçados de $5cm$ e que as fontes sonoras foram posicionadas a $1m$ de distância do ponto médio entre os sensores, em duas direções diferentes: -45° e 45° .

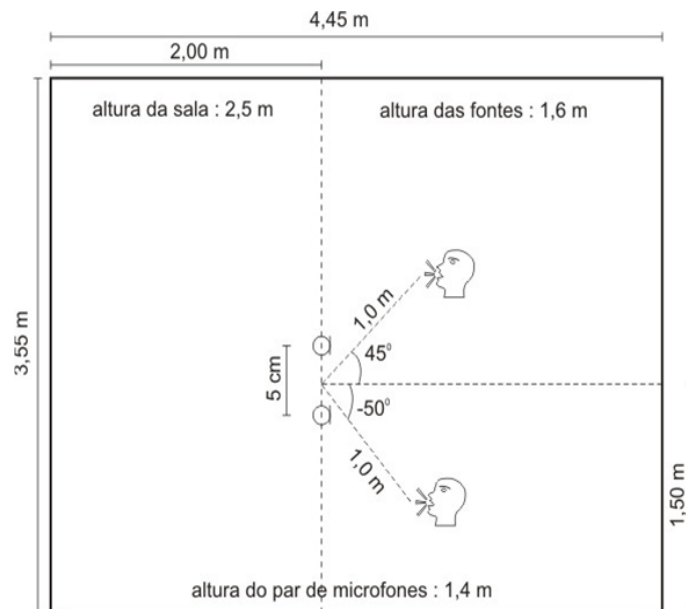


Figura 33 - Sala acústica virtual [63].

Destarte a aquisição das misturas ocorreu via convolução dos sinais de fala escolhidos com as respostas impulsivas sintéticas obtidas para a sala acústica virtual propostas em [63]

e [64], para diferentes tempos de reverberação do ambiente (T_{60}), truncando-se o resultado obtido (sinal misturado) para o tamanho desejado (igual ao sinal original).

As aludidas misturas obtidas para $T_{60} = 25ms$ são apresentadas na Figura 34, na qual é possível verificar que a Mistura 1 é desprovida de trechos de silêncio e a Mistura 2 apresenta intervalos de silêncio. Por questões de similaridade com a Figura 34, as misturas para $T_{60} = 100ms$ foram omitidas neste trabalho.

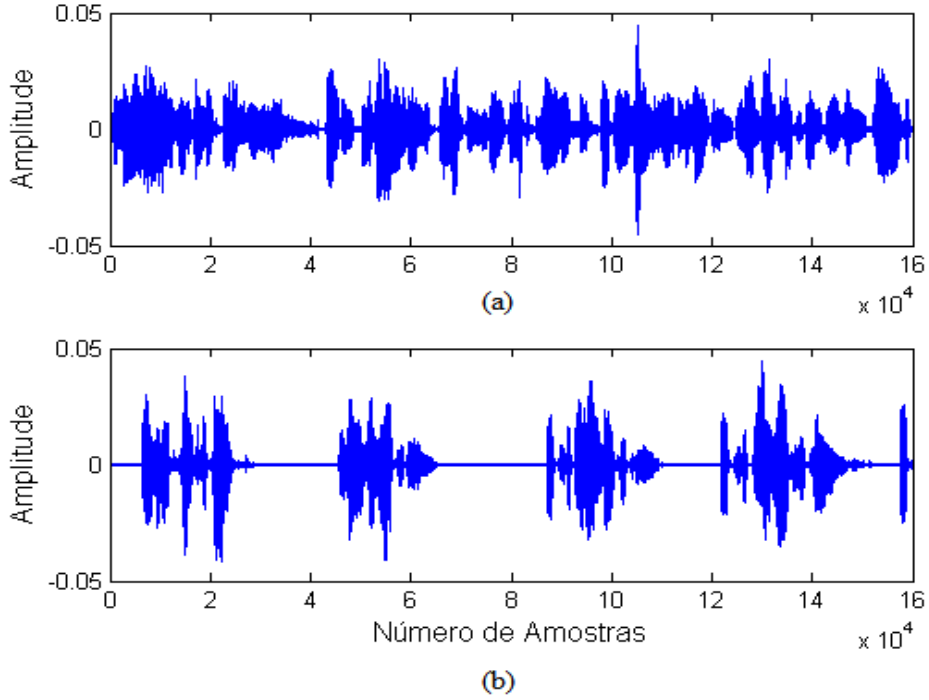


Figura 34 - Mistura obtidas para $T_{60} = 25ms$: (a) Mistura 1 e (b) Mistura 2.

Em todas as simulações do algoritmo de separação cega utilizamos 500 iterações. O valor inicial do fator de aprendizado adaptativo apresentado na Seção 2.4.4 em todos os casos foi $\mu^k(0) = 0,001$, para todas as sub-bandas. Os filtros de separação \mathbf{w}_{pq}^k foram sempre inicializados conforme exposto na Seção 2.4.5 e proposto em [14]. Conforme mencionado na complexidade computacional, adotamos o número de intervalos de tempo empregados nas estimativas de correlação $D_k = S_k$ e o tamanho dos blocos das saídas utilizados para estas estimativas $N_k = 2D_k$.

Para as implementações em sub-bandas foram utilizados bancos de filtros modulados por cosseno com $M = 2, 4$ e 8 bandas, implementados através de filtros protótipos PR com $N_p = 16M$, conforme supracitado na Seção 4.5, na qual investigamos o desempenho e a complexidade computacional do algoritmo de separação cega contemplando tanto a estrutura uniforme (experimento 1) quanto a estrutura não uniforme (experimento 2).

Em todos os experimentos expomos os resultados do desempenho do algoritmo de BSS escolhido e do custo computacional, tanto sem o emprego do detector de voz (SD), quanto com o uso do detector (CD), que neste trabalho conjugando eficiência e desempenho, escolhemos o ALED, cujos parâmetros e algoritmo são mencionados na Seção 3.2. A resposta do ALED (detector de voz empregado nesta proposta) mediante a aplicação nas misturas obtidas, são ilustradas na Figura 17 (Mistura 1) e na Figura 18 (Mistura 2).

5.3.2 Experimento 1

Neste experimento comparamos o desempenho e a complexidade computacional da separação cega de sinais para as misturas obtidas conforme supracitado na Seção 5.3.1 tanto em banda cheia quanto em sub-banda, empregando a estrutura uniforme apresentada na Seção 4.3, cuja configuração é ilustrada na Figura 25. Contemplamos diferentes condições de reverberação ($T_{60} = 25ms$ e $T_{60} = 100ms$) e também diferentes comprimentos de filtros FIR do sistema de separação, ($S_k = 256/FD_k$ e $768/FD_k$).

O fato dos filtros de separação serem atualizados de forma independente em cada sub-banda, possibilita trabalharmos com FD_k diferentes em bandas distintas, reduzindo assim, o *aliasing* causado pela atenuação finita dos filtros de análise e síntese. Assim, almejando um compromisso entre desempenho e complexidade computacional, simulamos versões mescladas para $M = 2$ ($M2$), onde a primeira banda (B_0) é implementada com $FD_0 = 1$ e a segunda banda (B_1) com $FD_1 = 2$; para $M = 4$ ($M4$), sendo a primeira banda (B_0) e a quarta banda (B_3) implementadas com $FD_{0e3} = 2$, e a segunda e terceira banda, respectivamente as bandas, B_1 e B_2 , realizadas com $FD_{1e2} = 1$; e para $M = 8$ ($M8$), nas quais a quarta banda ($B3$) e a quinta banda ($B4$) adotam $FD_{3e4} = 1$ e as demais bandas empregam $FD_{0,1,2,5,6,7} = 2$.

Apresentamos nas Tabelas 7 e 8 os resultados médios finais das métricas SIR, SAR e SDR (em dB), para diferentes tempos de reverberação e de comprimentos de filtros de separação (S_k) tanto em banda cheia quanto em sub-banda. Nas aludidas tabelas também constam os resultados sem o uso de detector (SD) e com o emprego de detector (CD).

Observando as Tabelas 7 e 8, verifica-se que a elevação dos FD_k promove queda nas métricas das estruturas em sub-bandas em relação aos casos sem decimação ($FD_k = 1$), mas mantendo ainda um desempenho superior aos resultados em banda cheia. Uma opção interessante, é usar diferentes FD_k nas diferentes sub-bandas, de modo a prover um bom compromisso entre complexidade e desempenho, tal sugestão é investigada nesta proposta, na qual denotamos por casos mesclados, os algoritmos que realizam a separação cega em

sub-banda em taxas distintas.

Tabela 7 - Resultados médios finais da SIR, SAR e SDR, para diferentes valores de FD e para $S_k = 256/FD_k$ e $768/FD_k$ para a Mistura 1.

M	FD	T_{60}	S	SIR		SAR		SDR	
				SD	CD	SD	CD	SD	CD
1	1	25ms	256	20,00	20,00	74,60	74,60	20,00	20,00
		100ms	768	18,30	18,30	20,97	20,97	16,40	16,40
2	1	25ms	256	23,80	23,80	70,72	70,72	23,80	23,80
		100ms	768	20,12	20,12	20,79	20,79	17,40	17,40
2	2	25ms	256	21,46	21,46	34,96	34,96	20,01	20,01
		100ms	768	18,34	18,34	20,31	20,31	16,16	16,16
2	M2	25ms	256	22,12	22,12	24,85	24,85	19,33	19,33
		100ms	768	18,97	18,97	18,29	18,29	15,44	15,44
4	1	25ms	256	28,33	28,33	73,87	73,87	28,33	28,33
		100ms	768	22,37	22,37	20,53	20,53	18,31	18,31
4	2	25ms	256	24,59	24,59	31,38	31,38	23,70	23,70
		100ms	768	20,17	20,17	20,47	20,47	17,28	17,28
4	M4	25ms	256	26,73	26,73	72,78	72,78	26,73	26,73
		100ms	768	21,92	21,92	20,61	20,61	18,15	18,15
8	1	25ms	256	33,07	33,07	74,92	74,92	33,07	33,07
		100ms	768	23,70	23,70	21,21	21,21	19,26	19,26
8	2	25ms	256	28,68	28,68	35,23	35,23	27,56	27,56
		100ms	768	21,31	21,31	21,28	21,28	18,21	18,21
8	4	25ms	256	19,44	19,44	27,37	27,37	18,60	18,60
		100ms	768	16,43	16,43	19,80	19,80	14,72	14,72
8	M8	25ms	256	31,58	31,58	74,47	74,49	31,58	31,58
		100ms	768	23,35	23,35	21,30	21,30	19,18	19,18

Tabela 8 - Resultados médios finais da SIR, SAR e SDR, para diferentes valores de FD e para $S_k = 256/FD_k$ e $768/FD_k$ para a Mistura 2.

M	FD	T_{60}	S	SIR		SAR		SDR	
				SD	CD	SD	CD	SD	CD
1	1	25ms	256	19,16	19,06	45,41	45,42	19,14	19,05
		100ms	768	16,95	16,95	21,38	21,38	15,57	15,57
2	1	25ms	256	25,98	25,99	51,24	50,89	25,87	25,87
		100ms	768	19,92	20,23	21,12	21,01	17,38	17,43
2	2	25ms	256	22,22	22,23	31,62	31,62	21,69	21,69
		100ms	768	18,32	18,33	20,72	20,72	16,29	16,29
2	M2	25ms	256	23,18	23,18	24,95	24,95	20,69	20,68
		100ms	768	18,88	18,89	19,40	19,39	16,04	16,04
4	1	25ms	256	29,26	29,27	53,40	53,41	29,05	29,05
		100ms	768	22,90	23,64	20,45	20,32	18,21	18,18
4	2	25ms	256	24,81	24,83	33,60	33,53	24,10	24,11
		100ms	768	20,09	20,46	20,33	20,29	17,05	17,17
4	M4	25ms	256	28,57	28,59	53,28	53,35	28,39	28,40
		100ms	768	22,50	23,16	20,27	20,12	17,99	17,97
8	1	25ms	256	32,67	32,67	55,47	55,44	32,10	32,08
		100ms	768	24,88	25,58	21,24	21,15	19,07	18,97
8	2	25ms	256	27,15	27,18	36,36	36,32	26,37	26,37
		100ms	768	22,29	22,71	21,07	21,07	18,16	18,19
8	4	25ms	256	14,87	14,87	24,11	24,10	14,35	14,35
		100ms	768	12,48	12,48	18,54	18,55	11,47	11,47
8	M8	25ms	256	31,57	31,59	55,44	55,42	31,17	31,18
		100ms	768	24,61	24,99	21,15	21,09	18,94	18,88

A Figura 35 apresenta a evolução da SIR média para $M = 1, 2$ e 4 , $FD_k = \lceil M/2 \rceil$ para tempo de reverberação de $25ms$ e $S_k = 256/FD_k$ para as misturas sem trechos de silêncio (Mistura 1) e com trechos de silêncio (Mistura 2), sem uso do detector (SD) e com uso detector (CD).

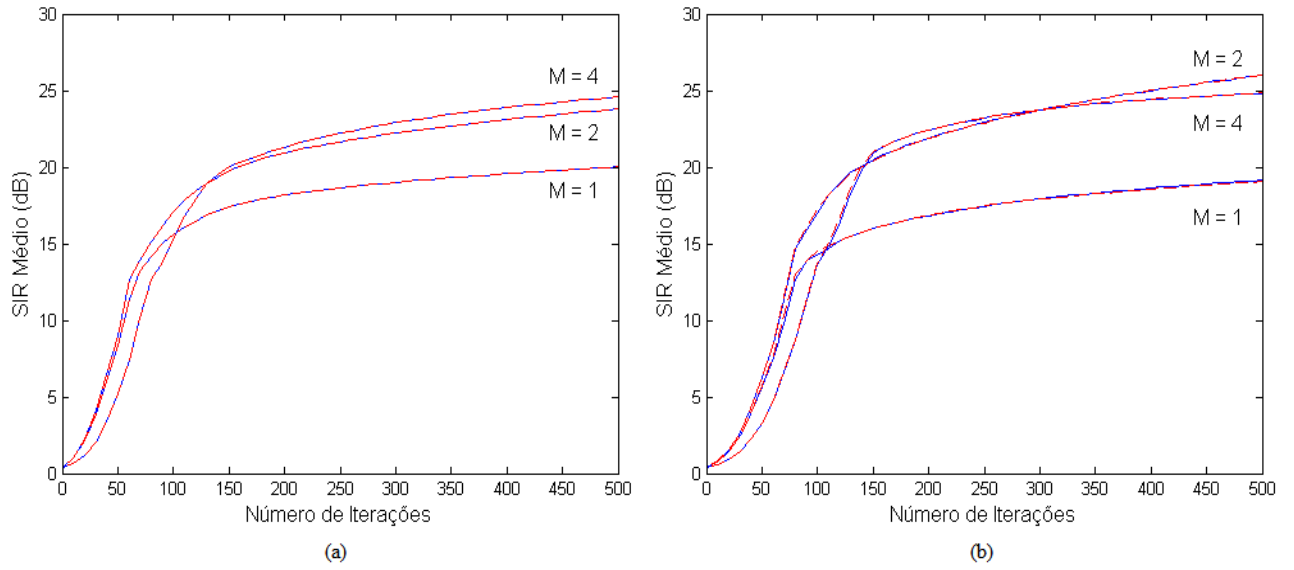


Figura 35 - Evolução da SIR média para $M = 1, 2$ e 4 , $FD_k = \lceil M/2 \rceil$ para $T_{60} = 25ms$ e $S_k = 256/FD_k$, SD em azul e CD em vermelho : (a) Mistura 1 e (b) Mistura 2.

A Figura 36 contempla a evolução da SIR média para $M = 1, 2$ e 4 , $FD_k = \lceil M/2 \rceil$ para tempo de reverberação de $100ms$ e $S_k = 768/FD_k$ para as Misturas 1 e 2, sem emprego do detector (SD) e com uso detector (CD).

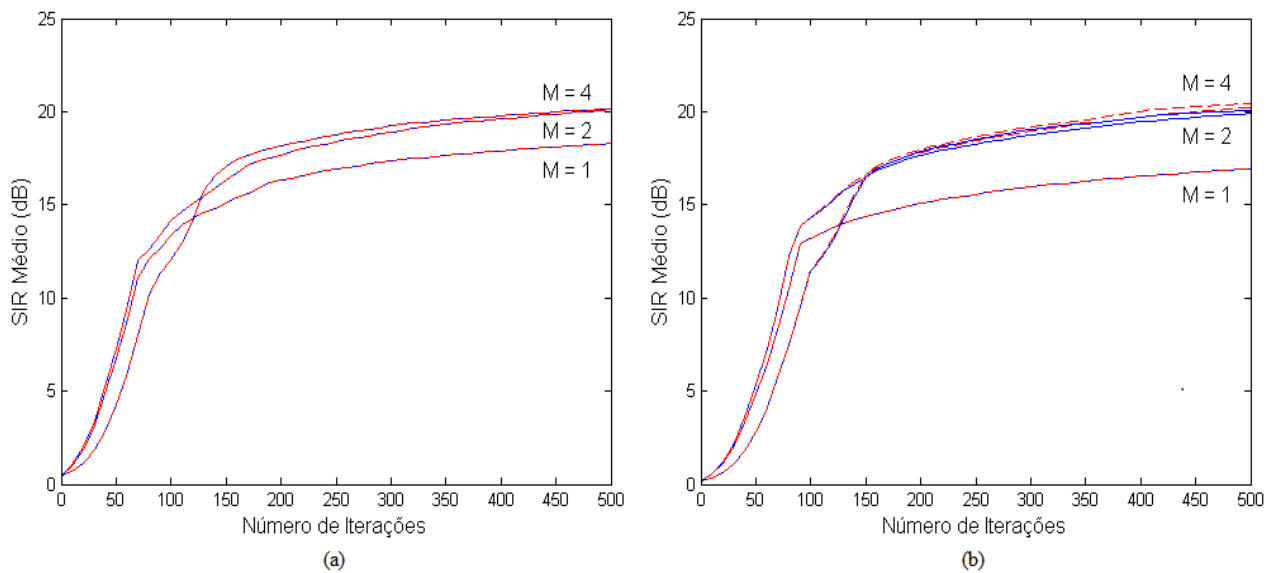


Figura 36 - Evolução da SIR média para $M = 1, 2$ e 4 , $FD_k = \lceil M/2 \rceil$ para $T_{60} = 100ms$ e $S_k = 768/FD_k$, SD em azul e CD em vermelho : (a) Mistura 1 e (b) Mistura 2.

Tabela 10 - Número de blocos empregados no algoritmo de separação, para diferentes valores de M e FD para a Mistura 2.

M	FD	S	S	Sem Detector	Com Detector							
				B_0	B_0							
1	1	256	256	621	573							
		768	768	204	204							
M	FD	T_{60}	S	Sem Detector	Com Detector							
				$B_{0,1}$	B_0	B_1						
2	1	25ms	256	621	616	501						
		100ms	768	204	204	172						
2	2	25ms	256	621	621	593						
		100ms	768	204	204	199						
2	$M2$	25ms	256	621	616	593						
		100ms	768	204	204	199						
M	FD	T_{60}	S	Sem Detector	Com Detector							
				$B_{0,1,2,3}$	B_0	B_1	B_2	B_3				
4	1	25ms	256	621	612	528	445	419				
		100ms	768	204	204	180	155	146				
4	2	25ms	256	621	621	596	490	461				
		100ms	768	204	204	198	168	161				
4	$M4$	25ms	256	621	621	528	445	461				
		100ms	768	204	204	180	155	161				
M	FD	T_{60}	S	Sem Detector	Com Detector							
				$B_{0,1,2,3,4,5,6,7}$	B_0	B_1	B_2	B_3	B_4	B_5	B_6	B_7
8	1	25ms	256	621	621	620	606	621	621	621	618	621
		100ms	768	204	204	204	201	204	204	204	204	204
8	2	25ms	256	621	621	620	606	621	621	621	618	621
		100ms	768	204	204	204	201	204	204	204	204	204
8	4	25ms	256	621	621	620	606	621	621	621	618	621
		100ms	768	204	204	204	201	204	204	204	204	204
8	$M8$	T_{60}	256	621	621	621	550	491	489	559	568	549
		25ms	768	204	204	204	185	168	169	190	192	187

As Tabelas 11 e 12 mostram o número de multiplicações por iteração tanto em implementação normal (ver Equação (89)) quanto na implementação eficiente (ver Equação (90)), para as simulações realizadas neste experimento tanto em banda cheia quanto em sub-banda, nas quais podemos observar que o emprego do detector, em misturas com presença de intervalos de silêncio (Mistura 2), promove importante redução do número de blocos utilizado pelo algoritmo de separação cega a cada iteração, principalmente nas bandas de frequência mais altas, acarretando em redução do custo computacional envolvido. Observa-se também nas tabelas supracitadas, que a implementação da estrutura em sub-banda, resulta em elevação da complexidade computacional em até M vezes, para os casos não maximamente decimados ($M \neq FD$), em relação à implementação em banda cheia, todavia almejando-se não somente

redução do custo computacional envolvido, mas também elevar o desempenho obtido, as versões mescladas, apresentam-se como soluções bastante atrativas (ver Tabelas 7 e 8).

Tabela 11 - Número de multiplicações por iteração em implementação comum e em implementação eficiente, para diferentes valores de M e FD para a Mistura 1.

M	FD	T_{60}	S	Sem Detector		Com Detector	
				NMPI	$NMPI_{IE}$	NMPI	$NMPI_{IE}$
1	1	25ms	256	$8,36 \times 10^{10}$	$3,35 \times 10^7$	$8,35 \times 10^{10}$	$3,35 \times 10^7$
		100ms	768	$7,40 \times 10^{11}$	$3,80 \times 10^7$	$7,40 \times 10^{11}$	$3,80 \times 10^7$
2	1	25ms	256	$1,67 \times 10^{11}$	$6,71 \times 10^7$	$1,67 \times 10^{11}$	$6,71 \times 10^7$
		100ms	768	$1,48 \times 10^{12}$	$7,60 \times 10^7$	$1,48 \times 10^{12}$	$7,60 \times 10^7$
2	2	25ms	256	$8,36 \times 10^{10}$	$3,35 \times 10^7$	$8,36 \times 10^{10}$	$3,35 \times 10^7$
		100ms	768	$7,40 \times 10^{11}$	$3,80 \times 10^7$	$7,40 \times 10^{11}$	$3,80 \times 10^7$
2	M2	25ms	256	$8,88 \times 10^{10}$	$4,11 \times 10^7$	$8,88 \times 10^{10}$	$4,11 \times 10^7$
		100ms	768	$7,86 \times 10^{11}$	$4,67 \times 10^7$	$7,86 \times 10^{11}$	$4,67 \times 10^7$
4	1	25ms	256	$3,34 \times 10^{11}$	$1,34 \times 10^8$	$3,32 \times 10^{11}$	$1,33 \times 10^8$
		100ms	768	$2,96 \times 10^{12}$	$1,52 \times 10^8$	$2,95 \times 10^{12}$	$1,52 \times 10^8$
4	2	25ms	256	$1,67 \times 10^{11}$	$6,71 \times 10^7$	$1,67 \times 10^{11}$	$6,70 \times 10^7$
		100ms	768	$1,48 \times 10^{12}$	$7,60 \times 10^7$	$1,48 \times 10^{12}$	$7,60 \times 10^7$
4	M4	25ms	256	$1,78 \times 10^{11}$	$8,23 \times 10^7$	$1,78 \times 10^{11}$	$8,22 \times 10^7$
		100ms	768	$1,57 \times 10^{12}$	$9,35 \times 10^7$	$1,57 \times 10^{12}$	$9,35 \times 10^7$
8	1	25ms	256	$6,69 \times 10^{11}$	$2,68 \times 10^8$	$6,64 \times 10^{11}$	$2,66 \times 10^8$
		100ms	768	$5,92 \times 10^{12}$	$3,04 \times 10^8$	$5,89 \times 10^{12}$	$3,03 \times 10^8$
8	2	25ms	256	$3,34 \times 10^{11}$	$1,34 \times 10^8$	$3,34 \times 10^{11}$	$1,34 \times 10^8$
		100ms	768	$2,96 \times 10^{12}$	$1,52 \times 10^8$	$2,96 \times 10^{12}$	$1,52 \times 10^8$
8	4	25ms	256	$1,67 \times 10^{11}$	$6,71 \times 10^7$	$1,67 \times 10^{11}$	$6,71 \times 10^7$
		100ms	768	$1,48 \times 10^{12}$	$7,60 \times 10^7$	$1,48 \times 10^{12}$	$7,58 \times 10^7$
8	M8	25ms	256	$1,99 \times 10^{11}$	$1,13 \times 10^8$	$1,99 \times 10^{11}$	$1,13 \times 10^8$
		100ms	768	$1,76 \times 10^{12}$	$1,28 \times 10^8$	$1,76 \times 10^{12}$	$1,28 \times 10^8$

Mediante os resultados apresentados, constatou-se que o fato de elevar o T_{60} do ambiente, não somente reduziu o número de blocos desprovidos de fala, mas também, para prover melhores resultados para o algoritmo de separação, foi necessário aumentar também o comprimento do filtro de separação (S_k).

Embora apresentamos apenas os resultados para $S_k = 256/FD_k$ e $S_k = 768/FD_k$, é importante mencionar que esses foram os melhores comprimentos dos filtros de separação encontrados, dentre os submetidos neste trabalho, que apresentaram resultados satisfatórios para respectivamente $T_{60} = 25ms$ e $T_{60} = 100ms$.

Tabela 12 - Número de multiplicações por iteração em implementação comum e em implementação eficiente, para diferentes valores de M e FD para a Mistura 2.

M	FD	T_{60}	S	Sem Detector		Com Detector	
				NMPI	$NMPI_{IE}$	NMPI	$NMPI_{IE}$
1	1	25ms	256	$8,36 \times 10^{10}$	$3,35 \times 10^7$	$7,71 \times 10^{10}$	$3,10 \times 10^7$
		100ms	768	$7,40 \times 10^{11}$	$3,80 \times 10^7$	$7,15 \times 10^{11}$	$3,67 \times 10^7$
2	1	25ms	256	$1,67 \times 10^{11}$	$6,71 \times 10^7$	$1,50 \times 10^{11}$	$6,03 \times 10^7$
		100ms	768	$1,48 \times 10^{12}$	$7,60 \times 10^7$	$1,36 \times 10^{12}$	$7,01 \times 10^7$
2	2	25ms	256	$8,36 \times 10^{10}$	$3,35 \times 10^7$	$8,17 \times 10^{10}$	$3,28 \times 10^7$
		100ms	768	$7,40 \times 10^{11}$	$3,80 \times 10^7$	$7,31 \times 10^{11}$	$3,76 \times 10^7$
2	M2	25ms	256	$8,88 \times 10^{10}$	$4,11 \times 10^7$	$8,79 \times 10^{10}$	$4,05 \times 10^7$
		100ms	768	$7,86 \times 10^{11}$	$4,67 \times 10^7$	$7,85 \times 10^{11}$	$4,65 \times 10^7$
4	1	25ms	256	$3,34 \times 10^{11}$	$1,34 \times 10^8$	$2,70 \times 10^{11}$	$1,08 \times 10^8$
		100ms	768	$2,96 \times 10^{12}$	$1,52 \times 10^8$	$2,48 \times 10^{12}$	$1,28 \times 10^8$
4	2	25ms	256	$1,67 \times 10^{11}$	$6,71 \times 10^7$	$1,46 \times 10^{11}$	$5,86 \times 10^7$
		100ms	768	$1,48 \times 10^{12}$	$7,60 \times 10^7$	$1,33 \times 10^{12}$	$6,81 \times 10^7$
4	M4	25ms	256	$1,78 \times 10^{11}$	$8,23 \times 10^7$	$1,40 \times 10^{11}$	$6,58 \times 10^7$
		100ms	768	$1,57 \times 10^{12}$	$9,35 \times 10^7$	$1,30 \times 10^{12}$	$7,80 \times 10^7$
8	1	25ms	256	$6,69 \times 10^{11}$	$2,68 \times 10^8$	$5,56 \times 10^{11}$	$2,23 \times 10^8$
		100ms	768	$5,92 \times 10^{12}$	$3,04 \times 10^8$	$5,11 \times 10^{12}$	$2,63 \times 10^8$
8	2	25ms	256	$3,34 \times 10^{11}$	$1,34 \times 10^8$	$3,12 \times 10^{11}$	$1,25 \times 10^8$
		100ms	768	$2,96 \times 10^{12}$	$1,52 \times 10^8$	$2,81 \times 10^{12}$	$1,44 \times 10^8$
8	4	25ms	256	$1,67 \times 10^{11}$	$6,71 \times 10^7$	$1,67 \times 10^{11}$	$6,68 \times 10^7$
		100ms	768	$1,48 \times 10^{12}$	$7,60 \times 10^7$	$1,48 \times 10^{12}$	$7,59 \times 10^7$
8	M8	25ms	256	$1,99 \times 10^{11}$	$1,13 \times 10^8$	$1,61 \times 10^{11}$	$9,53 \times 10^7$
		100ms	768	$1,76 \times 10^{12}$	$1,28 \times 10^8$	$1,49 \times 10^{12}$	$1,12 \times 10^8$

5.3.3 Experimento 2

Neste experimento comparamos o desempenho e a complexidade computacional da separação cega de sinais para as misturas obtidas conforme mencionamos na Seção 5.3.1 tanto em banda cheia quanto em sub-banda, empregando a estrutura não uniforme apresentada na Seção 4.4, cuja configuração é ilustrada na Figura 27. Contemplamos diferentes condições de reverberação ($T_{60} = 25ms$ e $T_{60} = 100ms$) e diferentes comprimentos de filtros FIR do sistema de separação, ($S_k = 256/FD_k$ e $768/FD_k$).

Analogamente a estrutura uniforme também contemplamos nesse experimento versões mescladas, todavia para a estrutura não uniforme escolhemos para $M = 3$ (M3) um $FD_{0e1} = 1$ para as duas primeiras bandas, ou seja, B_0 e B_1 , e a terceira banda (B_2), implementamos com um $FD_2 = 2$; para $M = 4$ (M4) a segunda banda (B_1) adota $FD_1 = 2$ as demais empregam $FD_{0,2e3} = 1$; e por fim, para $M = 5$ (M5) a terceira banda (B_2) foi implementada $FD_2 = 1$, a quarta banda, denotada (B_3), foi realizada com $FD_3 = 4$ e nas demais

inserirmos $FD_{0,1e4} = 2$, tal escolha, teve como fundamento os resultados experimentais obtidos, almejando sempre reduzir a complexidade computacional do algoritmo sem prejudicar consideravelmente o desempenho do mesmo.

Apresentamos nas Tabelas 13 e 14 os resultados médios finais das métricas SIR, SAR e SDR (em dB), para $T_{60} = 25ms$ empregando $S_k = 256/FD_k$ e $T_{60} = 100ms$ adotando $S_k = 768/FD_k$ tanto em banda cheia quanto em sub-banda adotando a estrutura não uniforme, para as Misturas 1 e 2. Nas tabelas supracitadas também constam os resultados sem o uso de detector (SD) e com o emprego de detector (CD).

Tabela 13 - Resultados médios finais da SIR, SAR e SDR, para diferentes valores de FD e para $S_k = 256/FD_k$ para a Mistura 1 em estrutura não uniforme.

M	DIV	SIR		SAR		SDR		FD				
		SD	CD	SD	CD	SD	CD	B_0	B_1	B_2	B_3	B_4
3	1	16,09	16,09	25,23	25,23	15,48	15,48	4	4	2	—	—
	2	25,52	25,52	24,70	24,70	21,47	21,47	2	2	1	—	—
	4	28,30	28,30	74,15	74,15	28,30	28,30	1	1	1	—	—
	$M3$	27,42	27,42	18,24	18,24	17,55	17,55	1	1	2	—	—
4	1	11,55	11,55	20,00	20,00	10,74	10,74	8	8	4	2	—
	2	22,53	22,53	19,92	19,92	17,85	17,85	4	4	2	1	—
	4	31,70	31,70	68,59	68,59	31,70	31,70	2	2	1	1	—
	8	32,93	32,93	75,07	75,07	32,93	32,93	1	1	1	1	—
	$M4$	30,67	30,67	15,59	15,59	15,04	15,04	1	2	1	1	—
5	1	5,77	5,77	13,27	13,27	4,88	4,88	16	16	8	4	2
	2	5,76	5,76	14,53	14,53	5,08	5,08	8	8	4	2	1
	4	28,58	28,58	58,24	58,24	28,57	28,57	4	4	2	1	1
	8	26,85	26,85	58,97	58,97	26,84	26,84	2	2	1	1	1
	16	28,64	28,64	58,89	58,89	28,63	28,63	1	1	1	1	1
	$M5$	24,84	24,84	9,21	9,21	8,98	8,98	2	2	1	4	2

Tabela 14 - Resultados médios finais da SIR, SAR e SDR, para diferentes valores de FD e para $S_k = 256/FD_k$ para a Mistura 2 em estrutura não uniforme.

M	DIV	SIR		SAR		SDR		FD				
		SD	CD	SD	CD	SD	CD	B_0	B_1	B_2	B_3	B_4
3	1	12,71	12,71	19,37	19,37	11,80	11,80	4	4	2	–	–
	2	26,06	26,09	25,95	25,94	22,78	22,77	2	2	1	–	–
	4	28,20	28,26	51,27	50,90	28,00	28,04	1	1	1	–	–
	$M3$	30,39	30,45	16,07	16,04	15,80	15,78	1	1	2	–	–
4	1	9,03	9,03	16,56	16,56	8,11	8,11	8	8	4	2	–
	2	18,36	18,34	16,35	16,35	14,12	14,12	4	4	2	1	–
	4	29,02	29,09	51,06	50,71	28,76	28,80	2	2	1	1	–
	8	29,59	29,69	51,27	50,90	29,28	29,34	1	1	1	1	–
	$M4$	30,44	30,35	10,06	10,16	9,99	10,08	1	2	1	1	–
5	1	4,69	4,69	16,54	16,54	4,30	4,30	16	16	8	4	2
	2	4,97	4,97	11,67	11,67	3,89	3,90	8	8	4	2	1
	4	29,77	29,86	49,41	49,15	29,45	29,51	4	4	2	1	1
	8	29,97	30,07	50,03	49,69	29,63	29,70	2	2	1	1	1
	16	30,02	30,12	49,95	49,57	29,68	29,73	1	1	1	1	1
	$M5$	26,34	26,34	3,68	3,68	3,64	3,64	2	2	1	4	2

Cabe ressaltar que na estrutura não uniforme introduzimos um novo parâmetro, denotado por DIV , para indicar a taxa de redução dos FD_k em cada sub-banda em relação à configuração em árvore maximamente decimada, a qual atribuímos $DIV = 1$. Para facilitar a compreensão, expomos nas últimas colunas das Tabelas 13 e 14 (e nas demais tabelas embasadas na estrutura não uniforme), os FD_k empregados em cada banda nos testes realizados).

Nas tabelas referente à estrutura não uniforme, omitimos os casos para $M = 2$, devido ao fato da estrutura não uniforme para $M = 2$, corresponder à mesma configuração para a estrutura uniforme adotando $M = 2$ (ver Seção 4.4), e por conseguinte apresenta os mesmos resultados, que podem ser observados nas tabelas supracitadas no experimento 1.

Contemplamos nas Tabelas 15 e 16 os resultados médios finais das métricas SIR, SAR e SDR (em dB), para $T_{60} = 100ms$ e $S_k = 768/FD_k$ tanto em banda cheia quanto em sub-banda adotando a estrutura não uniforme, respectivamente para as Misturas 1 e 2 com e sem o emprego de detector de voz.

Tabela 15 - Resultados médios finais da SIR, SAR e SDR, para diferentes valores de FD e para $S_k = 768/FD_k$ para a Mistura 1 em estrutura não uniforme.

M	DIV	SIR		SAR		SDR		FD				
		SD	CD	SD	CD	SD	CD	B_0	B_1	B_2	B_3	B_4
3	1	14,50	14,50	18,58	18,58	13,02	13,02	4	4	2	—	—
	2	20,88	20,88	18,00	18,00	16,14	16,14	2	2	1	—	—
	4	22,21	22,21	20,63	20,63	18,31	18,31	1	1	1	—	—
	$M3$	22,89	22,89	15,00	15,00	14,23	14,23	1	1	2	—	—
4	1	10,17	10,17	15,52	15,52	8,92	8,92	8	8	4	2	—
	2	18,83	18,83	16,46	16,46	14,37	14,37	4	4	2	1	—
	4	22,53	21,35	21,35	21,35	18,87	18,87	2	2	1	1	—
	8	22,78	22,78	21,28	21,28	18,93	18,93	1	1	1	1	—
	$M4$	25,73	25,73	12,67	12,67	12,33	12,33	1	2	1	1	—
5	1	4,94	4,94	13,29	13,29	4,11	4,11	16	16	8	4	2
	2	4,76	4,76	12,60	12,60	3,89	3,89	8	8	4	2	1
	4	21,57	21,57	23,13	23,13	18,91	18,91	4	4	2	1	1
	8	21,82	21,82	23,13	23,13	19,11	19,11	2	2	1	1	1
	16	22,30	22,30	22,99	22,99	19,43	19,43	1	1	1	1	1
	$M5$	23,03	23,03	7,18	7,18	6,97	6,97	2	2	1	4	2

Os resultados apresentados nas Tabelas 13, 14, 15 e 16 indicam que as versões mescladas para a estrutura não uniforme ($M3$, $M4$ e $M5$) embora os resultados médios finais da SIR sejam satisfatórios, as demais métricas provavelmente devido ao *aliasing* não apresentam bons resultados.

Tabela 16 - Resultados médios finais da SIR, SAR e SDR, para diferentes valores de FD e para $S_k = 768/FD_k$ para a Mistura 2 em estrutura não uniforme.

M	DIV	SIR		SAR		SDR		FD				
		SD	CD	SD	CD	SD	CD	B_0	B_1	B_2	B_3	B_4
3	1	12,14	12,14	16,69	16,68	10,77	10,77	4	4	2	—	—
	2	20,30	20,72	18,96	18,92	16,29	16,34	2	2	1	—	—
	4	21,64	22,21	20,53	20,73	17,73	17,84	1	1	1	—	—
	$M3$	22,93	22,93	14,40	14,51	13,59	13,67	1	1	2	—	—
4	1	8,21	8,21	14,20	14,19	7,00	7,00	8	8	4	2	—
	2	14,24	14,21	14,63	14,62	11,28	11,27	4	4	2	1	—
	4	22,16	22,87	21,49	21,47	18,26	18,30	2	2	1	1	—
	8	22,30	23,05	21,69	21,65	18,38	18,42	1	1	1	1	—
	$M4$	26,00	26,14	9,78	9,74	9,58	9,55	1	2	1	1	—
5	1	4,53	4,53	13,76	13,76	3,87	3,87	16	16	8	4	2
	2	4,57	4,59	10,51	10,52	3,29	3,31	8	8	4	2	1
	4	22,45	23,18	21,32	21,29	18,28	18,32	4	4	2	1	1
	8	22,55	23,31	21,41	21,38	18,32	18,36	2	2	1	1	1
	16	22,53	23,30	21,35	21,37	18,28	18,34	1	1	1	1	1
	$M5$	23,50	23,50	2,44	2,43	2,38	2,38	2	2	1	4	2

As Figura 37 e 38 apresentam a evolução da SIR média para para $M = 1, 3, 4$ e 5 , $DIV = \lceil DIV_{max}/2 \rceil$ respectivamente para $T_{60} = 25ms$ adotando $S_k = 256/FD_k$ e $T_{60} = 100ms$ utilizando $S_k = 768/FD_k$, para as Misturas 1 e 2, sem uso do detector (SD) e com uso detector (CD).

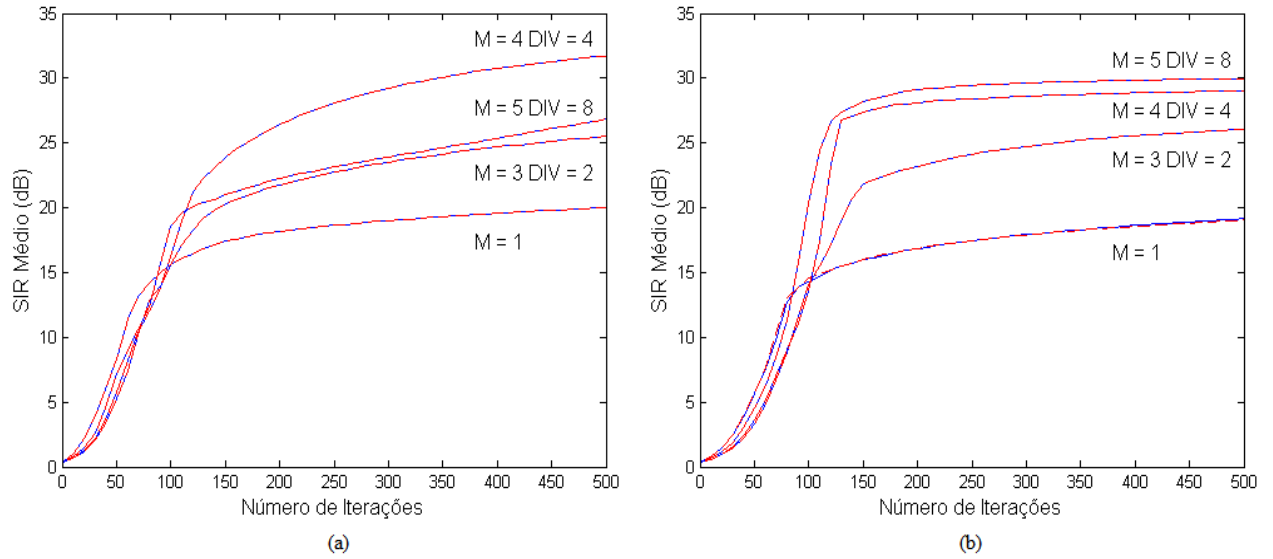


Figura 37 - Evolução da SIR média para $M = 1, 3, 4$ e 5 , $DIV = \lceil DIV_{max}/2 \rceil$ para $T_{60} = 25ms$ e $S_k = 256/FD_k$, SD em azul e CD em vermelho : (a) Mistura 1 e (b) Mistura 2.

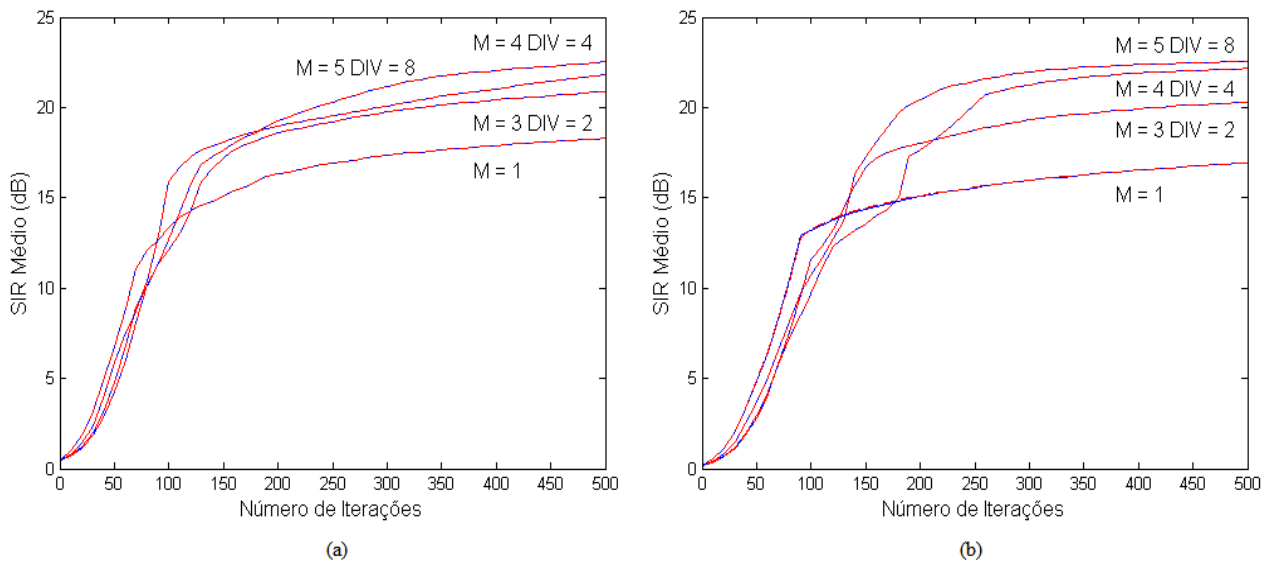


Figura 38 - Evolução da SIR média para $M = 1, 3, 4$ e 5 , $DIV = \lceil DIV_{max}/2 \rceil$ para $T_{60} = 100ms$ e $S_k = 768/FD_k$, SD em azul e CD em vermelho : (a) Mistura 1 e (b) Mistura 2.

Apresentamos nas Tabelas 17 e 18 o número de blocos utilizados pelo algoritmo de separação a cada iteração neste experimento para as Mistura 1 e 2 respectivamente para a estrutura não uniforme.

Tabela 17 - Número de blocos empregados no algoritmo de separação, para diferentes valores de M e FD para a Mistura 1.

M	DIV	T_{60}	S	Sem Detector	Com Detector					
				$B_{0,1,2}$	B_0	B_1	B_2			
3	1	25ms	256	621	621	621	621			
		100ms	768	204	204	204	204			
	2	25ms	256	621	621	621	621			
		100ms	768	204	204	204	204			
	4	25ms	256	621	621	621	621			
		100ms	768	204	204	204	204			
	$M3$	25ms	256	621	621	621	621			
		100ms	768	204	204	204	204			
M	DIV	T_{60}	S	Sem Detector	Com Detector					
				$B_{0,1,2}$	B_0	B_1	B_2	B_3		
4	1	25ms	256	621	621	621	621	621		
		100ms	768	204	204	204	204	204		
	2	25ms	256	621	621	621	621	621		
		100ms	768	204	204	204	204	204		
	4	25ms	256	621	621	621	621	621		
		100ms	768	204	204	204	204	204		
	8	25ms	256	621	621	621	621	621		
		100ms	768	204	204	204	204	204		
	$M4$	25ms	256	621	621	621	621	621		
		100ms	768	204	204	204	204	204		
	M	DIV	T_{60}	S	Sem Detector	Com Detector				
					$B_{0,1,2}$	B_0	B_1	B_2	B_3	B_4
5	1	25ms	256	621	621	621	621	621	621	
		100ms	768	204	204	204	204	204	204	
	2	25ms	256	621	621	621	621	621	621	
		100ms	768	204	204	204	204	204	204	
	4	25ms	256	621	621	621	621	621	621	
		100ms	768	204	204	204	204	204	204	
	8	25ms	256	621	621	621	621	621	621	
		100ms	768	204	204	204	204	204	204	
	16	25ms	256	621	621	621	621	621	621	
		100ms	768	204	204	204	204	204	204	
	$M5$	25ms	256	621	621	621	621	621	621	
		100ms	768	204	204	204	204	204	204	

Analogamente à abordagem em estrutura uniforme, podemos observar através das Figuras 37 e 38, que os sistemas em sub-bandas em estrutura não uniforme também apresentam desempenho significativamente superior à estrutura em banda cheia, exceto para valores muito pequenos de DIV , conforme é possível verificar nas Tabelas 13, 14, 15 e 16, uma vez que o fato de ocorrer $DIV = 1$ implica em uma estrutura maximamente decimada, o que por sua vez, resulta no aumento do *aliasing* entre os canais.

Tabela 18 - Número de blocos empregados no algoritmo de separação, para diferentes valores de M e FD para a Mistura 2.

M	DIV	T_{60}	S	Sem Detector	Com Detector					
				$B_{0,1,2}$	B_0	B_1	B_2			
3	1	25ms	256	621	621	621	593			
		100ms	768	204	204	204	199			
	2	25ms	256	621	621	596	506			
		100ms	768	204	204	198	172			
	4	25ms	256	621	612	519	506			
		100ms	768	204	202	180	172			
	M3	25ms	256	621	612	519	593			
		100ms	768	204	202	180	199			
M	DIV	T_{60}	S	Sem Detector	Com Detector					
				$B_{0,1,2}$	B_0	B_1	B_2	B_3		
4	1	25ms	256	621	621	621	621	593		
		100ms	768	204	204	204	204	199		
	2	25ms	256	621	621	621	596	506		
		100ms	768	204	204	204	198	172		
	4	25ms	256	621	621	621	519	506		
		100ms	768	204	204	204	180	172		
	8	25ms	256	621	621	610	519	506		
		100ms	768	204	204	204	180	172		
	M4	25ms	256	621	621	621	519	506		
		100ms	768	204	204	204	180	172		
	M	DIV	T_{60}	S	Sem Detector	Com Detector				
					$B_{0,1,2}$	B_0	B_1	B_2	B_3	B_4
5	1	25ms	256	621	621	621	621	621	593	
		100ms	768	204	204	204	204	204	199	
	2	25ms	256	621	621	621	621	596	506	
		100ms	768	204	204	204	204	198	172	
	4	25ms	256	621	621	621	621	519	506	
		100ms	768	204	204	204	204	180	172	
	8	25ms	256	621	621	614	610	519	506	
		100ms	768	204	204	204	204	180	172	
	16	25ms	256	621	605	604	610	519	506	
		100ms	768	204	203	200	204	180	172	
	M5	25ms	256	621	621	614	610	621	593	
		100ms	768	204	204	204	204	204	199	

Apresentamos nas Tabelas 19 e 20 os resultados do número de multiplicações por iteração (tanto em implementação normal (Equação (89)) quanto na implementação eficiente (Equação (90)), para as simulações apresentadas indicando também que a etapa de detecção de voz, no caso de BSS em sub-bandas, promove importante redução do número de blocos utilizado pelo algoritmo de separação a cada iteração, principalmente nas bandas de frequência mais altas (assim como ocorreu na estrutura uniforme), mediante a presença de trechos de silêncio nas misturas, resultando em redução da complexidade computacional envolvida.

Tabela 19 - Tabela Complexidade Computacional para a Mistura 1 - $S_k = 256/FD_k$ e $S_k = 768/FD_k$.

M	DIV	S	Sem Detector		Com Detector	
			NMPI	$NMPI_{IE}$	NMPI	$NMPI_{IE}$
3	1	256	$5,90 \times 10^9$	$1,10 \times 10^7$	$5,90 \times 10^9$	$1,10 \times 10^7$
		768	$5,21 \times 10^{10}$	$1,27 \times 10^7$	$5,21 \times 10^{10}$	$1,27 \times 10^7$
3	2	256	$9,41 \times 10^{10}$	$4,87 \times 10^7$	$9,41 \times 10^{10}$	$4,87 \times 10^7$
		768	$8,33 \times 10^{11}$	$5,55 \times 10^7$	$8,33 \times 10^{11}$	$5,55 \times 10^7$
3	4	256	$2,51 \times 10^{11}$	$1,01 \times 10^8$	$2,51 \times 10^{11}$	$1,01 \times 10^8$
		768	$2,22 \times 10^{12}$	$1,14 \times 10^8$	$2,22 \times 10^{12}$	$1,14 \times 10^8$
3	M3	256	$1,72 \times 10^{11}$	$7,47 \times 10^7$	$1,72 \times 10^{11}$	$7,47 \times 10^7$
		768	$1,53 \times 10^{12}$	$8,48 \times 10^7$	$1,53 \times 10^{12}$	$8,48 \times 10^7$
4	1	256	$5,61 \times 10^9$	$1,00 \times 10^7$	$5,61 \times 10^9$	$1,00 \times 10^7$
		768	$4,96 \times 10^{10}$	$1,16 \times 10^7$	$4,96 \times 10^{10}$	$1,16 \times 10^7$
4	2	256	$8,95 \times 10^{10}$	$4,45 \times 10^7$	$8,95 \times 10^{10}$	$4,45 \times 10^7$
		768	$7,92 \times 10^{11}$	$5,07 \times 10^7$	$7,92 \times 10^{11}$	$5,07 \times 10^7$
4	4	256	$1,78 \times 10^{11}$	$8,23 \times 10^7$	$1,78 \times 10^{11}$	$8,23 \times 10^7$
		768	$1,57 \times 10^{12}$	$9,35 \times 10^7$	$1,57 \times 10^{12}$	$9,35 \times 10^7$
4	8	256	$3,34 \times 10^{11}$	$1,34 \times 10^8$	$3,34 \times 10^{11}$	$1,34 \times 10^8$
		768	$2,96 \times 10^{12}$	$1,52 \times 10^8$	$2,96 \times 10^{12}$	$1,52 \times 10^8$
4	M4	256	$2,56 \times 10^{11}$	$1,08 \times 10^8$	$2,56 \times 10^{11}$	$1,08 \times 10^8$
		768	$2,27 \times 10^{12}$	$1,23 \times 10^8$	$2,27 \times 10^{12}$	$1,23 \times 10^8$
5	1	256	$5,59 \times 10^9$	$9,83 \times 10^6$	$5,59 \times 10^9$	$9,83 \times 10^6$
		768	$4,94 \times 10^{10}$	$1,14 \times 10^7$	$4,94 \times 10^{10}$	$1,14 \times 10^7$
5	2	256	$8,92 \times 10^{10}$	$4,36 \times 10^7$	$8,92 \times 10^{10}$	$4,36 \times 10^7$
		768	$7,90 \times 10^{11}$	$4,96 \times 10^7$	$7,90 \times 10^{11}$	$4,96 \times 10^7$
5	4	256	$1,73 \times 10^{11}$	$7,81 \times 10^7$	$1,73 \times 10^{11}$	$7,81 \times 10^7$
		768	$1,53 \times 10^{12}$	$8,87 \times 10^7$	$1,53 \times 10^{12}$	$8,87 \times 10^7$
5	8	256	$2,61 \times 10^{11}$	$1,16 \times 10^8$	$2,61 \times 10^{11}$	$1,16 \times 10^8$
		768	$2,31 \times 10^{12}$	$1,32 \times 10^8$	$2,31 \times 10^{12}$	$1,32 \times 10^8$
5	16	256	$4,18 \times 10^{11}$	$1,68 \times 10^8$	$4,18 \times 10^{11}$	$1,68 \times 10^8$
		768	$3,70 \times 10^{12}$	$1,90 \times 10^8$	$3,70 \times 10^{12}$	$1,90 \times 10^8$
5	M5	256	$9,96 \times 10^{10}$	$5,80 \times 10^7$	$9,96 \times 10^{10}$	$5,80 \times 10^7$
		768	$8,82 \times 10^{11}$	$6,62 \times 10^7$	$8,82 \times 10^{11}$	$6,62 \times 10^7$

Os resultados para a estrutura não uniforme também evidenciam que o emprego de

detectores de voz não compromete o desempenho do algoritmo de BSS, mesmo em casos de misturas sem trechos de silêncio, revelando-se importante, quando lidamos com misturas que apresentam instantes de silêncio, não somente pelo desempenho ligeiramente superior, mas principalmente pela redução da complexidade computacional quando empregamos estruturas em sub-bandas.

Tabela 20 - Tabela Complexidade Computacional para a Mistura 2 - $S_k = 256/FD_k$ e $S_k = 768/FD_k$.

M	DIV	S	Sem Detector		Com Detector	
			NMPI	$NMPI_{IE}$	NMPI	$NMPI_{IE}$
3	1	256	$5,90 \times 10^9$	$1,10 \times 10^7$	$5,66 \times 10^9$	$1,06 \times 10^7$
		768	$5,21 \times 10^{10}$	$1,27 \times 10^7$	$5,10 \times 10^{10}$	$1,25 \times 10^7$
3	2	256	$9,41 \times 10^{10}$	$4,87 \times 10^7$	$7,84 \times 10^{10}$	$4,22 \times 10^7$
		768	$8,33 \times 10^{11}$	$5,55 \times 10^7$	$7,15 \times 10^{11}$	$4,92 \times 10^7$
3	4	256	$2,51 \times 10^{11}$	$1,01 \times 10^8$	$2,20 \times 10^{11}$	$8,84 \times 10^7$
		768	$2,22 \times 10^{12}$	$1,14 \times 10^8$	$2,01 \times 10^{12}$	$1,03 \times 10^8$
3	M3	256	$1,72 \times 10^{11}$	$7,47 \times 10^7$	$1,57 \times 10^{11}$	$6,83 \times 10^7$
		768	$1,53 \times 10^{12}$	$8,48 \times 10^7$	$1,43 \times 10^{12}$	$7,97 \times 10^7$
4	1	256	$5,61 \times 10^9$	$1,00 \times 10^7$	$5,38 \times 10^9$	$9,70 \times 10^6$
		768	$4,96 \times 10^{10}$	$1,16 \times 10^7$	$4,84 \times 10^{10}$	$1,14 \times 10^7$
4	2	256	$8,95 \times 10^{10}$	$4,45 \times 10^7$	$7,38 \times 10^{10}$	$3,80 \times 10^7$
		768	$7,92 \times 10^{11}$	$5,07 \times 10^7$	$6,75 \times 10^{11}$	$4,45 \times 10^7$
4	4	256	$1,78 \times 10^{11}$	$8,23 \times 10^7$	$1,48 \times 10^{11}$	$7,05 \times 10^7$
		768	$1,57 \times 10^{12}$	$9,35 \times 10^7$	$1,37 \times 10^{12}$	$8,31 \times 10^7$
4	8	256	$3,34 \times 10^{11}$	$1,34 \times 10^8$	$3,04 \times 10^{11}$	$1,22 \times 10^8$
		768	$2,96 \times 10^{12}$	$1,52 \times 10^8$	$2,76 \times 10^{12}$	$1,42 \times 10^8$
4	M4	256	$2,56 \times 10^{11}$	$1,08 \times 10^8$	$2,27 \times 10^{11}$	$9,65 \times 10^7$
		768	$2,27 \times 10^{12}$	$1,23 \times 10^8$	$2,06 \times 10^{12}$	$1,12 \times 10^8$
5	1	256	$5,59 \times 10^9$	$9,83 \times 10^6$	$5,36 \times 10^9$	$9,49 \times 10^6$
		768	$4,94 \times 10^{10}$	$1,14 \times 10^7$	$4,83 \times 10^{10}$	$1,11 \times 10^7$
5	2	256	$8,92 \times 10^{10}$	$4,36 \times 10^7$	$7,35 \times 10^{10}$	$3,71 \times 10^7$
		768	$7,90 \times 10^{11}$	$4,96 \times 10^7$	$6,72 \times 10^{11}$	$4,34 \times 10^7$
5	4	256	$1,73 \times 10^{11}$	$7,81 \times 10^7$	$1,44 \times 10^{11}$	$6,64 \times 10^7$
		768	$1,53 \times 10^{12}$	$8,87 \times 10^7$	$1,33 \times 10^{12}$	$7,83 \times 10^7$
5	8	256	$2,61 \times 10^{11}$	$1,16 \times 10^8$	$2,31 \times 10^{11}$	$1,03 \times 10^8$
		768	$2,31 \times 10^{12}$	$1,32 \times 10^8$	$2,11 \times 10^{12}$	$1,21 \times 10^8$
5	16	256	$4,18 \times 10^{11}$	$1,68 \times 10^8$	$3,83 \times 10^{11}$	$1,54 \times 10^8$
		768	$3,70 \times 10^{12}$	$1,90 \times 10^8$	$3,48 \times 10^{12}$	$1,79 \times 10^8$
5	M5	256	$9,96 \times 10^{10}$	$5,80 \times 10^7$	$9,79 \times 10^{10}$	$5,70 \times 10^7$
		768	$8,82 \times 10^{11}$	$6,62 \times 10^7$	$8,81 \times 10^{11}$	$6,60 \times 10^7$

5.3.4 Experimento 3

O experimento a seguir almeja apresentar uma outra abordagem para reduzir a complexidade computacional do algoritmo de separação sem comprometer significativamente o desempenho do mesmo, para as misturas obtidas conforme supracitado na Seção 5.3.1. A estratégia, baseia-se em reduzir o comprimento dos filtros S_k nas versões mescladas por um fator de redução (FR_k), propositalmente, reduzimos à metade o comprimento dos filtros S_k que encontram-se sem decimação ($FD_k = 1$), nas versões mescladas. Destarte para casos mesclados o comprimento de filtro é reduzido de tal forma que:

$$S_k^{Red} = \frac{S_k}{FD_k FR_k}, \quad (94)$$

sendo FR_k o Fator de Redução na k -ésima sub-banda.

Por simplicidade apresentamos nas Tabelas 21 e 22 os resultados tanto para o desempenho, quanto os blocos empregados no algoritmo e a complexidade computacional, respectivamente para as Misturas 1 e 2, apenas para o caso mesclado $M4$ em estrutura uniforme, logo implica em $FD = [2 \ 1 \ 1 \ 2]$.

Tabela 21 - Tabela apresentando a estratégia para reduzir a complexidade computacional para a Mistura 1 - $S = 256$ e 768 .

$M = 4 - FD = [2 \ 1 \ 1 \ 2]$ - Fator de Redução (FR) = $[1 \ 2 \ 2 \ 1]$									
	S	Sem Detector				Com Detector			
		B_0	B_1	B_2	B_3	B_0	B_1	B_2	B_3
$M4$	256	621	1246	1246	621	621	1246	1245	618
	768	204	412	412	204	204	412	412	204
		SIR	NMPI	$NMPI_{IE}$	SIR	NMPI	$NMPI_{IE}$		
	256	26,70	$3,15 \times 10^{10}$	$4,56 \times 10^7$	26,70	$3,15 \times 10^{10}$	$4,56 \times 10^7$		
	768	21,62	$2,80 \times 10^{11}$	$5,27 \times 10^7$	21,62	$2,80 \times 10^{11}$	$5,27 \times 10^7$		

Tabela 22 - Tabela apresentando a estratégia para reduzir a complexidade computacional para a Mistura 2 - $S = 256$ e 768 .

$M = 4 - FD = [2 \ 1 \ 1 \ 2]$ - Fator de Redução (FR) = $[1 \ 2 \ 2 \ 1]$									
	S	Sem Detector				Com Detector			
		B_0	B_1	B_2	B_3	B_0	B_1	B_2	B_3
$M4$	256	621	1246	1246	621	621	1052	881	461
	768	204	412	412	204	204	353	301	161
		SIR	NMPI	$NMPI_{IE}$	SIR	NMPI	$NMPI_{IE}$		
	256	28,46	$3,15 \times 10^{10}$	$4,56 \times 10^7$	28,48	$2,54 \times 10^{10}$	$3,69 \times 10^7$		
	768	21,67	$2,80 \times 10^{11}$	$5,27 \times 10^7$	21,99	$2,31 \times 10^{11}$	$4,36 \times 10^7$		

A Figura 39 apresenta a evolução da SIR média para a implementação em banda cheia $M = 1$ e para a versão mesclada em estrutura uniforme com $FD = M4$, para $T_{60} = 25ms$

e S_k reduzido (conforme apresentado na Equação (94)) para as misturas sem trechos de silêncio (Mistura 1) e com trechos de silêncio (Mistura 2), sem uso do detector (SD) e com uso detector (CD), (devido às similaridades com a Figura 39, omitimos neste trabalho a evolução da SIR média para $T_{60} = 768ms$).

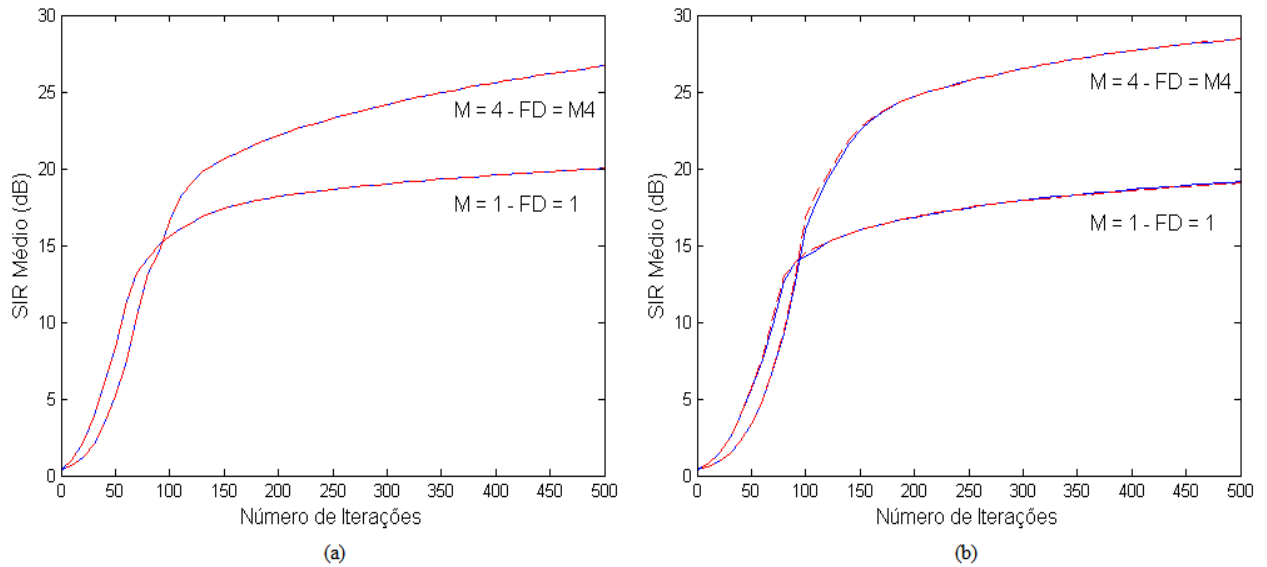


Figura 39 - Evolução da SIR média para a implementação em banda cheia e para a versão mesclada em estrutura uniforme com $FD = M4$, para $T_{60} = 25ms$ e S_k reduzido, SD em azul e CD em vermelho : (a) Mistura 1 e (b) Mistura 2.

Cabe ressaltar que a redução do comprimento dos filtros de separação implica no aumento do número de blocos, todavia mesmo com uma quantidade maior de blocos, conseguimos obter a redução do custo. Isso foi possível uma vez que no cálculo da complexidade computacional o comprimento dos filtros de separação apresenta um influência na complexidade computacional muito mais significativa em relação ao número de blocos (ver Equação (89)) desde que a condição $S_k \gg 1$ seja satisfeita.

Nas Tabelas 21 e 22 em comparação com os resultados obtidos na experiência 1, indicam que é possível obter uma complexidade computacional menor associado à um ganho de desempenho considerável, em relação à implementação em banda cheia através da implementação em estrutura uniforme para as versões mescladas via redução dos comprimentos dos filtros que não apresentam decimação ($FD = 1$).

5.3.5 Experimento 4

As misturas empregadas neste experimento foram obtidas via convolução dos sinais de fala escolhidos na Seção 5.3.1, entretanto com as respostas impulsivas extraídas de um banco de dados denotado por MARDY (*Multichannel Acoustic Reverberation Database at York*) proposto em [65]. As respostas impulsivas presentes neste banco de dados, modelam a sala acústica real, apresentada na Figura 40, considerando um conjunto de oito microfones alinhados, formando um vetor (array), espaçados entre si de 5cm , dentre os quais escolhemos as respostas impulsivas referentes aos microfones localizados nas posições dois e sete deste array, para duas fontes distantes de 1m , uma à esquerda e outra à direita, do eixo do aludido array.

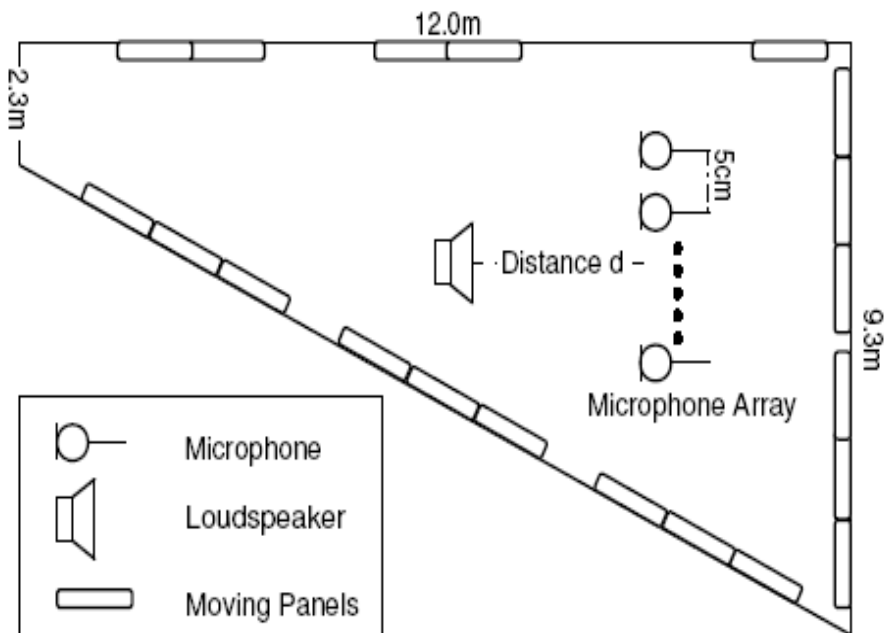


Figura 40 - Sala acústica real [65].

A partir destas novas misturas, apresentadas na Figura 41, que denotamos por Misturas 3 e 4, realizamos a implementação tanto em banda cheia quanto em sub-bandas para a estrutura uniforme investigada no primeiro experimento.

Cabe ressaltar que as Mistura 3 e 4 empregam respectivamente os mesmos sinais de voz que as Mistura 1 e 2, entretanto as Misturas 3 e 4 foram obtidas via emprego do banco de dados MARDY.

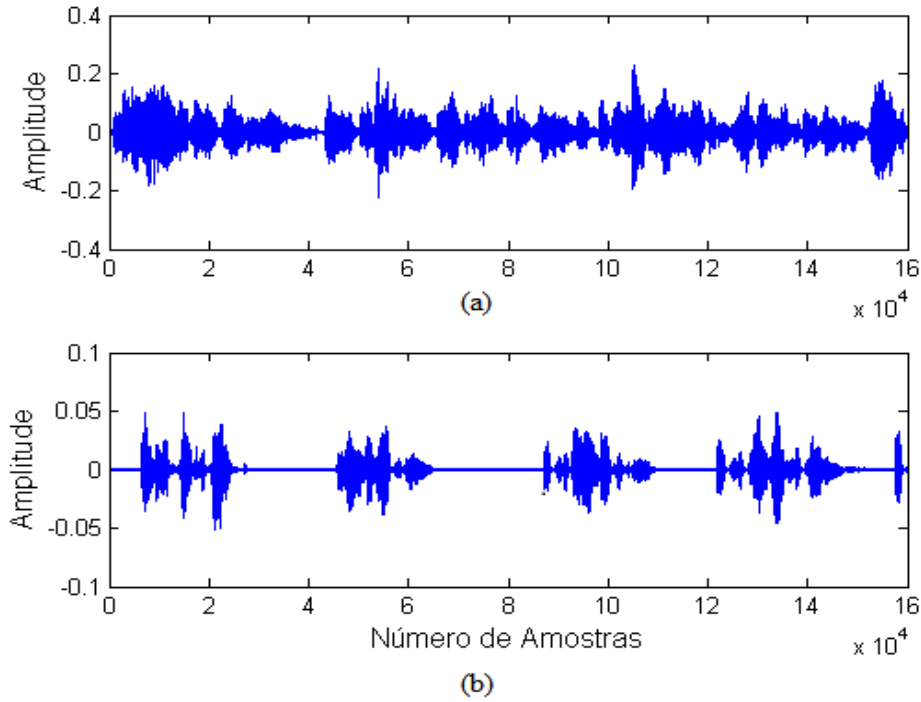


Figura 41 - Mistura obtidas: (a) Mistura 3 e (b) Mistura 4.

Os testes com estas novas misturas são importantes, uma vez que elas simulam um ambiente real conforme proposto em [65], apresentando um tempo de reverberação elevado em relação aos experimentos anteriores, e como consequência foi necessário, para prover bons resultados, empregar $S_k = 2048/FD_k$.

Apresentamos nas Tabelas 23 e 24 os resultados médios finais das métricas SIR, SAR e SDR (em dB), para $S_k = 2048/FD_k$ tanto em banda cheia quanto em sub-banda em estrutura uniforme, respectivamente para as Misturas 3 e 4, sem o uso de detector (SD) e com o emprego de detector (CD).

Tabela 23 - Resultados médios finais da SIR, SAR e SDR, para diferentes valores de FD e para $S_k = 2048/FD_k$ para a Mistura 3.

M	FD	S	SIR		SAR		SDR	
			SD	CD	SD	CD	SD	CD
1	1	2048	19,29	19,29	10,34	10,34	9,67	9,67
2	1	2048	21,36	21,36	10,35	10,35	9,82	9,82
2	2	2048	21,65	21,65	10,30	10,30	9,78	9,78
2	$M2$	2048	21,20	21,20	10,24	10,24	9,70	9,70
4	1	2048	22,87	22,87	10,37	10,37	10,05	10,05
4	2	2048	23,07	23,07	10,33	10,33	10,01	10,01
4	$M4$	2048	23,27	23,27	10,34	10,34	10,03	10,03
8	1	2048	24,60	24,60	10,35	10,35	10,13	10,13
8	2	2048	24,37	24,37	10,32	10,32	10,08	10,08
8	4	2048	23,07	23,07	10,28	10,28	9,96	9,96
8	$M8$	2048	24,54	24,54	10,32	10,32	10,09	10,09

Tabela 24 - Resultados médios finais da SIR, SAR e SDR, para diferentes valores de FD e para $S_k = 2048/FD_k$ para a Mistura 4.

M	FD	S	SIR		SAR		SDR	
			SD	CD	SD	CD	SD	CD
1	1	2048	15,59	15,56	18,17	18,49	13,64	13,72
2	1	2048	18,16	18,22	18,35	18,32	15,20	15,22
2	2	2048	16,92	16,92	18,01	18,01	14,36	14,36
2	$M2$	2048	17,56	17,56	17,30	17,30	14,37	14,37
4	1	2048	20,00	20,13	18,29	18,31	16,02	16,07
4	2	2048	17,94	17,97	17,95	17,96	14,90	14,91
4	$M4$	2048	19,23	19,30	18,13	18,15	15,60	15,65
8	1	2048	21,11	21,13	18,17	18,15	16,36	16,36
8	2	2048	19,34	19,36	18,28	18,27	15,74	15,74
8	4	2048	11,47	11,47	16,97	16,97	10,28	10,28
8	$M8$	2048	20,82	20,86	18,33	18,32	16,36	16,37

Na Figura 42 apresentamos a evolução da SIR média para $M = 1, 2$ e 4 , $FD_k = \lceil M/2 \rceil$ para $S_k = 2048/FD_k$ para as misturas sem trechos de silêncio (Mistura 3) e com trechos de silêncio (Mistura 4), sem uso do detector (SD) e com uso detector (CD).

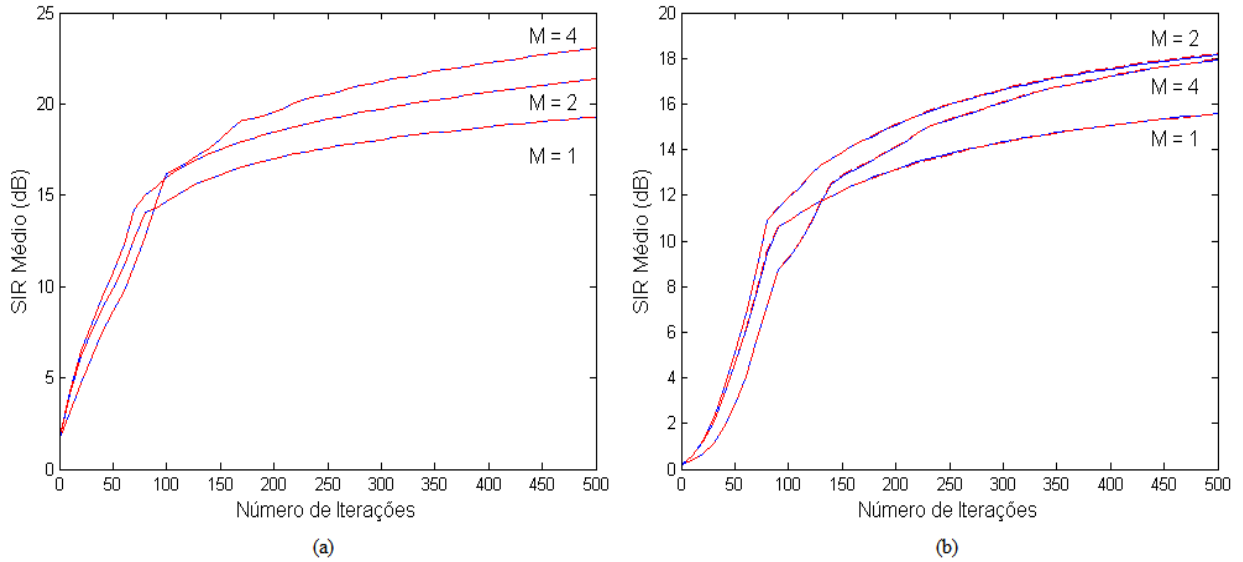


Figura 42 - Evolução da SIR média para $M = 1, 2$ e 4 , $FD_k = \lceil M/2 \rceil$ para $S_k = 2048/FD_k$, SD em azul e CD em vermelho : (a) Mistura 3 e (b) Mistura 4.

Os resultados apresentados na Figura 42 e nas Tabelas 23 e 24, permitem também concluir que a elevação dos FD_k promove queda nas métricas das estruturas em sub-bandas em relação aos casos sem decimação ($FD_k = 1$), mas mantendo ainda um desempenho superior aos resultados em banda cheia, embora não tão expressivo como os resultados obtidos no experimento 1.

Contemplamos nas Tabelas 25 e 26 o número de blocos utilizados pelo algoritmo de separação a cada iteração respectivamente para as Misturas 3 e 4.

Tabela 25 - Número de blocos empregados no algoritmo de separação, para diferentes valores de M e FD para a Mistura 3.

M	FD	S	Sem Detector	Com Detector							
			B_0	B_0							
1	1	2048	74	74							
M	FD	S	Sem Detector	Com Detector							
			$B_{0,1}$	B_0	B_1						
2	1	2048	74	74	74						
2	2	2048	74	74	74						
2	M2	2048	74	74	74						
M	FD	S	Sem Detector	Com Detector							
			$B_{0,1,2,3}$	B_0	B_1	B_2	B_3				
4	1	2048	74	74	74	74	74				
4	2	2048	74	74	74	74	74				
4	M4	2048	74	74	74	74	74				
M	FD	S	Sem Detector	Com Detector							
			$B_{0,1,2,3,4,5,6,7}$	B_0	B_1	B_2	B_3	B_4	B_5	B_6	B_7
8	1	2048	74	74	74	74	74	74	74	74	74
8	2	2048	74	74	74	74	74	74	74	74	74
8	4	2048	74	74	74	74	74	74	74	74	74
8	M8	2048	74	74	74	74	74	74	74	74	74

Tabela 26 - Número de blocos empregados no algoritmo de separação, para diferentes valores de M e FD para a Mistura 4.

M	FD	S	Sem Detector	Com Detector							
			B_0	B_0							
1	1	2048	74	71							
M	FD	S	Sem Detector	Com Detector							
			$B_{0,1}$	B_0	B_1						
2	1	2048	74	74	65						
2	2	2048	74	74	73						
2	M2	2048	74	74	73						
M	FD	S	Sem Detector	Com Detector							
			$B_{0,1,2,3}$	B_0	B_1	B_2	B_3				
4	1	2048	74	74	67	61	59				
4	2	2048	74	74	73	64	62				
4	M4	2048	74	74	67	61	62				
M	FD	S	Sem Detector	Com Detector							
			$B_{0,1,2,3,4,5,6,7}$	B_0	B_1	B_2	B_3	B_4	B_5	B_6	B_7
8	1	2048	74	74	74	66	64	64	63	62	62
8	2	2048	74	74	74	69	71	73	71	72	70
8	4	2048	74	74	74	74	74	74	74	74	74
8	M8	2048	74	74	74	69	64	64	71	72	70

Apresentamos nas Tabelas 27 e 28 o número de multiplicações por iteração tanto em implementação normal (ver Equação (89)) quanto na implementação eficiente (ver Equação (90)), para as simulações realizadas neste experimento. Nas quais observa-se que o aumento de S_k , reduziu número total de blocos por iteração, o que dificultou ao algoritmo de separação retirar blocos considerados desprovidos de fala, o que resultou em uma sutil diferença quando empregou-se o detector de voz apenas para a Mistura 4, por apresentar intervalos de silêncio.

Tabela 27 - Número de multiplicações por iteração em implementação comum e em implementação eficiente, para diferentes valores de M e FD para a Mistura 3.

M	FD	S	Sem Detector		Com Detector	
			NMPI	$NMPI_{IE}$	NMPI	$NMPI_{IE}$
1	1	2048	$5,09 \times 10^{12}$	$4,11 \times 10^7$	$5,09 \times 10^{12}$	$4,11 \times 10^7$
2	1	2048	$1,02 \times 10^{13}$	$8,21 \times 10^7$	$1,02 \times 10^{13}$	$8,21 \times 10^7$
2	2	2048	$5,09 \times 10^{12}$	$4,11 \times 10^7$	$5,09 \times 10^{12}$	$4,11 \times 10^7$
2	$M2$	2048	$5,41 \times 10^{12}$	$5,06 \times 10^7$	$5,41 \times 10^{12}$	$5,06 \times 10^7$
4	1	2048	$2,03 \times 10^{13}$	$1,64 \times 10^8$	$2,03 \times 10^{13}$	$1,64 \times 10^8$
4	2	2048	$1,02 \times 10^{13}$	$8,21 \times 10^7$	$1,02 \times 10^{13}$	$8,21 \times 10^7$
4	$M4$	2048	$1,08 \times 10^{13}$	$1,01 \times 10^8$	$1,08 \times 10^{13}$	$1,01 \times 10^8$
8	1	2048	$4,07 \times 10^{13}$	$3,29 \times 10^8$	$4,07 \times 10^{13}$	$3,29 \times 10^8$
8	2	2048	$2,03 \times 10^{13}$	$1,64 \times 10^8$	$2,03 \times 10^{13}$	$1,64 \times 10^8$
8	4	2048	$1,02 \times 10^{13}$	$8,21 \times 10^7$	$1,02 \times 10^{13}$	$8,21 \times 10^7$
8	$M8$	2048	$1,21 \times 10^{13}$	$1,39 \times 10^8$	$1,21 \times 10^{13}$	$1,39 \times 10^8$

Tabela 28 - Número de multiplicações por iteração em implementação comum e em implementação eficiente, para diferentes valores de M e FD para a Mistura 4.

M	FD	S	Sem Detector		Com Detector	
			NMPI	$NMPI_{IE}$	NMPI	$NMPI_{IE}$
1	1	2048	$5,09 \times 10^{12}$	$4,11 \times 10^7$	$4,88 \times 10^{12}$	$3,94 \times 10^7$
2	1	2048	$1,02 \times 10^{13}$	$8,21 \times 10^7$	$9,56 \times 10^{12}$	$7,71 \times 10^7$
2	2	2048	$5,09 \times 10^{12}$	$4,11 \times 10^7$	$5,05 \times 10^{12}$	$4,08 \times 10^7$
2	$M2$	2048	$5,41 \times 10^{12}$	$5,06 \times 10^7$	$5,40 \times 10^{12}$	$5,05 \times 10^7$
4	1	2048	$2,03 \times 10^{13}$	$1,64 \times 10^8$	$1,79 \times 10^{13}$	$1,45 \times 10^8$
4	2	2048	$1,02 \times 10^{13}$	$8,21 \times 10^7$	$9,38 \times 10^{12}$	$7,58 \times 10^7$
4	$M4$	2048	$1,08 \times 10^{13}$	$1,01 \times 10^8$	$9,38 \times 10^{12}$	$8,85 \times 10^7$
8	1	2048	$4,07 \times 10^{13}$	$3,29 \times 10^8$	$3,64 \times 10^{13}$	$2,94 \times 10^8$
8	2	2048	$2,03 \times 10^{13}$	$1,64 \times 10^8$	$1,97 \times 10^{13}$	$1,59 \times 10^8$
8	4	2048	$1,02 \times 10^{13}$	$8,21 \times 10^7$	$1,02 \times 10^{13}$	$8,21 \times 10^7$
8	$M8$	2048	$1,21 \times 10^{13}$	$1,39 \times 10^8$	$1,06 \times 10^{13}$	$1,26 \times 10^8$

Neste experimento, em um procedimento análogo aos demais, realizamos diversos testes para diferentes valores de S_k , em que $S_k = 2048/FD_k$ foi o comprimento dos filtros encontrado, capaz de promover bons resultados para o algoritmo de separação, para separar as misturas deste experimento.

5.3.6 Experimento 5

Empregamos neste experimento as misturas obtidas no experimento 4 e implementamos a separação cega tanto em banda cheia quanto em sub-bandas para a estrutura não uniforme investigada no segundo experimento para $S_k = 2048/FD_k$.

Face ao exposto apresentamos nas Tabelas 29 e 30 os resultados médios finais das

métricas SIR, SAR e SDR (em dB), para $S_k = 2048/FD_k$ tanto em banda cheia quanto em sub-banda adotando a estrutura não uniforme, para as Misturas 3 e 4.

Tabela 29 - Resultados médios finais da SIR, SAR e SDR, para diferentes valores de FD e para $S_k = 2048/FD_k$ para a Mistura 3 em estrutura não uniforme.

M	DIV	SIR		SAR		SDR		FD				
		SD	CD	SD	CD	SD	CD	B_0	B_1	B_2	B_3	B_4
3	1	21,82	21,82	10,27	10,27	9,81	9,81	4	4	2	–	–
	2	23,01	23,01	10,22	10,22	9,87	9,87	2	2	1	–	–
	4	22,87	22,87	10,36	10,36	10,04	10,04	1	1	1	–	–
	$M3$	22,79	22,79	10,25	10,25	9,93	9,93	1	1	2	–	–
4	1	16,56	16,56	9,98	9,98	8,96	8,96	8	8	4	2	–
	2	23,78	23,78	10,10	10,10	9,79	9,79	4	4	2	1	–
	4	24,56	24,56	10,30	10,30	10,08	10,08	2	2	1	1	–
	8	24,64	24,64	10,33	10,33	10,11	10,11	1	1	1	1	–
	$M4$	24,31	24,31	9,21	9,21	9,02	9,02	1	2	1	1	–
5	1	9,10	9,10	8,95	8,95	5,72	5,72	16	16	8	4	2
	2	20,63	20,63	9,53	9,53	9,10	9,10	8	8	4	2	1
	4	26,46	26,46	10,32	10,32	10,15	10,15	4	4	2	1	1
	8	26,52	26,52	10,34	10,34	10,17	10,17	2	2	1	1	1
	16	26,60	26,60	10,37	10,37	10,21	10,21	1	1	1	1	1
	$M5$	26,04	26,04	8,08	8,08	7,98	7,98	2	2	1	4	2

Tabela 30 - Resultados médios finais da SIR, SAR e SDR, para diferentes valores de FD e para $S_k = 2048/FD_k$ para a Mistura 4 em estrutura não uniforme.

M	DIV	SIR		SAR		SDR		FD				
		SD	CD	SD	CD	SD	CD	B_0	B_1	B_2	B_3	B_4
3	1	11,55	11,55	15,87	15,87	10,06	10,06	4	4	2	–	–
	2	18,11	18,18	17,16	17,16	14,56	14,58	2	2	1	–	–
	4	19,69	19,77	18,33	18,35	15,90	15,93	1	1	1	–	–
	$M3$	20,98	20,97	13,73	13,77	12,88	12,90	1	1	2	–	–
4	1	8,20	8,20	13,65	13,65	6,83	6,83	8	8	4	2	–
	2	12,92	12,92	13,95	13,95	10,22	10,22	4	4	2	1	–
	4	19,71	19,79	18,14	18,15	15,81	15,83	2	2	1	1	–
	8	19,80	19,88	18,30	18,32	15,94	15,96	1	1	1	1	–
	$M4$	22,10	22,12	9,31	9,32	9,04	9,05	1	2	1	1	–
5	1	4,47	4,47	12,49	12,49	3,61	3,61	16	16	8	4	2
	2	4,19	4,20	10,12	10,12	2,88	2,88	8	8	4	2	1
	4	19,90	19,99	17,72	17,74	15,62	15,65	4	4	2	1	1
	8	20,00	20,08	17,95	17,97	15,81	15,84	2	2	1	1	1
	16	20,08	20,16	17,95	17,96	15,84	15,87	1	1	1	1	1
	$M5$	22,35	22,35	2,23	2,23	2,16	2,16	2	2	1	4	2

Apresentamos nas Tabelas 31 e 32 o número de blocos utilizados pelo algoritmo de separação a cada iteração neste experimento para as Mistura 3 e 4 respectivamente para a estrutura não uniforme.

Tabela 31 - Número de blocos empregados no algoritmo de separação, para diferentes valores de M e FD para a Mistura 3.

M	DIV	S	Sem Detector	Com Detector				
			$B_{0,1,2}$	B_0	B_1	B_2		
3	1	2048	74	74	74	74		
	2	2048	74	74	74	74		
	4	2048	74	74	74	74		
	$M3$	2048	74	74	74	74		
M	DIV	S	Sem Detector	Com Detector				
			$B_{0,1,2}$	B_0	B_1	B_2	B_3	
4	1	2048	74	74	74	74	74	
	2	2048	74	74	74	74	74	
	4	2048	74	74	74	74	74	
	8	2048	74	74	74	74	74	
	$M4$	2048	74	74	74	74	74	
M	DIV	S	Sem Detector	Com Detector				
			$B_{0,1,2}$	B_0	B_1	B_2	B_3	B_4
5	1	2048	74	74	74	74	74	74
	2	2048	74	74	74	74	74	74
	4	2048	74	74	74	74	74	74
	8	2048	74	74	74	74	74	74
	16	2048	74	74	74	74	74	74
	$M5$	768	74	74	74	74	74	74

Tabela 32 - Número de blocos empregados no algoritmo de separação, para diferentes valores de M e FD para a Mistura 4.

M	DIV	S	Sem Detector	Com Detector				
			$B_{0,1,2}$	B_0	B_1	B_2		
3	1	2048	74	74	74	73		
	2	2048	74	74	73	65		
	4	2048	74	74	67	65		
	$M3$	2048	74	74	67	73		
M	DIV	S	Sem Detector	Com Detector				
			$B_{0,1,2}$	B_0	B_1	B_2	B_3	
4	1	2048	74	74	74	74	73	
	2	2048	74	74	74	73	65	
	4	2048	74	74	74	67	65	
	8	2048	74	74	74	67	65	
	$M4$	2048	74	74	74	67	65	
M	DIV	S	Sem Detector	Com Detector				
			$B_{0,1,2}$	B_0	B_1	B_2	B_3	B_4
5	1	2048	74	74	74	74	74	73
	2	2048	74	74	74	74	73	65
	4	2048	74	74	74	74	67	65
	8	2048	74	74	74	74	67	65
	16	2048	74	74	74	74	67	65
	$M5$	768	74	74	74	74	74	73

Apresentamos nas Tabelas 33 e 34 os resultados do número de multiplicações por iteração, para as simulações apresentadas para $S_k = 2048/FD_k$.

Tabela 33 - Tabela Complexidade Computacional para a Mistura 3 - $S_k = 2048/FD_k$.

M	DIV	S	Sem Detector		Com Detector	
			NMPI	$NMPI_{IE}$	NMPI	$NMPI_{IE}$
3	1	2048	$3,58 \times 10^{11}$	$1,39 \times 10^7$	$3,58 \times 10^{11}$	$1,39 \times 10^7$
3	2	2048	$5,72 \times 10^{12}$	$6,01 \times 10^7$	$5,72 \times 10^{12}$	$6,01 \times 10^7$
3	4	2048	$1,53 \times 10^{13}$	$1,23 \times 10^8$	$1,53 \times 10^{13}$	$1,23 \times 10^8$
3	$M3$	2048	$1,05 \times 10^{13}$	$9,17 \times 10^7$	$1,05 \times 10^{13}$	$9,17 \times 10^7$
4	1	2048	$3,40 \times 10^{11}$	$1,27 \times 10^7$	$3,40 \times 10^{11}$	$1,27 \times 10^7$
4	2	2048	$5,44 \times 10^{12}$	$5,50 \times 10^7$	$5,44 \times 10^{12}$	$5,50 \times 10^7$
4	4	2048	$1,08 \times 10^{13}$	$1,01 \times 10^8$	$1,08 \times 10^{13}$	$1,01 \times 10^8$
4	8	2048	$2,03 \times 10^{13}$	$1,64 \times 10^8$	$2,03 \times 10^{13}$	$1,64 \times 10^8$
4	$M4$	2048	$1,56 \times 10^{13}$	$1,33 \times 10^8$	$1,56 \times 10^{13}$	$1,33 \times 10^8$
5	1	2048	$3,39 \times 10^{11}$	$1,24 \times 10^7$	$3,39 \times 10^{11}$	$1,24 \times 10^7$
5	2	2048	$5,43 \times 10^{12}$	$5,38 \times 10^7$	$5,43 \times 10^{12}$	$5,38 \times 10^7$
5	4	2048	$1,05 \times 10^{13}$	$9,60 \times 10^7$	$1,05 \times 10^{13}$	$9,60 \times 10^7$
5	8	2048	$1,59 \times 10^{13}$	$1,42 \times 10^8$	$1,59 \times 10^{13}$	$1,42 \times 10^8$
5	16	2048	$2,54 \times 10^{13}$	$2,05 \times 10^8$	$2,54 \times 10^{13}$	$2,05 \times 10^8$
5	$M5$	2048	$6,06 \times 10^{12}$	$7,18 \times 10^7$	$6,06 \times 10^{12}$	$7,18 \times 10^7$

Tabela 34 - Tabela Complexidade Computacional para a Mistura 4 - $S_k = 2048/FD_k$.

M	DIV	S	Sem Detector		Com Detector	
			NMPI	$NMPI_{IE}$	NMPI	$NMPI_{IE}$
3	1	2048	$3,58 \times 10^{11}$	$1,39 \times 10^7$	$3,54 \times 10^{11}$	$1,38 \times 10^7$
3	2	2048	$5,72 \times 10^{12}$	$6,01 \times 10^7$	$5,10 \times 10^{12}$	$5,50 \times 10^7$
3	4	2048	$1,53 \times 10^{13}$	$1,23 \times 10^8$	$1,42 \times 10^{13}$	$1,14 \times 10^8$
3	$M3$	2048	$1,05 \times 10^{13}$	$9,17 \times 10^7$	$1,00 \times 10^{13}$	$8,76 \times 10^7$
4	1	2048	$3,40 \times 10^{11}$	$1,27 \times 10^7$	$3,36 \times 10^{11}$	$1,26 \times 10^7$
4	2	2048	$5,44 \times 10^{12}$	$5,50 \times 10^7$	$4,82 \times 10^{12}$	$4,98 \times 10^7$
4	4	2048	$1,08 \times 10^{13}$	$1,01 \times 10^8$	$9,71 \times 10^{12}$	$9,23 \times 10^7$
4	8	2048	$2,03 \times 10^{13}$	$1,64 \times 10^8$	$1,92 \times 10^{13}$	$1,55 \times 10^8$
4	$M4$	2048	$1,56 \times 10^{13}$	$1,33 \times 10^8$	$1,45 \times 10^{13}$	$1,24 \times 10^8$
5	1	2048	$3,39 \times 10^{11}$	$1,24 \times 10^7$	$3,35 \times 10^{11}$	$1,23 \times 10^7$
5	2	2048	$5,43 \times 10^{12}$	$5,38 \times 10^7$	$4,80 \times 10^{12}$	$4,86 \times 10^7$
5	4	2048	$1,05 \times 10^{13}$	$9,60 \times 10^7$	$9,43 \times 10^{12}$	$8,71 \times 10^7$
5	8	2048	$1,59 \times 10^{13}$	$1,42 \times 10^8$	$1,48 \times 10^{13}$	$1,33 \times 10^8$
5	16	2048	$2,54 \times 10^{13}$	$2,05 \times 10^8$	$2,43 \times 10^{13}$	$1,96 \times 10^8$
5	$M5$	2048	$6,06 \times 10^{12}$	$7,18 \times 10^7$	$6,06 \times 10^{12}$	$7,17 \times 10^7$

CONCLUSÃO

Neste trabalho, a detecção de voz como uma etapa de pré-processamento na tarefa de separação cega de fontes no domínio do tempo foi investigada. Foram utilizadas misturas convolutivas com presença e com ausência de trechos de silêncio, para diferentes condições de reverberação. Para verificação da qualidade das estimativas foram usadas a SIR, SAR e SDR.

Simulações computacionais foram realizadas envolvendo sinais de fala, mostrando o desempenho superior das estruturas em sub-bandas em relação à banda cheia, considerando diferentes fatores de decimação.

A utilização dos detectores de voz conforme mencionamos nos experimentos, mostrou-se bastante interessante porque o algoritmo de separação teve um desempenho similar nos casos de misturas sem trechos de silêncio e ligeiramente superior quando as misturas continham trechos de silêncio; e para os casos de misturas com ausência de fala promoveu significativa redução do número de blocos utilizado a cada iteração e conseqüentemente redução da complexidade computacional envolvida, principalmente nas bandas mais altas das implementações em sub-bandas, onde a concentração de energia do sinal de voz é menor.

Almejando dar continuidade a este trabalho, podemos sugerir para futuras pesquisas as seguintes ideias que podem ser exploradas:

- Implementar o emprego do detector de voz na técnica de separação cega empregada neste trabalho, bem como da estrutura em sub-bandas, em implementações *online* e *block online*;
- Implementar e avaliar o desempenho de outros detectores de voz não contemplados neste trabalho;
- Explorar outras estruturas em sub-bandas;
- Avaliar o desempenho do detector de voz em outras técnicas de separação cega, como estratégia de desativar o algoritmo na ausência de fala;
- Explorar os resultados com detectores de voz e estruturas em sub-bandas através da aquisição de misturas reais;
- Investigar a redução do comprimento dos filtros de separação, bem como trabalhar em diferentes bandas com fatores de decimação distintos, como estratégia de compensar a complexidade computacional.

APÊNDICE A

Neste apêndice serão apresentados conceitos e propriedades matemáticas responsáveis em auxiliar na prova ou derivação de alguns resultados importantes utilizados nesta tese.

A.1 Algumas Propriedades

As propriedades apresentadas a seguir serão empregadas posteriormente nas deduções utilizadas neste trabalho e podem ser encontradas, bem como suas definições, em [3], [46] e [66]

$$(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T \quad (95)$$

$$\text{Tr}(\mathbf{A}) = \sum_{i=1}^m a_{ii} \quad (96)$$

$$\text{Tr}(\mathbf{A} + \mathbf{B}) = \text{Tr}(\mathbf{A}) + \text{Tr}(\mathbf{B}) \quad (97)$$

$$\text{Tr}(\mathbf{BC}) = \text{Tr}(\mathbf{CB}) \quad (98)$$

$$\frac{\partial \{\log [\det (\mathbf{A})]\}}{\partial \mathbf{A}} = (\mathbf{A}^T)^{-1} \quad (99)$$

$$\partial \{\log \det(\mathbf{A})\} = \text{Tr}(\mathbf{A}^{-1} \partial \mathbf{A}) \quad (100)$$

$$\partial (\mathbf{A} + \mathbf{B}) = \partial (\mathbf{A}) + \partial (\mathbf{B}) \quad (101)$$

$$\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}^*} = \frac{1}{2} \left[\frac{\partial f(\mathbf{A})}{\partial \Re \mathbf{A}} + i \frac{\partial f(\mathbf{A})}{\partial \Im \mathbf{A}} \right] \quad (102)$$

$$\frac{\partial \text{Tr}(\mathbf{A} \mathbf{X}^H)}{\partial \Re \mathbf{X}} = \mathbf{A} \quad (103)$$

$$i \frac{\partial \text{Tr}(\mathbf{A} \mathbf{X}^H)}{\partial \Im \mathbf{X}} = \mathbf{A} \quad (104)$$

$$\frac{\partial \text{Tr}(\mathbf{A}\mathbf{X})}{\partial \Re \mathbf{X}} = \mathbf{A}^T \quad (105)$$

$$i \frac{\partial \text{Tr}(\mathbf{A}\mathbf{X})}{\partial \Im \mathbf{X}} = -\mathbf{A}^T \quad (106)$$

A.2 Provas e derivações

Apresentaremos a seguir, as manipulações matemáticas realizadas na função de custo na Equação (41) supracitadas no Capítulo 2 e contempladas em [23], assim temos:

$$\begin{aligned} \nabla_{\mathbf{W}} \Phi(m) &= 2 \frac{\partial \left\{ \sum_{i=0}^{\infty} [\beta(i, m) \log \det \text{bdiag}(\mathbf{Y}^H \cdot \mathbf{Y}) - \log \det(\mathbf{Y}^H \cdot \mathbf{Y})] \right\}}{\partial \mathbf{W}^*} \\ &= 2 \sum_{i=0}^{\infty} \beta(i, m) \left\{ \frac{\partial [\log \det \text{bdiag}(\mathbf{Y}^H \cdot \mathbf{Y})]}{\partial \mathbf{W}^*} - \frac{\partial [\log \det(\mathbf{Y}^H \cdot \mathbf{Y})]}{\partial \mathbf{W}^*} \right\} \\ &= 2 \sum_{i=0}^{\infty} \beta(i, m) \left\{ \frac{\partial [\log \det \text{bdiag}((\mathbf{X}\mathbf{W})^H \mathbf{X}\mathbf{W})]}{\partial \mathbf{W}^*} - \frac{\partial [\log \det((\mathbf{X}\mathbf{W})^H \mathbf{X}\mathbf{W})]}{\partial \mathbf{W}^*} \right\} \\ &= 2 \sum_{i=0}^{\infty} \beta(i, m) \left\{ \frac{\partial [\log \det \text{bdiag}(\mathbf{W}^H \mathbf{X}^H \mathbf{X}\mathbf{W})]}{\partial \mathbf{W}^*} - \frac{\partial [\log \det(\mathbf{W}^H \mathbf{X}^H \mathbf{X}\mathbf{W})]}{\partial \mathbf{W}^*} \right\} \\ &= 2 \sum_{i=0}^{\infty} \beta(i, m) \left\{ \underbrace{\frac{\partial [\log \det \text{bdiag}(\mathbf{W}^H \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W})]}{\partial \mathbf{W}^*}}_{\mathbf{A}_1} - \underbrace{\frac{\partial [\log \det(\mathbf{W}^H \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W})]}{\partial \mathbf{W}^*}}_{\mathbf{A}_2} \right\}. \quad (107) \end{aligned}$$

Podemos deduzir os termos na Equação (107) separadamente, e por simplicidade começaremos por \mathbf{A}_2 , que através da Equação (100) obtemos:

$$\begin{aligned} \mathbf{A}_2 &= \frac{\partial [\log \det(\mathbf{W}^H \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W})]}{\partial \mathbf{W}^*} = \frac{\partial f(\mathbf{A})}{\partial \mathbf{A}^*} \Rightarrow \\ \partial f(\mathbf{A}) &= \partial [\log \det(\mathbf{W}^H \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W})] = \text{Tr} \left[(\mathbf{W}^H \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W})^{-1} \partial (\mathbf{W}^H \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W}) \right]. \quad (108) \end{aligned}$$

Face ao exposto, podemos aplicar a Regra da Cadeia na Equação (108), e assim encontramos:

$$\partial f(\mathbf{A}) = \text{Tr} \left\{ (\mathbf{W}^H \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W})^{-1} [\partial (\mathbf{W}^H) \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W} + \mathbf{W}^H \partial (\mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W})] \right\}. \quad (109)$$

Empregando-se as propriedades contidas nas equações (97) e (98) respectivamente e sabendo que $\mathbf{R}_{\mathbf{X}\mathbf{X}}$ não depende de \mathbf{W} encontramos:

$$\begin{aligned}
\partial f(\mathbf{A}) &= \text{Tr} \left[\underbrace{(\mathbf{W}^H \mathbf{R}_{\text{XX}} \mathbf{W})^{-1}}_{\mathbf{B}} \partial (\mathbf{W}^H) \underbrace{\mathbf{R}_{\text{XX}} \mathbf{W}}_{\mathbf{C}} \right] + \text{Tr} \left[(\mathbf{W}^H \mathbf{R}_{\text{XX}} \mathbf{W})^{-1} \mathbf{W}^H \partial (\mathbf{R}_{\text{XX}} \mathbf{W}) \right] \\
&= \text{Tr} \left[\mathbf{R}_{\text{XX}} \mathbf{W} (\mathbf{W}^H \mathbf{R}_{\text{XX}} \mathbf{W})^{-1} \partial (\mathbf{W}^H) \right] + \text{Tr} \left[(\mathbf{W}^H \mathbf{R}_{\text{XX}} \mathbf{W})^{-1} \mathbf{W}^H \mathbf{R}_{\text{XX}} \partial (\mathbf{W}) \right]. \quad (110)
\end{aligned}$$

Para aplicamos a propriedade da Equação (102) precisamos primeiro encontrar a derivada em relação a parte real de \mathbf{W} , e aplicando as equações (103) e (105),obtemos:

$$\begin{aligned}
\frac{\partial f(\mathbf{A})}{\partial \Re \mathbf{W}} &= \frac{\text{Tr} \left[\mathbf{R}_{\text{XX}} \mathbf{W} (\mathbf{W}^H \mathbf{R}_{\text{XX}} \mathbf{W})^{-1} \partial (\mathbf{W}^H) \right]}{\partial \Re \mathbf{W}} + \frac{\text{Tr} \left[(\mathbf{W}^H \mathbf{R}_{\text{XX}} \mathbf{W})^{-1} \mathbf{W}^H \mathbf{R}_{\text{XX}} \partial (\mathbf{W}) \right]}{\partial \Re \mathbf{W}} \\
&= \text{Tr} \left[\mathbf{R}_{\text{XX}} \mathbf{W} (\mathbf{W}^H \mathbf{R}_{\text{XX}} \mathbf{W})^{-1} \frac{\partial (\mathbf{W}^H)}{\partial \Re \mathbf{W}} \right] + \text{Tr} \left[(\mathbf{W}^H \mathbf{R}_{\text{XX}} \mathbf{W})^{-1} \mathbf{W}^H \mathbf{R}_{\text{XX}} \frac{\partial (\mathbf{W})}{\partial \Re \mathbf{W}} \right] \\
&= \mathbf{R}_{\text{XX}} \mathbf{W} (\mathbf{W}^H \mathbf{R}_{\text{XX}} \mathbf{W})^{-1} + \left[(\mathbf{W}^H \mathbf{R}_{\text{XX}} \mathbf{W})^{-1} \mathbf{W}^H \mathbf{R}_{\text{XX}} \right]^T. \quad (111)
\end{aligned}$$

Agora precisamos encontrar a derivada em relação a parte imaginária de \mathbf{W} , e aplicando as equações (104) e (106) encontramos:

$$\begin{aligned}
\frac{\partial f(\mathbf{A})}{\partial \Im \mathbf{W}} &= \frac{\text{Tr} \left[\mathbf{R}_{\text{XX}} \mathbf{W} (\mathbf{W}^H \mathbf{R}_{\text{XX}} \mathbf{W})^{-1} \partial (\mathbf{W}^H) \right]}{\partial \Im \mathbf{W}} + \frac{\text{Tr} \left[(\mathbf{W}^H \mathbf{R}_{\text{XX}} \mathbf{W})^{-1} \mathbf{W}^H \mathbf{R}_{\text{XX}} \partial (\mathbf{W}) \right]}{\partial \Im \mathbf{W}} \\
&= \text{Tr} \left[\mathbf{R}_{\text{XX}} \mathbf{W} (\mathbf{W}^H \mathbf{R}_{\text{XX}} \mathbf{W})^{-1} \frac{\partial (\mathbf{W}^H)}{\partial \Im \mathbf{W}} \right] + \text{Tr} \left[(\mathbf{W}^H \mathbf{R}_{\text{XX}} \mathbf{W})^{-1} \mathbf{W}^H \mathbf{R}_{\text{XX}} \frac{\partial (\mathbf{W})}{\partial \Im \mathbf{W}} \right] \\
&= \mathbf{R}_{\text{XX}} \mathbf{W} (\mathbf{W}^H \mathbf{R}_{\text{XX}} \mathbf{W})^{-1} - \left[(\mathbf{W}^H \mathbf{R}_{\text{XX}} \mathbf{W})^{-1} \mathbf{W}^H \mathbf{R}_{\text{XX}} \right]^T. \quad (112)
\end{aligned}$$

Podemos então, aplicar a propriedade da Equação (102) e assim:

$$\begin{aligned}
\mathbf{A}_2 = \frac{\partial f(\mathbf{A})}{\partial \mathbf{W}^*} &= \frac{1}{2} \left[\frac{\partial f(\mathbf{A})}{\partial \Re \mathbf{W}} + i \frac{\partial f(\mathbf{A})}{\partial \Im \mathbf{W}} \right] = \frac{1}{2} \left[2 \mathbf{R}_{\text{XX}} \mathbf{W} (\mathbf{W}^H \mathbf{R}_{\text{XX}} \mathbf{W})^{-1} \right] \\
&= \mathbf{R}_{\text{XX}} \mathbf{W} (\mathbf{W}^H \mathbf{R}_{\text{XX}} \mathbf{W})^{-1}. \quad (113)
\end{aligned}$$

A Equação (113) apresenta o resultado do segundo termo indicado na Equação (41). Agora para a dedução \mathbf{A}_1 na Equação (41), também no Capítulo 2, será necessário reescrevermos a equação para facilitar a dedução, conforme proposto em [14], e assim, temos que:

$$\mathbf{A}_1 = \frac{\partial [\log \det \text{bdiag}(\mathbf{W}^H \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W})]}{\partial \mathbf{W}^*} = \frac{\partial f(\mathbf{A})}{\partial \mathbf{W}^*} \therefore \mathbf{A} = \text{bdiag}(\mathbf{W}^H \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W}), \quad (114)$$

sendo $\mathbf{A} = \text{bdiag}(\mathbf{W}^H \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W})$.

Definindo w_{pq}^{PQ} como um elemento localizado na p -ésima linha e q -ésima coluna da submatriz \mathbf{W}_{PQ} que por sua vez, encontra-se na P -ésima linha e Q -ésima coluna da matriz de separação \mathbf{W} . A mesma nomenclatura aplica-se ao termo a_{kj}^{KJ} referente a matriz global \mathbf{A} na Equação (114).

Utilizando a Regra da Cadeia para expressar \mathbf{A}_1 em relação a cada um de seus elementos, encontramos:

$$\mathbf{A}_1|_{w_{pq}^{PQ}} = \frac{\partial f(\mathbf{A})}{\partial \mathbf{W}^*} = \sum_{k,j,K,J} \left[\underbrace{\frac{\partial f(\mathbf{A})}{\partial a_{kj}^{KJ}}}_{\mathbf{B}_1} \underbrace{\frac{\partial a_{kj}^{KJ}}{\partial (w_{pq}^{PQ})^*}}_{\mathbf{B}_2} \right]. \quad (115)$$

Podemos expressar o termo \mathbf{B}_1 no formato matricial, e assim, aplicando a propriedade da Equação (99), obtemos que:

$$\mathbf{B}_1 = \frac{\partial f(\mathbf{A})}{\partial \mathbf{A}} = \frac{\partial \{\log [\det(\mathbf{A})]\}}{\partial \mathbf{A}} = (\mathbf{A}^T)^{-1}. \quad (116)$$

Considerando a matriz \mathbf{A} uma matriz real ($\mathbf{A}^H = \mathbf{A}^T$) e simétrica ($\mathbf{A}^T = \mathbf{A}$) podemos concluir que:

$$\mathbf{B}_1 = \frac{\partial f(\mathbf{A})}{\partial \mathbf{A}} = (\mathbf{A}^T)^{-1} = (\mathbf{A}^H)^{-1} = (\mathbf{A})^{-1} = \text{bdiag}^{-1}(\mathbf{W}^H \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W}). \quad (117)$$

Para o cálculo do termo \mathbf{B}_2 , precisamos definir b_{ij}^{IJ} como elemento de $\mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W}$, seguindo a mesma nomenclatura de w_{pq}^{PQ} e a_{kj}^{KJ} . Visando representar o termo a_{kj}^{KJ} em função de w_{ki}^{KI} e b_{ij}^{IJ} temos:

$$\mathbf{A} = \sum_i \sum_I [(w_{ki}^{KI})^H b_{ij}^{IJ} \delta_{KJ}] = \sum_i \sum_I [(w_{ki}^{KI})^* b_{ij}^{IJ} \delta_{KJ}], \quad (118)$$

em que empregamos o símbolo de Kronecker para representa o operador $\text{bdiag}(\cdot)$ sendo que $\delta_{KJ} = 1$ para $K = J$ e $\delta_{KJ} = 0$ caso contrário. O termo \mathbf{B}_2 pode então ser representado

por [14]:

$$\begin{aligned}
\mathbf{B}_2 &= \frac{\partial a_{kj}^{KJ}}{\partial (w_{pq}^{PQ})^*} = \frac{\partial}{\partial (w_{pq}^{PQ})^*} \sum_i \sum_I [(w_{ki}^{KI})^* b_{ij}^{IJ} \delta_{KJ}] \\
&= \sum_i \sum_I (\delta_{ip} \delta_{kq} \delta_{IP} \delta_{KQ} \delta_{KJ} b_{ij}^{IJ}) \\
&= \delta_{kq} \delta_{KQ} \delta_{KJ} b_{pj}^{PJ}.
\end{aligned} \tag{119}$$

Combinando as equações (117) e (119) na Equação (115) e adaptando-as obtemos:

$$\begin{aligned}
\mathbf{A}_1 &= \sum_{k,j,K,J} \left\{ [\text{bdiag}^{-1} (\mathbf{W}^H \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W})]_{kj}^{KJ} \delta_{kq} \delta_{KQ} \delta_{KJ} [\mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W}]_{pj}^{PJ} \right\} \\
&= \sum_j \left\{ [\text{bdiag}^{-1} (\mathbf{W}^H \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W})]_{qj}^{QQ} [\mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W}]_{pj}^{PJ} \right\} \\
&= [(\mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W}) \text{bdiag}^{-1} (\mathbf{W}^H \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W})]_{pq}^{PQ} \\
&= \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W} \text{bdiag}^{-1} (\mathbf{W}^H \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W}).
\end{aligned} \tag{120}$$

Para concluir esta análise, podemos substituir \mathbf{A}_1 e \mathbf{A}_2 na Equação (107), obtendo como resultado:

$$\begin{aligned}
\nabla_{\mathbf{W}} \Phi(m) &= 2 \sum_{i=0}^{\infty} \beta(i, m) \left\{ \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W} [\text{bdiag}^{-1} (\mathbf{W}^H \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W})] - \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W} (\mathbf{W}^H \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W})^{-1} \right\} \\
&= 2 \sum_{i=0}^{\infty} \beta(i, m) \left\{ \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W} \left[\text{bdiag}^{-1} \left(\underbrace{\mathbf{W}^H \mathbf{X}^H}_{\mathbf{Y}^H} \underbrace{\mathbf{X} \mathbf{W}}_{\mathbf{Y}} \right) \right] - \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W} (\mathbf{W}^H \mathbf{X}^H \mathbf{X} \mathbf{W})^{-1} \right\} \\
&= 2 \sum_{i=0}^{\infty} \beta(i, m) \left\{ \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W} [\text{bdiag}^{-1} (\mathbf{Y}^H \mathbf{Y})] - \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W} (\mathbf{Y}^H \mathbf{Y})^{-1} \right\} \\
&= 2 \sum_{i=0}^{\infty} \beta(i, m) \mathbf{R}_{\mathbf{X}\mathbf{X}} \mathbf{W} \left[(\text{bdiag}^{-1} \mathbf{R}_{\mathbf{Y}\mathbf{Y}}) - \mathbf{R}_{\mathbf{Y}\mathbf{Y}}^{-1} \right] \\
&= 2 \sum_{i=0}^{\infty} \beta(i, m) \mathbf{X}^H \mathbf{X} \mathbf{W} \left[(\text{bdiag}^{-1} \mathbf{R}_{\mathbf{Y}\mathbf{Y}}) - \mathbf{R}_{\mathbf{Y}\mathbf{Y}}^{-1} \right] \\
&= 2 \sum_{i=0}^{\infty} \beta(i, m) \underbrace{\mathbf{X}^H \mathbf{Y}}_{\mathbf{R}_{\mathbf{X}\mathbf{Y}}} \left[(\text{bdiag}^{-1} \mathbf{R}_{\mathbf{Y}\mathbf{Y}}) - \mathbf{R}_{\mathbf{Y}\mathbf{Y}}^{-1} \right] \\
&= 2 \sum_{i=0}^{\infty} \beta(i, m) \mathbf{R}_{\mathbf{X}\mathbf{Y}} \mathbf{R}_{\mathbf{Y}\mathbf{Y}}^{-1} \left[\mathbf{R}_{\mathbf{Y}\mathbf{Y}} (\text{bdiag}^{-1} \mathbf{R}_{\mathbf{Y}\mathbf{Y}}) - \mathbf{I} \right]
\end{aligned}$$

$$= 2 \sum_{i=0}^{\infty} \beta(i, m) \mathbf{R}_{xy} \mathbf{R}_{yy}^{-1} (\mathbf{R}_{yy} - \text{bdiag} \mathbf{R}_{yy}) \text{bdiag}^{-1} \mathbf{R}_{yy}. \quad (121)$$

O resultado final da expressão na Equação (121) equivale ao resultado exposto na Equação (41).

REFERÊNCIAS

- [1] CAVALCANTE, C. C. *Sobre Separação Cega de Fontes: Proposições e Análises de Estratégias para Processamento Multi-Usuário*. 268 p. Tese (Doutorado) — Faculdade de Engenharia Elétrica e Computação, Universidade Estadual de Campinas, Campinas, Abril 2004.
- [2] PETRAGLIA, M.; BATALHEIRO, P.; HADDAD, D. Métodos de Separação Cega de Fontes. *Tutoriais do XVII Congresso Brasileiro de Automata*, v. X, n. X, p. 25, Abril 2008.
- [3] CICHOCKI, A.; AMARI, S. ichi. *Adaptive Blind Signal and Image Processing*. [S.l.]: John Wiley & Sons, Ltd, 2002. 554 p.
- [4] LATHAUWER, L. D.; MOOR, B. D.; VANDEWALLE, J. Fetal electrocardiogram extraction by source subspace separation. In: *Proc. IEEE SP/ATHOS Workshop on Higher-Order Statistics*. Girona, Spain: [s.n.], 1995. p. 134–138.
- [5] MAKEIG, S. et al. Independent component analysis of electroencephalographic data. *Advances in Neural Information Processing Systems*, Denver, CO: MIT Press, v. 8, 1995.
- [6] KARHUNEN, J. et al. Applications of neural blind separation to signal and image processing. In: *Proc. ICASSP*. [S.l.: s.n.], 1997. v. 1, p. 131–134.
- [7] KIM, T. et al. Blind Source Separation Exploiting Higher-Order Frequency Dependencies. *IEEE Transactions on Audio, Speech and Language Processing*, v. 15, n. 1, p. 70–78, January 2007.
- [8] SAWADA, H. et al. A robust and precisemethod for solving the permutation problem of frequency-domain blind source separation. *IEEE Transactions on Speech And Audio Processing*, v. 12, n. 5, p. 530–538, 2004.
- [9] HADDAD, D. B.; PETRAGLIA, M. R.; BATALHEIRO, P. B. Técnicas para Separação Cega de Fontes aplicadas na Melhoria de Desempenho de um Classificador de Palavras Isoladas. In: *5o. Congresso de Engenharia de Áudio da AES-Brasil*. São Paulo: [s.n.], 2007. v. 1, p. 20–27.

- [10] ACERNESE, F. et al. Neural networks for blind source separation of stromboli explosion quakes. *IEEE Transactions on Neural Networks*, v. 14.
- [11] ICHIR, M. M.; MOHAMMAD-DJAFARI, A. Hidden markov models for wavelet-based blind source separation. *IEEE Transactions on Image Processing*, v. 15.
- [12] ERDOGAN, A. T. Globally convergent deflationary instantaneous blind source separation algorithm for digital communication signals. *IEEE Transactions on Signal Processing*, v. 55, n. 5, p. 2182–2192, 2007.
- [13] BOURGEOIS, J.; MINKER, W. *Time-Domain Beamforming and Blind Source Separation – Speech Input in the Car Environment*. [S.l.]: Springer (v. 3), 2007. 225 p.
- [14] BUCHNER, H.; AICHNER, R.; KELLERMANN, W. A Generalization of Blind Source Separation Algorithms for Convolutional Mixtures Based on Second-Order Statistics. *IEEE Transactions on Speech and Audio Processing*, v. 13, n. 1, p. 120–134, January 2005.
- [15] BATALHEIRO, P. B.; ROCHA, R. A.; HADDAD, D. B. Separação cega de sinais de fala em sub-bandas utilizando detectores de voz. In: *XXXI Simpósio Brasileiro de Telecomunicações – SBrT2013*. Fortaleza, Brasil: [s.n.], 2013. p. 1–5.
- [16] KIM, T. Real-Time Independent Vector Analysis for Convolutional Blind Source Separation. *IEEE Transactions on Circuits and Systems*, v. 57, n. 7, p. 1431–1438, July 2010.
- [17] VINCENT, E.; GRIBONVAL, R.; FÉVOTTE, C. Performance measurement in blind audio source separation. *IEEE Transactions on Audio, Speech & Language Processing*, v. 14, n. 4, p. 1462–1469, July 2006.
- [18] SAKHNOV, K.; VERTELETSKAYA, E.; SIMAK, B. Dynamical Energy-Based Speech/Silence Detector for Speech Enhancement Applications. In: *World Congress on Engineering*. [S.l.: s.n.], 2009. v. 1, p. 6.
- [19] BATALHEIRO, P. B. *Novos Algoritmos em Subbandas para Filtragem Adaptativa e Separação Cega de Fontes*. 138 p. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, COPPE, Rio de Janeiro, Brasil, Julho 2009.
- [20] VAIDYANATHAN, P. P. *Multirate Systems and Filter Banks*. [S.l.]: Prentice Hall, 1993. 911 p.

- [21] VAIDYANATHAN, P. P. Multirate Digital Filters, Filter Banks, Polyphase Networks, and Applications: A Tutorial. *Proceedings of the IEEE*, v. 78, n. 1, p. 56–93, January 1990.
- [22] CAO, X.-R.; LIU, R.-W. General Approach to Blind Source Separation. *IEEE Transactions on Signal Processing*, v. 44, n. 3, p. 562–571, March 1996.
- [23] HADDAD, D. B. *Propostas para Separação Cega e Supervisionada de Fontes*. 155 p. Dissertação (Mestrado) — Universidade Federal do Rio de Janeiro, COPPE, Rio de Janeiro, Brasil, Junho 2008.
- [24] TALEB, A.; JUTTEN, C.; OLYMPIEFF, S. Source separation in post-nonlinear mixtures: An entropy-based algorithm. In: *ICASSP'98*. Seattle, USA: [s.n.], 1998. p. 2089–2092.
- [25] TALEB, A.; JUTTEN, C. Source Separation in Post-Nonlinear Mixtures. *IEEE Transactions on Signal Processing*, v. 47, n. 10, p. 2807–2820, October 1999.
- [26] ACHARD, S.; PHAM, D. T.; JUTTEN, C. Blind Source Separation in Post-Nonlinear Mixtures. In: *Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA2001)*. [S.l.: s.n.], 2001. p. 295–300.
- [27] DAMASCENO, N. C. *Separação Cega de Fontes Lineares e Não Lineares Usando Algoritmo Genético, Redes Neurais Artificiais RBF e Negentropia de Rényi como Medida de Independência*. 109 p. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Norte, Natal, Brasil, 2010.
- [28] ZHENG, C.-H. et al. Nonlinear blind source separation using hybrid neural networks. In: *ISNN (1)*. [S.l.: s.n.], 2006. p. 1165–1170.
- [29] ROJAS, I. et al. Nonlinear blind source separation using genetic algorithms. In: *in Proc. Int. Conf. on Independent Component Analysis and Signal Separation (ICA2001)*. [S.l.: s.n.], 2001. p. 400–405.
- [30] AICHNER, R. et al. Real-Time Convolutional Blind Source Separation Based on a Broadband Approach. In: *Proc. IEEE Int. Symposium on Independent Component Analysis and Blind Source Separation (ICA)*. Granada, Spain: Springer, 2004. v. 3195 of LNCS, p. 840–848.

- [31] DOUGLAS, S. C.; GUPTA, M. Blind Speech Separation. In: MAKINO, S. et al. (Ed.). *Convolutional Blind Source Separation for Audio Signals*. [S.l.]: Springer, 2007. cap. 1, p. 3–45.
- [32] LIANG, Y. *Optimal Algorithms for Blind Source Separation - Application to Acoustic Echo Cancellation*. 112 p. Dissertação (Mestrado) — National University of Ireland, Maynooth, September 2010.
- [33] MOLGEDEY, L.; SCHUSTER, H. G. Separation of a Mixture of Independent Signals Using Time Delayed Correlations. In: *Physical Review Letters*. [S.l.: s.n.], 1994. v. 72, n. 23, p. 3634–3637.
- [34] BUCHNER, H.; AICHNER, R.; KELLERMANN, W. Blind Speech Separation. In: MAKINO, S. et al. (Ed.). *TRINICON-based Blind System Identification with Application to Multiple-Source Localization and Separation*. [S.l.]: Springer, 2007. cap. 4, p. 101–147.
- [35] LOTTER, T.; VARY, P. Speech Enhancement by MAP Spectral Amplitude Estimation Using a Super-Gaussian Speech Model. In: *EURASIP Journal on Applied Signal Processing*. [S.l.: s.n.], 2005. v. 2005, n. 7, p. 1110–1126.
- [36] ARAKI, S. et al. The Fundamental Limitation of Frequency Domain Blind Source Separation for Convolutional Mixtures of Speech. *IEEE Transactions on Speech and Audio Processing*, v. 11, n. 2, p. 109–116, March 2003.
- [37] IKRAM, M. Z.; MORGAN, D. R. Exploring permutation inconsistency in blind separation of speech signals in a reverberant environment. In: *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*. Istanbul, Turkey: [s.n.], 2000. v. 2, p. 1041–1044.
- [38] AICHNER, R. *Acoustic Blind Source Separation in Reverberant and Noisy Environments*. 267 p. Tese (Doutorado) — University Erlangen-Nuremberg, Germany, October 2007.
- [39] TONG, L. et al. Indeterminacy and Identifiability of Blind Identification. *IEEE Transactions on Circuits and Systems*, v. 38, n. 5, p. 499–509, May 1991.
- [40] FIGUEIREDO, F. A. P. de. *Estudo da melhoria da taxa de aprendizagem de um algoritmo para separação cega de fontes utilizando técnicas vindas da teoria de redes neurais*. 135 p. Dissertação (Mestrado) — Instituto Nacional de Telecomunicações, Santa Rita do Sapucaí, Maio 2011.

- [41] PEDERSEN, M. S. et al. A Survey of Convolutional Blind Source Separation Methods. In: *Springer Handbook on Speech Processing and Speech Communication*. [S.l.]: Springer, 2007. p. 1–34.
- [42] BELL, A. J.; SEJNOWSKI, T. J. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. In: *Neural Computation*. [S.l.: s.n.], 1995. v. 7, n. 6, p. 1129–1159.
- [43] HYVÄRINEN, A.; OJA, E. A fast fixed-point algorithm for independent component analysis. In: *Neural Computation*. MIT Press Cambridge, MA: [s.n.], 1997. v. 9, p. 1483–1492.
- [44] BUCHNER, H.; AICHNER, R.; KELLERMANN, W. A generalization of a class of blind source separation algorithms for convolutional mixtures. In: *4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*. Nara, Japan: [s.n.], 2003. p. 945–950.
- [45] OPPENHEIM, A. *Inequalities connected with definite hermitian forms*. 5. ed. New Jersey: J. London Math. Soc., 1930. 114-119 p.
- [46] Kaare Brandt Petersen and Michael Syskind Pedersen. *The Matrix Cookbook*. Disponível em: <<http://orion.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf>>. Acesso em: 16 de Setembro de 2013.
- [47] BUCHNER, H.; AICHNER, R.; KELLERMANN, W. Blind source separation for convolutional mixtures: a unified treatment. In: *Audio Signal Processing, Kluwer Academic Publishers*. Boston: [s.n.], 2004. p. 255–294.
- [48] CLARK, F.; PETRAGLIA, M.; HADDAD, D. Comparação de Detectores de Atividade de Voz em Ambiente Ruidoso. In: *XXIX Simpósio Brasileiro de Telecomunicações*. Curitiba, Paraná: [s.n.], 2011. p. 5.
- [49] CLARK, F. S. P. *Cancelamento de Eco Acústico e Separação Cega de Fontes Aplicados à Telefonia Viva-Voz*. Rio de Janeiro, RJ: [s.n.], Dezembro 2010. 92 p. Monografia (Graduação) - Escola Politécnica - Departamento de Eletrônica e de Computação, Universidade Federal do Rio de Janeiro.
- [50] POLLÁK, P.; SOVKA, P.; UHLÍR, J. Noise suppression system for a car. In: *Proc. Third European Conf. on Speech, Communication and Technology*. [S.l.: s.n.], 1993. p. 1073–1076.

- [51] DAVIS, A.; NORDHOLM, S. A Low Complexity Statistical Voice Activity Detector with Performance Comparisons to ITU-T/ETSI Voice Activity Detector. *Joint Conference of The Fourth International Conference on Information, Communications & Signal Processing and Fourth Pacific-Rim Conference on Multimedia*, v. 1, p. 119–123, December 2003.
- [52] CHO, Y. D.; KONDOZ, A. Analysis and Improvement of a Statistical Model-Based Voice Activity Detector. *IEEE Signal Processing Letters on Audio, Speech and Language Processing*, v. 8, n. 10, p. 276–278, October 2001.
- [53] PRASAD, R. V. et al. Comparison of voice activity detection algorithms for VoIP. In: *Proceedings of the Seventh International Symposium on Computers and Communications - ISCC 2002*. Taormina, Italy: [s.n.], 2002. p. 530–535.
- [54] SOHN, J.; KIM, N. S.; SUNG, W. A Statistical Model-Based Voice Activity Detection. *IEEE Signal Processing Letters*, v. 6, n. 1, p. 1–3, January 1999.
- [55] SOON, I. Y.; KOH, S. N.; YEO, C. K. Improved noise suppression filter using self-adaptive estimator of probability of speech absence. *Signal Processing*, v. 75, n. 2, p. 151–159, June 1999.
- [56] EPHRAIM, Y.; MALAH, D. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. *IEEE Trans. Acoust., Speech, Signal Processing*, v. 32, p. 1109–1121, December 1984.
- [57] CHO, Y. D.; AL-NAIMI, K.; KONDOZ, A. Improved voice activity detection based on a smoothed statistical likelihood ratio. *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, v. 2, p. 737–740, May 2001.
- [58] GUEDES, K. B. *Simulação Eficiente do Sistema Elétrico de Potência Utilizando Filtragem Digital Multitaxa*. 110 p. Tese (Doutorado) — Pontifícia Universidade Católica do Rio de Janeiro, PUC-RJ, Rio de Janeiro, Brasil, Setembro 2002.
- [59] PINTO, D. F. *Uso de Banco de Filtros em Separação Cega de Fontes*. 67 p. Dissertação (Mestrado) — Universidade Federal do Rio de Janeiro, COPPE, Rio de Janeiro, Brasil, Outubro 2012.
- [60] DINIZ, P. S. R. *Adaptive Filtering: Algorithms and Practical Implementation*. [S.l.]: Kluwer Academic Publishers, 1997. 443 p.

- [61] KOILPILLAI, R. D.; VAIDYANATHAN, P. P. Cosine-Modulated FIR Filter Banks Satisfying Perfect Reconstruction. *IEEE Transactions on Signal Processing*, v. 40, n. 4, p. 770–783, April 1992.
- [62] NGUYEN, T. Q. Digital Filter Banks Design Quadratic-Constrained Formulation. *IEEE Transactions on Signal Processing*, v. 43, n. 9, p. 2103–2108, September 1995.
- [63] Disponível em: <<http://sassec.gforge.inria.fr/>>. Acesso em: 15 de Fevereiro de 2013.
- [64] LEHMAN, 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, p. 269–277, June 2008.
- [65] WEN, J. et al. Evaluation of speech dereverberation algorithms using the MARDY database. In: *Proc. Intl. Workshop Acoust. Echo Noise Control - (IWAENC)*. Paris, France: [s.n.], 2006. p. 1–4.
- [66] HARVILLE, D. A. *Matrix Algebra From a Statistician's Perspective*. [S.l.]: Springer, 1997. 630 p.