



SISTEMA DE LOCALIZAÇÃO DE MÚLTIPLAS FONTES SONORAS EM
RECINTOS FECHADOS BASEADO EM ALGORITMOS DE SEPARAÇÃO
CEGA DE FONTES

Luiz Carlos Figueira Nogueira

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

Orientadora: Mariane Rembold Petraglia

Rio de Janeiro
Abril de 2015

SISTEMA DE LOCALIZAÇÃO DE MÚLTIPLAS FONTES SONORAS EM
RECINTOS FECHADOS BASEADO EM ALGORITMOS DE SEPARAÇÃO
CEGA DE FONTES

Luiz Carlos Figueira Nogueira

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

Examinada por:

Prof^a. Mariane Rembold Petraglia, Ph.D.

Prof. Amaro Azevedo de Lima, Ph.D.

Prof. José Gabriel Rodríguez Carneiro Gomes, Ph.D.

Prof. Jules Ghislain Slama, D.Sc.

Prof. Julio Cesar Boscher Torres, D.Sc.

Prof. Tadeu Nagashima Ferreira, D.Sc.

RIO DE JANEIRO, RJ – BRASIL

ABRIL DE 2015

Nogueira, Luiz Carlos Figueira

Sistema de localização de múltiplas fontes sonoras em recintos fechados baseado em algoritmos de separação cega de fontes/Luiz Carlos Figueira Nogueira. – Rio de Janeiro: UFRJ/COPPE, 2015.

XX, 127 p.: il.; 29,7cm.

Orientadora: Mariane Rembold Petraglia

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2015.

Referências Bibliográficas: p. 121 – 127.

1. localização de fontes sonoras. 2. separação cega de fontes. 3. estimação de direção de chegada. I. Petraglia, Mariane Rembold. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

À minha família e amigos.

Agradecimentos

Agradeço à minha orientadora por todos incentivos e contribuições técnicas, sempre com ampla competência e gentileza em todo o período de pesquisa, a qual se tornou mais um grande espelho na minha vida acadêmica.

Agradeço à minha família e amigos por todos os incentivos e paciência empregados em função dos meus momentos de pesquisa.

Agradeço aos meus colegas de trabalho e amigos que sempre colaboraram para o meu empenho na pesquisa.

Agradeço ao CEFET/RJ por colaborar com minhas pesquisas.

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

SISTEMA DE LOCALIZAÇÃO DE MÚLTIPLAS FONTES SONORAS EM
RECINTOS FECHADOS BASEADO EM ALGORITMOS DE SEPARAÇÃO
CEGA DE FONTES

Luiz Carlos Figueira Nogueira

Abril/2015

Orientadora: Mariane Rembold Petraglia

Programa: Engenharia Elétrica

Investiga-se, nesta tese, um sistema de localização de múltiplas fontes sonoras baseado no encontro de retas correspondentes às direções de chegadas (DOAs) dos sinais em dois conjuntos de microfones. As estimativas das DOAs são obtidas a partir de parâmetros fornecidos por algoritmos de separação cega de fontes (BSS). Modificações nos algoritmos BSS convencionais são propostas visando a obtenção de estimativas mais acuradas das DOAs em um menor intervalo de tempo. As condições necessárias para tornar o sistema robusto são investigadas, resultando em um procedimento para identificar a região em que a localização pode ser considerada correta e para determinar as posições ótimas dos arranjos de microfones. Esse procedimento utiliza informações da acurácia das estimativas das DOAs e do erro máximo admitido nas estimativas das posições das fontes. O sistema de localização proposto é avaliado por simulações em recintos com baixa e alta reverberações e em experimentos reais com dados medidos em uma sala de aula. O método de análise proposto pode ser estendido a outras técnicas tradicionais de localização, viabilizando as reduções da quantidade de microfones utilizados e do tempo de processamento. A pesquisa desenvolvida encontra aplicação no projeto de ambientes interativos inteligentes, como salas de aula virtual e de vídeo-conferência, e em outros sistemas de comunicação e de robótica nos quais a localização de fontes acústicas é fundamental.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

MULTIPLE SOURCE LOCALIZATION BASED ON BLIND SOURCE SEPARATION ALGORITHMS

Luiz Carlos Figueira Nogueira

April/2015

Advisor: Mariane Rembold Petraglia

Department: Electrical Engineering

This thesis investigates a system of multiple sound sources localization based on the encounter of the lines corresponding to the directions of arrival (DOAs) of the sources signals into a pair of microphone arrays. The DOAs estimates are derived from parameters provided by blind source separation (BSS) algorithms. Improvements in the conventional BSS algorithms are proposed with the purpose of producing more accurate DOA estimates in a shorter time interval. The necessary conditions for making the system robust are investigated, resulting in a procedure to identify the region in which the location can be considered correct and to determine the optimal positions of the microphone arrays. This procedure uses informations regarding the accuracy of the DOA estimates and the maximum permissible error in the estimates of the source positions. The proposed localization system is evaluated by simulations in low and high reverberation rooms and by real experiments with data measured in a classroom. The proposed analysis method can be extended to other traditional localization techniques, enabling a reduction in the number of microphones and processing time. This research finds application in the design of intelligent interactive environments, such as virtual classrooms and videoconferencing rooms, and other communications and robotics systems in which the location of sound sources is essential.

Sumário

Lista de Figuras	xi
Lista de Tabelas	xvii
Lista de Símbolos	xviii
Lista de Abreviaturas	xix
1 Introdução	1
1.1 Objetivos da Pesquisa	2
1.2 Revisão Bibliográfica	4
1.3 Estrutura da Tese	7
2 Misturas de Sinais Acústicos	8
2.1 Canal de Propagação Acústica	9
2.2 Simulação do Canal	10
2.3 Decaimento de Energia da RIR	13
2.4 Considerações Finais sobre as Misturas	14
3 Algoritmos de Separação Cega de Fontes	15
3.1 Separação de Misturas Instantâneas	15
3.1.1 Pré-processamento	16
3.1.2 Análise de Componentes Independentes	16
3.2 Separação de Misturas Convolutivas	18
3.2.1 Separação no Domínio da Frequência	19
3.2.2 Separação no Domínio do Tempo	21
3.3 Avaliação do Sistema de Separação	24
4 Algoritmos de Estimação de Direção de Chegada	27
4.1 Estimativa da DOA com Uma Fonte Ativa	28
4.2 Estimativa da DOA com Múltiplas Fontes Ativas	31
4.2.1 Método do Campo Distante	31
4.2.2 Método SCT	32

4.2.3	Método TRINICON	34
4.3	Modificações nos Métodos de BSS para Utilização na Estimação das DOAs	35
4.3.1	Algoritmo BSS no Domínio da Frequência	35
4.3.2	Algoritmo TRINICON	36
5	Sistema de Localização de Fontes e Determinação da Região de Localização Robusta	39
5.1	Fundamentos Teóricos para Estimar a Região de Localização Robusta	40
5.2	Estimação da Região Adequada em Ambientes Simulados	46
5.2.1	Ambiente Simulado	46
5.2.2	Região Adequada para uma Fonte em Ambiente com Baixa Reverberação	47
5.2.3	Região Adequada em Ambiente com Alta Reverberação	51
5.3	Estimação da Região Adequada em Ambiente Real	55
5.3.1	Descrição do Ambiente Real	56
5.3.2	Região Adequada para o Ambiente Real	58
6	Resultados da Localização para Uma Fonte Ativa	62
6.1	Localização de uma Fonte em Ambiente Simulado com Baixa Reverberação	62
6.2	Localização de uma Fonte em Ambiente Simulado com Alta Reverberação	66
6.3	Influência da Distância entre Arranjos e da Reverberação na Localização	69
6.4	Localização de Uma Fonte no Ambiente Real	70
6.5	Considerações Finais sobre a Localização de Uma Fonte	75
7	Resultados da Localização de Duas Fontes Ativas	77
7.1	Método do Campo Distante	77
7.1.1	Ambiente Simulado com Baixa Reverberação	78
7.1.2	Ambiente Simulado com Alta Reverberação	81
7.1.3	Ambiente Real	84
7.1.4	Considerações Finais sobre Localização com o Método do Campo Distante	87
7.2	Método SCT	88
7.2.1	Ambiente Simulado com Baixa Reverberação	88
7.2.2	Ambiente Simulado com Alta Reverberação	91
7.2.3	Ambiente Real	94
7.2.4	Considerações Finais sobre a Localização com o Método SCT	97
7.3	Método TRINICON	97

7.3.1	Ambiente Simulado com Baixa Reverberação	98
7.3.2	Ambiente Simulado com Alta Reverberação	100
7.3.3	Ambiente Real	103
7.3.4	Considerações Finais sobre a Localização com o Método TRI- NICON	106
7.4	Comparações entre Métodos de Localização de Duas Fontes	107
7.4.1	Ambientes Simulados	107
7.4.2	Ambiente Real	111
7.5	Emprego das Modificações Propostas nos Algoritmos de BSS	113
7.5.1	Método SCT	113
7.5.2	Método TRINICON	116
8	Conclusões e Trabalhos Futuros	118
	Referências Bibliográficas	121

Lista de Figuras

2.1	Diagrama dos processos de mistura e separação.	9
2.2	Ilustração simplificada do canal de áudio.	10
2.3	Modelo “Image-Source” simplificado.	11
3.1	Mistura convolutiva com duas fontes e dois sensores.	18
3.2	Etapas do processo de separação de misturas convolutivas.	20
3.3	Sistema de mistura e separação com os parâmetros utilizados no método TRINICON.	22
3.4	Ilustração da decomposição, num contexto de misturas instantâneas, do vetor (a) \mathbf{y}_1 ; (b) \mathbf{y}_2 , supondo três amostras consecutivas das estimativas das fontes.	26
4.1	Sistema para estimação da posição com 2 pares de microfones.	28
4.2	Frente de onda na condição de campo distante.	29
4.3	Exemplo ilustrativo do desempenho do algoritmo GCC-PHAT. (a) trecho de sinal de voz e sua versão atrasada de 14 amostras; (b) correlação cruzada dos sinais; (c) GCC-PHAT.	30
4.4	Modelo de campo distante.	31
4.5	Exemplo de espectrograma de sinal de voz.	36
5.1	Sistema de localização baseado em BSS.	40
5.2	Elementos do sistema de localização.	41
5.3	Estimativa da região de localização apropriada para (a) $d_{PAR} = 6$ m e (b) $d_{PAR} = 3$ m, sendo $\Delta\theta = 4^\circ$ e $D_{MAX} = 1$ m.	43
5.4	Área apropriada percentual, em escala de cinza, para diferentes valores de d_{PAR} e $\Delta\theta$, com (a) $D_{MAX} = 1$ m e (c) $D_{MAX} = 0.5$ m. Valores médios das áreas apropriadas, R_m , para diferentes valores de d_{PAR} , com (b) $D_{MAX} = 1$ m e (d) $D_{MAX} = 0.5$ m.	45
5.5	Posicionamento dos microfones e fontes no ambiente dos testes simulados.	47
5.6	Taxa de acerto dado o erro máximo admitido.	48

5.7	Valor médio da área apropriada ao variar $\Delta\theta$ entre 0.1° e 1.5° , para d_{PAR} entre 1 m e 10 m e D_{MAX} entre 0.2 m e 0.5 m.	49
5.8	Valores médios da área apropriada para diferentes erros de localização D_{MAX}	50
5.9	Região apropriada para $d_{PAR} = 7$ m (Caso A) com $D_{MAX} \leq 0.25$ m, e $\Delta\theta$ igual a (a) 0.5° , (b) 1° e (c) 1.5°	51
5.10	Região apropriada para $d_{PAR} = 1$ m (Caso B) com $D_{MAX} \leq 0.25$ m, e $\Delta\theta$ igual a (a) 0.5° , (b) 1° e (c) 1.5°	52
5.11	Valor médio da área apropriada ao variar $\Delta\theta$ entre 1° e 7° , para d_{PAR} entre 1 m e 10 m e D_{MAX} entre 0.6 m e 1.4 m.	53
5.12	Valores médios da área apropriada para diferentes erros de localização D_{MAX}	53
5.13	Região apropriada para $d_{PAR} = 7$ m (Caso A) com $D_{MAX} \leq 1.2$ m, e $\Delta\theta$ igual a (a) 3° , (b) 5° e (c) 7°	54
5.14	Região apropriada para $d_{PAR} = 1$ m (Caso B) com $D_{MAX} \leq 1.2$ m, e $\Delta\theta$ igual a (a) 3° , (b) 5° e (c) 7°	55
5.15	Fotografias da sala de aula utilizada nos testes.	56
5.16	Posicionamento dos microfones e fontes nos testes em ambiente real.	57
5.17	Valor médio da área apropriada ao variar $\Delta\theta$ entre 1° e 9°	58
5.18	Valores médios da área apropriada para diferentes erros de localização D_{MAX}	59
5.19	Área considerada apropriada para a localização com $D_{MAX} = 1.2$ m, $d_{PAR} = 5$ m (Caso A) e $\Delta\theta$ igual a (a) 5° , (b) 7° , (c) 9° e (d) 11°	60
5.20	Região apropriada para $d_{PAR} = 2$ m (Caso B) com $D_{MAX} \leq 1.2$ m, e $\Delta\theta$ igual a (a) 2° , (b) 5° e (c) 7°	61
6.1	Taxa de sucesso nas estimativas das DOAs em cada possível posição da fonte considerando $\Delta\theta = 1^\circ$ para (a) Caso 1 e (b) Caso 2.	63
6.2	Taxa de sucesso média ao variar o erro máximo das estimativas das DOAs $\Delta\theta$ admitido.	64
6.3	Taxa de sucesso da localização em cada possível posição da fonte considerando erro máximo admitido $D_{MAX} = 25$ cm para (a) Caso A e (b) Caso (B).	64
6.4	Taxa de sucesso média da localização ao variar o erro máximo de localização admitido.	65
6.5	Taxa de sucesso nas estimativas das DOAs em cada possível posição da fonte para $T_{60} = 800$ ms, considerando (a) Caso 1 com $\Delta\theta = 3^\circ$ e (b) Caso 2 para $\Delta\theta = 1^\circ$	67

6.6	Taxa de sucesso média ao variar o erro máximo das estimativas das DOAs $\Delta\theta$ admitido para $T_{60} = 800$ ms.	67
6.7	Taxa de sucesso da localização em cada possível posição da fonte, com $T_{60} = 800$ ms, considerando erro máximo admitido $D_{MAX} = 1.2$ m para (a) Caso A e (b) Caso (B).	68
6.8	Taxa de sucesso média da localização ao variar o erro máximo de localização admitido para $T_{60} = 800$ ms.	69
6.9	Taxa de sucesso média na localização de uma fonte ao variar D_{MAX} para o ambiente simulado com T_{60} igual a (a) 200, (b) 400, (c) 600 e (d) 800 ms.	70
6.10	Taxa de sucesso nas estimativas das DOAs em cada possível posição da fonte considerando $\Delta\theta = 7^\circ$ para (a) Caso 1 e (b) Caso 2, e $\Delta\theta = 2^\circ$ para (c) Caso 3 e (d) Caso 4.	72
6.11	Taxa de sucesso média nas estimativas das DOAs ao variar o erro máximo admitido para (a) Caso 1, (b) Caso 2, (c) Caso 3 e (d) Caso 4.	73
6.12	Taxa de sucesso da localização em cada possível posição da fonte considerando $D_{MAX} = 1.2$ m para (a) Caso A e (b) Caso B.	74
6.13	Taxa de sucesso média da localização ao variar D_{MAX} para o ambiente real.	75
7.1	Taxa de sucesso nas estimativas das DOAs usando o método do Campo Distante em cada possível posição da fonte para $T_{60} = 200$ ms, considerando (a) Caso 1 com $\Delta\theta = 5^\circ$ e (b) Caso 2 para $\Delta\theta = 2^\circ$	79
7.2	Taxa de sucesso média ao variar $\Delta\theta$, com $T_{60} = 200$ ms e utilizando o método do Campo Distante.	80
7.3	Taxa de sucesso da localização em cada possível posição da fonte, com $T_{60} = 200$ ms e usando o método do Campo Distante, considerando $D_{MAX} = 1.2$ m para (a) Caso A e (b) Caso (B).	80
7.4	Taxa de sucesso média da localização ao variar D_{MAX} no ambiente simulado com $T_{60} = 200$ ms usando o método do Campo Distante.	80
7.5	Taxa de sucesso nas estimativas das DOAs usando o método do Campo Distante em cada possível posição da fonte para $T_{60} = 800$ ms e $\Delta\theta = 7^\circ$, considerando (a) Caso 1 e (b) Caso 2.	82
7.6	Taxa de sucesso média ao variar $\Delta\theta$, com $T_{60} = 800$ ms e utilizando o método do Campo Distante.	82
7.7	Taxa de sucesso da localização em cada possível posição da fonte, com $T_{60} = 800$ ms e usando o método do Campo Distante, considerando $D_{MAX} = 1.2$ m para (a) Caso A e (b) Caso (B).	83

7.8	Taxa de sucesso média da localização ao variar D_{MAX} no ambiente simulado com $T_{60} = 800$ ms usando o método do Campo Distante. . .	83
7.9	Taxa de sucesso nas estimativas das DOAs em cada possível posição da fonte considerando $\Delta\theta = 7^\circ$ para (a) Caso 1 e (b) Caso 2, e $\Delta\theta = 5^\circ$ para (c) Caso 3 e (d) Caso 4.	85
7.10	Taxa de sucesso média nas estimativas das DOAs ao variar $\Delta\theta$ para (a) Caso 1, (b) Caso 2, (c) Caso 3 e (d) Caso 4.	86
7.11	Taxa de sucesso da localização em cada possível posição da fonte, usando o método do Campo Distante no Ambiente Real, considerando $D_{MAX} = 1.2$ m para (a) Caso A e (b) Caso B.	86
7.12	Taxa de sucesso média para a localização ao variar D_{MAX} para o ambiente real usando o método do Campo Distante.	87
7.13	Taxa de sucesso nas estimativas das DOAs usando o método SCT em cada possível posição da fonte para $T_{60} = 200$ ms, considerando (a) Caso 1 com $\Delta\theta = 5^\circ$ e (b) Caso 2 com $\Delta\theta = 2^\circ$	89
7.14	Taxa de sucesso média ao variar $\Delta\theta$, com $T_{60} = 200$ ms e utilizando o método SCT.	89
7.15	Taxa de sucesso da localização em cada possível posição da fonte, com $T_{60} = 200$ ms e usando o método SCT, considerando $D_{MAX} = 1.2$ m para (a) Caso A e (b) Caso (B).	90
7.16	Taxa de sucesso média da localização ao variar D_{MAX} no ambiente simulado com $T_{60} = 200$ ms usando o método SCT.	91
7.17	Taxa de sucesso nas estimativas das DOAs em cada posição, usando o método SCT, para $T_{60} = 800$ ms e $\Delta\theta = 7^\circ$, considerando (a) Caso 1 e (b) Caso 2.	92
7.18	Taxa de sucesso média ao variar $\Delta\theta$ entre 1° e 20° , com $T_{60} = 800$ ms usando o método SCT.	92
7.19	Taxa de sucesso para a localização em cada posição no ambiente simulado com T_{60} igual a 800 ms usando o método método SCT.	93
7.20	Taxa de sucesso média da localização ao variar D_{MAX} no ambiente simulado com $T_{60} = 800$ ms usando o método SCT.	93
7.21	Taxa de sucesso nas estimativas das DOAs em cada possível posição da fonte no ambiente real, usando o método SCT e considerando $\Delta\theta = 7^\circ$ para (a) Caso 1 e (b) Caso 2, e $\Delta\theta = 2^\circ$ para (c) Caso 3 e (d) Caso 4.	94
7.22	Taxa de sucesso média ao variar $\Delta\theta$, no ambiente real e usando o método SCT, para (a) Caso 1, (b) Caso 2, (c) Caso 3 e (d) Caso 4. . .	95

7.23	Taxa de sucesso da localização em cada possível posição da fonte no ambiente real, usando o algoritmo do SCT e considerando $D_{MAX} = 1.2$ m, para (a) Caso A e (b) Caso B.	96
7.24	Taxa de sucesso média da localização ao variar D_{MAX} no ambiente real usando o método SCT.	97
7.25	Taxa de sucesso nas estimativas das DOAs usando o método TRINICON em cada possível posição da fonte para $T_{60} = 200$ ms, considerando (a) Caso 1 com $\Delta\theta = 5^\circ$ e (b) Caso 2 para $\Delta\theta = 2^\circ$	98
7.26	Taxa de sucesso média ao variar $\Delta\theta$, com $T_{60} = 200$ ms e utilizando o método TRINICON.	99
7.27	Taxa de sucesso da localização, com $T_{60} = 200$ ms e usando o algoritmo do TRINICON, considerando $D_{MAX} = 1.2$ m para (a) Caso A e (b) Caso (B).	99
7.28	Taxa de acerto média da localização ao variar D_{MAX} no ambiente simulado com $T_{60} = 200$ ms usando o método TRINICON.	100
7.29	Taxa de sucesso nas estimativas das DOAs usando o método TRINICON em cada possível posição da fonte para $T_{60} = 800$ ms, considerando (a) Caso 1 com $\Delta\theta = 5^\circ$ e (b) Caso 2 para $\Delta\theta = 2^\circ$	101
7.30	Taxa de acerto média ao variar $\Delta\theta$ entre 1° e 20° , com $T_{60} = 800$ ms usando o método TRINICON.	102
7.31	Taxa de sucesso da localização no ambiente simulado com $T_{60} = 800$ ms usando o método TRINICON, considerando $D_{MAX} = 1.2$ m para (a) Caso A e (b) Caso (B).	102
7.32	Taxa de acerto média da localização ao variar D_{MAX} no ambiente simulado com $T_{60} = 800$ ms usando o método TRINICON.	103
7.33	Taxa de sucesso nas estimativas das DOAs em cada possível posição da fonte, usando o método TRINICON no ambiente real, considerando $\Delta\theta = 7^\circ$ para (a) Caso 1 e (b) Caso 2, e $\Delta\theta = 5^\circ$ para (c) Caso 3 e (d) Caso 4.	104
7.34	Taxa de sucesso ao variar $\Delta\theta$ para (a) Caso 1, (b) Caso 2, (c) Caso 3 e (d) Caso 4.	105
7.35	Taxa de sucesso da localização em cada possível posição da fonte, usando o método TRINICON no ambiente real, considerando $D_{MAX} = 1.2$ m para (a) Caso A e (b) Caso B.	105
7.36	Taxa de acerto média da localização ao variar D_{MAX} no ambiente real usando o método TRINICON.	106
7.37	Taxa de sucesso média ao variar $\Delta\theta$ obtidas com os diversos métodos de estimação das DOAs para o Caso 1 nos ambientes simulados.	108

7.38	Taxa de sucesso média ao variar D_{MAX} obtidas com os diversos métodos de estimação das DOAs para o Caso A nos ambientes simulados.	109
7.39	Taxa de sucesso média da localização ao variar D_{MAX} no ambiente simulado para T_{60} entre 200 ms e 900 ms, usando o método SCT. . .	110
7.40	Taxa de sucesso da localização em cada possível posição da fonte usando o método SCT e considerando $D_{MAX} = 1.2$ m para (a) $T_{60} = 300$ ms, (b) $T_{60} = 400$ ms, (c) $T_{60} = 500$ ms e (d) $T_{60} = 600$ ms. . .	111
7.41	Taxa de sucesso média ao variar $\Delta\theta$ obtidas com os diversos métodos de estimação das DOAs para os Casos 1 (em azul) e 2 (em vermelho) no ambiente real.	112
7.42	Taxa de sucesso média das estimativas das posições ao variar D_{MAX} para os diversos algoritmos no Caso A em ambiente real.	113
7.43	Taxa de sucesso média ao variar $\Delta\theta$ do método SCT com N_p raias de frequência e diferentes comprimentos dos sinais no ambiente simulado com $T_{60} = 200$ ms.	114
7.44	Taxa de sucesso média ao variar $\Delta\theta$ do método SCT com N_p raias de frequência e diferentes comprimentos dos sinais no ambiente simulado com $T_{60} = 800$ ms.	115
7.45	Taxa de sucesso média ao variar $\Delta\theta$ do método TRINICON no ambiente simulado com $T_{60} = 200$ ms, para diferenças de DOAs das fontes entre 45° e 135°	116
7.46	Taxa de sucesso média ao variar $\Delta\theta$ do método TRINICON no ambiente simulado com $T_{60} = 800$ ms, para diferenças de DOAs das fontes entre 45° e 135°	117

Lista de Tabelas

6.1	Taxa de sucesso médio das estimativas das posições no ambiente real com 1 fonte.	75
-----	--	----

Lista de Símbolos

D_E	erro absoluto médio das posições estimadas, p. 43
D_{MAX}	erro máximo admitido na localização, p. 43
L	comprimento do filtro h_{ij} , p. 9
M	número de sensores, p. 9
N	número de fontes, p. 9
N_p	quantidade de raias de maior potência, p. 115
P_F	posição da fonte, p. 42
P_{E_i}	posição estimada da fonte, p. 42
P_{E_n}	posição estimada da fonte, p. 42
θ_{A_R}	ângulo correspondente a DOA real nos conjuntos de sensores M_A , p. 42
θ_{A_n}	ângulo correspondente a DOA estimada nos conjuntos de sensores M_A , p. 42
θ_{B_R}	ângulo correspondente a DOA real nos conjuntos de sensores M_B , p. 42
θ_{B_n}	ângulo correspondente a DOA estimada nos conjuntos de sensores M_B , p. 42
d_{PAR}	distância entre os conjuntos de sensores, p. 42
$h_{ij}(n)$	resposta ao impulso do canal entre a fonte i e o sensor j , p. 9
$s_i(n)$	sinal da i -ésima fonte, p. 9
$x_j(n)$	sinal da i -ésima mistura, p. 9

Lista de Abreviaturas

BSI	<i>Blind System Identification</i> , p. 6
BSS	<i>Blind Source Separation</i> , p. 2
CD	Campo Distante, p. 32
CMM	<i>Cepstral Mean Normalization</i> , p. 7
DOA	<i>Direction of Arrival</i> , p. 1
EDC	<i>Energy Decay Curve</i> , p. 14
FIR	<i>Finite Impulse Response</i> , p. 24
GCC-PHAT	<i>Generalized Cross Correlation with Phase Transform</i> , p. 3
GCC	<i>Generalized Cross Correlation</i> , p. 4
GMM	<i>Gaussian Mixture Model</i> , p. 7
ICA	<i>Independent Component Analysis</i> , p. 2
ILD	<i>Interaural Level Difference</i> , p. 7
INFOMAX	<i>INFORmation MAXimization</i> , p. 5
ISM	<i>Image-Source Model</i> , p. 6
ITD	<i>Interaural Time Difference</i> , p. 4
JADE	<i>Joint Approximate Diagonalization of Eigen-matrices</i> , p. 5
MCBD	<i>Multichannel Blind Deconvolution</i> , p. 23
MDP	<i>Minimal Distortion Principle</i> , p. 22
MIMO	<i>Multiple-Input Multiple-Output</i> , p. 6
PCA	<i>Principal Component Analysis</i> , p. 2

RIR	<i>Room Impulse Response</i> , p. 11
SAR	<i>Source to Artifacts Ratio</i> , p. 25
SCT	<i>State Coherence Transform</i> , p. 5
SDR	<i>Source to Distortion Ratio</i> , p. 25
SIR	<i>Source to Interferences Ratio</i> , p. 25
SRP-PHAT	<i>Steered Response Power using PHAse Transform</i> , p. 6
STFT	<i>Short Time Fourier Transform</i> , p. 2
TDOA	<i>Time Difference of Arrival</i> , p. 3
TRINICON	<i>TRIPLE-N Independent component analysis for CONVolutive mixtures</i> , p. 6
cSCT	<i>Cumulated SCT</i> , p. 35

Capítulo 1

Introdução

O desenvolvimento de algoritmos que utilizam sinais capturados por múltiplos sensores para estimar as posições de fontes é uma linha de pesquisa com destaque na área de processamento de sinais [1] com diversas aplicações, tais como no rastreamento de sinais de fontes acústicas, de telefonia móvel, de radares e sonares, permitindo a melhoria da qualidade desses sinais ou a localização de usuários ou intrusos. A partir dessas estimativas, é possível projetar e otimizar filtros espaciais (“beamformers”) para remoção de ruídos e interferências.

Na era atual de informação e computação, uma comunicação mais natural entre homem e máquina é desejada em muitos sistemas inteligentes. Conseqüentemente, a habilidade de localizar e rastrear sinais de voz em uma sala se torna essencial. A operação de equipamentos por comando de voz, a seleção automática de locutores ativos em sistemas de teleconferências e o uso de sistemas viva-voz (“hands-free”) em telefonia móvel são exemplos de aplicações que se beneficiam com o crescente desenvolvimento da área de processamento de sinais multi-sensores.

Em particular, cabe ao sistema de localização fornecer estimativas confiáveis das posições das fontes acústicas, onde a robustez das estimativas dependerá de uma série de fatores, tais como distribuição dos microfones no ambiente, posições das fontes, características acústicas do ambiente, método utilizado para determinar a direção de chegada (DOA, do inglês “Direction of Arrival”) de cada sinal em cada arranjo de sensores, método utilizado para separar as fontes, entre outros. Em um ambiente real, onde há ruído, diversas fontes e reverberação, a tarefa de localização se torna mais complexa, gerando desafios para a derivação de novos métodos ou a otimização dos já existentes.

No caso de múltiplas fontes, técnicas de separação cega de fontes (BSS, do inglês “Blind Source Separation”) [2]-[4] permitem fazer estimativas de seus sinais através de combinações dos mesmos, obtidas de um conjunto de sensores, utilizando a informação de diversidade espacial presente nas misturas. Os casos mais simples, de misturas instantâneas com fontes independentes e quantidade de fontes igual ou

menor que a de sensores, podem ser resolvidos através da análise de componentes independentes (ICA, do inglês “Independent Component Analysis”) [5]. Nos casos mais complexos, onde ocorrem misturas convolutivas ou a quantidade de fontes ativas é maior que a de sensores, há um aumento na complexidade das técnicas de separação, sendo necessária a utilização de um conjunto de ferramentas adicionais, com parâmetros específicos, como a STFT (do inglês “Short Time Fourier Transform”) e/ou a PCA (do inglês “Principal Component Analysis”), sendo em geral a separação feita no domínio da frequência, surgindo ainda os problemas da permutação e do escalamento das estimativas das fontes nas diferentes raias de frequência.

1.1 Objetivos da Pesquisa

O objetivo principal deste trabalho é elaborar um sistema de localização de múltiplas fontes sonoras em recintos fechados e derivar um procedimento para otimizar seu desempenho e avaliar a viabilidade da sua utilização, dadas as características do recinto e a expectativa do erro nas estimativas das DOAs. O sistema proposto é baseado nas estimativas das direções de chegada de cada sinal em dois arranjos de microfones, apresentando baixo custo quando comparado a outros sistemas que utilizam um grande número de microfones espalhados no recinto [6]-[8]. A aplicação principal do sistema de localização proposto é em salas de teleconferências, como as utilizadas para aulas à distância, com o objetivo de separar e/ou melhorar a qualidade dos sinais de voz de locutores provenientes de diferentes direções e auxiliar no posicionamento automático de câmera(s) para focalização no(s) locutor(es).

O modelo de análise proposto para viabilizar e/ou otimizar a localização utiliza a acurácia das estimativas das DOAs, o erro máximo admitido nas estimativas das posições e a distância entre os arranjos de microfones, para avaliar a região de localização robusta resultante do emprego do princípio do encontro das DOAs dos sinais das fontes nos dois arranjos. Com base nesse modelo, podem-se estabelecer diretrizes para orientar no tratamento acústico do ambiente e no posicionamento otimizado dos microfones, de forma a garantir que a localização seja bem sucedida.

Em [1] a localização de uma única fonte ativa é investigada, sendo utilizado o algoritmo GCC-PHAT (do inglês “Generalized Cross Correlation with Phase Transform”) [9] para estimar a direção de chegada do sinal em arranjos de microfones, e o princípio do encontro das DOAs para estimar a posição da fonte. É verificado que a localização é bem sucedida em ambientes com baixa reverberação (com $T_{60} = 200$ ms), não sendo avaliado o seu comportamento em outros ambientes. Além de não abordar casos de múltiplas fontes, o referido trabalho não avalia a acurácia da localização em relação à posição da fonte no ambiente e o posicionamento otimizado

dos microfones.

A utilização de algoritmos baseados em separação cega de fontes para estimação da direção de chegada foi tema de pesquisas recentes [5],[10]-[17]. Esses trabalhos avaliam apenas o desempenho dos algoritmos propostos nas estimativas das DOAs para ambientes de baixa reverberação. Em nenhum deles é investigada a acurácia da localização resultante do emprego de apenas dois conjuntos de sensores nas diferentes posições do ambiente e a otimização do posicionamento dos microfones.

O modelo proposto para viabilizar e/ou otimizar a localização é válido para sistemas de localização desenvolvidos para uma ou mais fontes, podendo ser aplicado:

- nos sistemas de localização baseados no algoritmo GCC-PHAT [1],[18],[19], fornecendo diretrizes para posicionar os microfones e indicando a área apropriada para a localização;
- como forma alternativa aos métodos propostos em [6],[20], auxiliando na escolha da quantidade de microfones e no seu posicionamento;
- no sistema proposto em [8],[21] para localizar e estimar a orientação da fonte (direção em que ela está apontada), auxiliando no posicionamento e determinação da quantidade de microfones e na delimitação da região de atuação dos arranjos;
- em [22], onde o autor utiliza sensores no centro de uma mesa e busca a localização e orientação das fontes, indicando a precisão da localização e seu alcance no entorno da mesa;
- como forma alternativa ao método proposto em [23],[24], onde a localização é feita utilizando TDOA (do inglês “Time Difference of Arrival”) nos microfones;
- nos processos de rastreamento de sinais, como os utilizados em [25],[26];
- na localização de fontes sonoras através de uma rede de sensores [27],[28], auxiliando na seleção dos microfones (ou arranjos).

Em [7] é proposta uma forma de estimar o erro de localização, cuja finalidade é semelhante à apresentada neste trabalho. Porém, o modelo de estimação de erro proposto nesta tese é mais abrangente, permitindo outras análises, como a da região de funcionamento apropriado do sistema e da distância ideal entre os arranjos de microfones.

O modelo proposto foi avaliado tanto em condições favoráveis, considerando um ambiente simulado de baixa reverberação com T_{60} de 200 ms, como em condições desafiadoras, em ambiente simulado com T_{60} de 800 ms e em ambiente real com T_{60} medido de 900 ms, para localização de uma única fonte e de duas fontes simultâneas.

As contribuições deste trabalho encontram aplicação em sistemas nos quais a determinação da posição de uma ou mais fontes acústicas é uma função importante, como em equipamentos de teleconferência, multimídia e robótica, permitindo a melhoria da qualidade da comunicação e da interação homem/máquina. Em particular, a aplicação analisada nesta tese é a de salas de teleconferência inteligentes, como as usadas em ensino a distância, avaliando a acurácia da localização, estabelecendo diretrizes que viabilizam e otimizam o desempenho do sistema e a região apropriada para sua operação.

1.2 Revisão Bibliográfica

A localização de uma ou mais fontes é o tema de várias pesquisas [8],[18],[19],[22],[26],[29]-[32]. As técnicas propostas na literatura para localização de fontes acústicas abordam o problema de diferentes formas.

Em muitas linhas de pesquisa, busca-se estimar a TDOA/DOA do sinal de cada fonte em um conjunto de sensores através da correlação cruzada generalizada (GCC, do inglês “Generalized Cross-Correlation”) [9],[33]. Nessa abordagem, descrita no Cap. 4, busca-se o pico dominante da função correlação cruzada generalizada dos sinais das misturas. Diversas técnicas de redução de picos espúrios, causados pela reverberação em ambientes fechados, pelo espectro colorido dos sinais de voz e por ruídos, foram propostas na literatura, destacando-se o uso da transformada de fase, resultando no método GCC-PHAT [9]. Outros métodos de estimação da TDOA/DOA obtêm a diferença de tempo entre sensores (ITD, do inglês “Interaural Time Difference”) por meio do casamento de atrasos usando sistemas adaptativos [30],[34]-[36]. Existem técnicas que buscam a posição das fontes através de outros parâmetros que não a DOA, como a energia com que os sinais atingem os sensores espalhados pelo ambiente [37]. No entanto, esses métodos em geral são mais sensíveis à presença de ruído no ambiente.

A localização de uma fonte a partir do cálculo da TDOA/DOA pode ser feita através da distribuição de dois pares de sensores no ambiente, como proposto em [18],[19],[29], sendo a estimativa da posição da fonte determinada pelo encontro das retas obtidas das direções de chegada do sinal em cada sensor. Nesses trabalhos as estimativas dos TDOAs foram obtidas pelo algoritmo GCC-PHAT, tendo apresentado bons resultados em ambientes de baixa reverberação. Outras técnicas utilizam um número grande de sensores espalhados pelo ambiente [6],[38], obtendo menores erros de localização, mas com um aumento significativo do custo do sistema e da complexidade do processamento, o que limita o seu emprego.

A presença de múltiplas fontes aumenta enormemente a complexidade do problema e impõe algumas restrições, como a admissão de que as fontes sejam su-

ficientemente esparsas. Essas restrições podem ser desconsideradas, empregando-se métodos de BSS. A estimação das DOAs de múltiplas fontes é abordada em [26],[33],[39],[40], sendo essas técnicas versões aprimoradas do método que busca a DOA através da GCC, sem utilizar métodos de BSS associados.

Em [41], uma das primeiras publicações em BSS, investigou-se o problema da separação para misturas instantâneas, estabelecendo-se o conceito de BSS através da ICA. O problema de separação cega de fontes já havia sido tratado anteriormente em [42], porém o seu foco era em codificação. Em [43], a identificação de fontes foi abordada através de estatísticas de ordem superior. Essa abordagem é de grande relevância em BSS, dando origem, dentre outros, ao algoritmo JADE (do inglês “Joint Approximate Diagonalization of Eigen-matrices”)[44] e ao INFOMAX (do inglês “INFORMATION MAXimization”) [45]. O desenvolvimento dos algoritmos de BSS baseados no gradiente natural (Natural ICA) teve início com o trabalho publicado em [46], gerando um algoritmo de grande importância na área, que evoluiu em diversas vertentes conforme apresentado em [5]. O algoritmo de ponto fixo FastICA, desenvolvido em [47], tem também grande destaque, sendo muito utilizado devido à sua eficiência e baixo custo computacional. Esse algoritmo evoluiu em diversas vertentes também apresentadas em [5].

A separação cega de fontes aplicada a sinais de áudio envolve, em geral, misturas convolutivas, sendo tratada em diversos trabalhos [2],[48]-[50]. A complexidade da separação nesse caso é maior do que a de misturas instantâneas, devido ao fato de ser necessário aplicar uma série de técnicas adicionais ao ICA, sendo a separação feita, em geral, no domínio da frequência, o que requer a correção de possíveis permutações (ordem aleatória dos sinais separados) e diferentes escalamentos (ganhos aleatórios) nas diversas raias de frequência. A solução para o problema da permutação pode ser obtida através da estimação da direção de chegada dos sinais nos sensores [48] ou da correlação espectral das fontes [51],[52]. A solução para a questão do escalamento é obtida através do princípio da distorção mínima [53],[54].

A associação entre os conceitos de BSS e DOA foi estabelecida em [16], onde a estimativa da DOA é feita através de um diagrama de irradiação, traçado a partir de dados fornecidos pela BSS, sendo a DOA resultante utilizada como um recurso auxiliar da BSS, com o objetivo de resolver o problema da permutação. Em [48], a DOA também é estimada a partir de dados da BSS usando o modelo de campo distante, sendo o objetivo também resolver a questão da permutação. Em [11],[12], o método que busca a DOA emprega técnicas de BSS como recurso auxiliar, sendo utilizada a transformada SCT após aplicar o algoritmo ICA recursivo às misturas.

Em [10],[55], são propostas técnicas de BSS no domínio do tempo, onde os autores abordam o problema da separação como o da identificação cega de sistema (BSI, do inglês “Blind System Identification”) de múltiplas entradas e múltiplas

saídas (MIMO, do inglês “Multiple-Input Multiple-Output”), criando o método conhecido como TRINICON (do inglês “TRIPLE-N Independent component analysis for CONVolutive mixtures), aplicável a sinais não-brancos, não-estacionários e não-gaussianos, como os sinais de voz utilizados neste trabalho. Em [55],[56], os autores abordam métodos para otimizar esse processo. Em [10] são propostos métodos de estimação da TDOA através de dados fornecidos pelo algoritmo TRINICON.

A análise de canais de propagação acústica é também um tema relevante para esta pesquisa, pois suas características, como tempo de reverberação e resposta em frequência, influenciam nos desempenhos dos métodos de BSS e de estimação da DOA. Esse tema é abordado em [57]-[59]. As simulações da propagação acústica de salas empregadas neste trabalho foram empregadas usando o método ISM (do inglês “Image-Source Model”) descrito em [60]-[62].

Em [7] foram comparados os resultados de diferentes posicionamentos dos microfones em salas de reuniões e videoconferências, através de um método de predição de erros considerando a localização de uma única fonte. No primeiro caso, em uma sala de reunião, as posições da fonte foram distribuídas de forma uniforme no ambiente, sendo a localização efetuada a partir de um arranjo com 10 microfones posicionados em uma única parede. No segundo caso, foram utilizados 4 arranjos, com 2 microfones cada, posicionados no centro de cada uma das paredes, o qual apresentou melhor rendimento do que o primeiro. Um ambiente de videoconferência também foi considerado, sendo a localização obtida através de dois arranjos com 4 sensores posicionados de 7 formas diferentes, e com 3 arranjos com 4 sensores para 8 formas de posicionamentos diferentes. Esta tese mostra a importância da escolha adequada do posicionamento dos arranjos em ambientes específicos.

Em [6],[20] foi apresentado o método SRP-PHAT (do inglês “Steered Response Power using PHASE Transform”), que mostrou performance robusta em ambientes com alta reverberação, utilizando arranjo com muitos sensores (ao menos 8). Em [25] o SRP-PHAT foi utilizado no rastreamento da posição de um locutor em uma sala inteligente, sendo utilizado um algoritmo de mascaramento de frequência na etapa do GCC, permitindo melhores resultados.

Em [23] foi proposta a utilização de uma rede neural baseada no CMN (do inglês “Cepstral Mean Normalization”) para a localização de fontes distantes dos arranjos, pelo fato dos seus sinais apresentarem drásticas distorções introduzidas pelos canais. Em [8] foi apresentado um método de localização e estimativa da direção orientada da fonte utilizando redes neurais baseadas no método SRP-PHAT, enquanto que em [31] foram utilizadas redes neurais estruturadas, empregando GMM (do inglês “Gaussian Mixture Model”) em conjunto com CMN. Em [22] foi proposto um modelo de localização e estimativa da direção orientada de fonte em torno de um par de microfones, através de uma rede neural estruturada empregando ITD e ILD (do

inglês “Interaural Level Difference”).

Em [26] foi proposto um método de localização com múltiplas fontes, utilizando o princípio de que sempre há interrupções na fala humana, o que permite a localização de múltiplas fontes simultaneamente ativas.

1.3 Estrutura da Tese

O restante do texto desta tese está organizado da seguinte forma:

- o Cap. 2 aborda a teoria do canal de propagação acústica, descrevendo as características físicas da propagação do som em um recinto, a modelagem do canal empregada no desenvolvimento dos algoritmos e nas simulações, e os parâmetros que são relevantes ao seu estudo, como o tempo de reverberação;
- o Cap. 3 aborda os principais fundamentos da separação cega de fontes, sendo descritos algoritmos de BSS para misturas instantâneas baseados na análise de componentes independentes e para misturas convolutivas, considerando tanto a abordagem em banda estreita (domínio da frequência) quanto a em banda larga (no domínio do tempo);
- o Cap. 4 descreve as principais técnicas de estimação de ângulo de chegada, abordando tanto o caso mais simples, em que há somente uma fonte ativa, como o caso de múltiplas fontes, onde as DOAs são obtidas a partir dos parâmetros de algoritmos de BSS; em seguida, são propostas modificações nos algoritmos de BSS para torná-los mais eficientes na localização das DOAs;
- o Cap. 5 descreve o sistema empregado para localização de múltiplas fontes, e apresenta o modelo que determina a região de localização apropriada e auxilia na otimização do posicionamento dos microfones;
- o Cap. 6 apresenta os resultados experimentais nos ambientes simulados e real para a localização de uma fonte, sendo analisadas as influências da reverberação, do posicionamento dos arranjos de microfones e da posição da fonte na sala nas taxas de sucesso da localização;
- o Cap. 7 apresenta os resultados experimentais nos ambientes simulados e real para a localização de duas fontes ativas pelo sistema proposto com os diferentes métodos baseados em BSS, além de avaliar o modelo proposto de otimização e determinação da região apropriada com os dados experimentais;
- o Cap. 8 apresenta as conclusões do trabalho e as propostas de trabalhos futuros.

Capítulo 2

Misturas de Sinais Acústicos

Para que as técnicas de separação cega de fontes sejam bem sucedidas, é fundamental considerar algumas características do processo de mistura.

A Fig. 2.1 mostra os processos de mistura e separação, onde s_i representa o sinal da i -ésima fonte, x_j é o sinal adquirido pelo j -ésimo sensor e representa a mistura j , $y_i(n)$ representa a estimativa da fonte i , h_{ij} representa os filtros de misturas e w_{ij} representa os filtros de separação. Assumindo-se que as misturas são lineares, os sinais adquiridos pelos sensores podem ser representados por

$$x_j(n) = \sum_{i=1}^N \sum_{l=0}^{L-1} h_{ij}(l) s_i(n-l), \quad j = 1, \dots, M \quad (2.1)$$

onde $h_{ij}(n)$ representa a resposta ao impulso do canal entre a fonte i e o sensor j .

No desenvolvimento do sistema de separação, é importante saber se a mistura é instantânea ou convolutiva, sendo que nas misturas instantâneas os filtros que representam os canais entre as fontes e os sensores podem ser representados por constantes (uma vez que as misturas num determinado instante de tempo são combinações lineares das amostras das fontes no mesmo instante de tempo), enquanto nas misturas convolutivas os canais são representados por filtros com diversos coeficientes (as misturas ocorrem através da convolução dos sinais das fontes com as respostas ao impulso dos filtros).

Outra classificação importante é em relação à quantidade de fontes e sensores. Se a quantidade de sensores for maior do que a de fontes, o sistema de misturas é denominado superdeterminado; quando a quantidade de fontes é igual à quantidade de sensores, diz-se que o sistema é determinado; e quando a quantidade de fontes é maior que a quantidade de sensores, o sistema é denominado subdeterminado.

As classificações citadas acima são importantes, já que a análise de componentes independentes é adequada apenas para misturas instantâneas e determinadas ou sobredeterminadas. Existem extensões específicas dessa técnica para os casos de misturas convolutivas e subdeterminadas, sendo que misturas convolutivas comumente

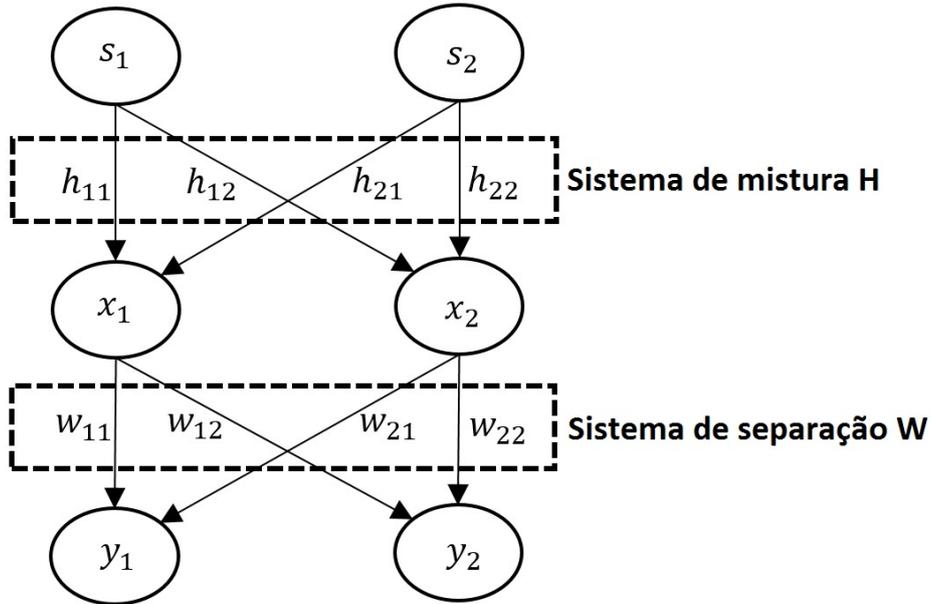


Figura 2.1: Diagrama dos processos de mistura e separação.

ocorrem em canais de áudio.

2.1 Canal de Propagação Acústica

As ondas acústicas, ao se propagarem em um determinado ambiente onde há obstáculos, sofrem reflexão, difração e interferências, conforme descrito em [57].

A Fig. 2.2 mostra a representação física de um canal de áudio com quatro possíveis percursos para a onda entre a fonte de áudio e o microfone. As atenuações sofridas pelo sinal nesses diversos percursos formam os coeficientes do filtro que representa o canal. Outro fato importante é que esses percursos geram o fenômeno conhecido como reverberação.

A dificuldade ao se projetar um filtro que represente matematicamente um canal de áudio está na obtenção dos seus coeficientes de forma que esses correspondam às atenuações e atrasos sofridos durante os percursos, pois essa tarefa depende de diversos fatores, como a geometria da sala, o tipo de material das paredes, teto e piso, e as posições de objetos no ambiente.

É importante destacar que os canais de áudio avaliados neste trabalho correspondem a ambientes simulados e reais, com uma quantidade significativamente maior de percursos entre a fonte e o microfone do que os quatro mostrados na Fig. 2.2, gerando filtros com um grande número de coeficientes, o que está diretamente associado ao alto nível de complexidade das tarefas de separação e localização.

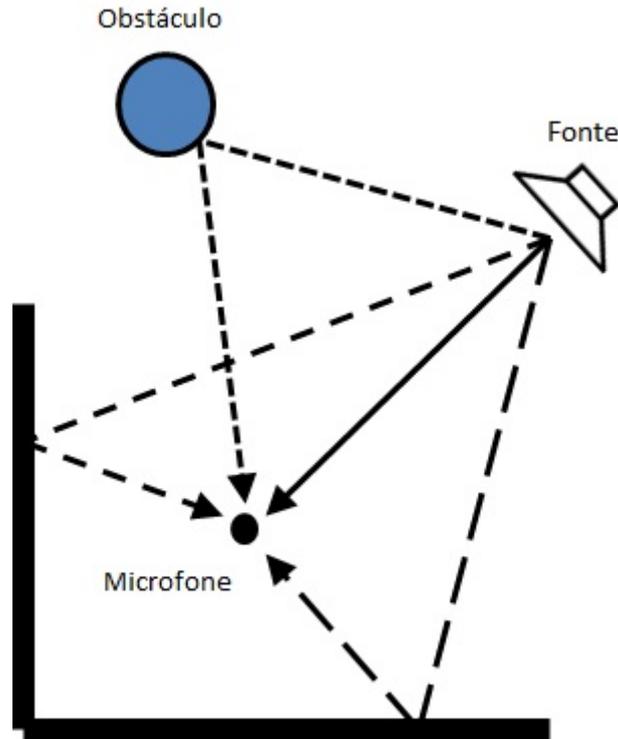


Figura 2.2: Ilustração simplificada do canal de áudio.

2.2 Simulação do Canal

Os canais de propagação acústica avaliados nesse trabalho foram salas, sendo as suas respostas ao impulso conhecidas como RIR (do inglês “Room Impulse Response”). O modelo usado para gerar a RIR nas simulações foi o “Image-Source”, originalmente proposto em [62], sendo nesse trabalho utilizada a sua versão modificada desenvolvida em [61], que possui excelente qualidade de simulação do canal do ambiente e de otimização computacional.

O modelo “Image-Source” se vale do princípio de que as paredes, o piso e o teto da sala funcionam como “espelhos” para as ondas acústicas, de modo que cada imagem refletida nesses espelhos funciona como uma imagem da fonte original, podendo em cada reflexão haver uma perda de energia do sinal, em função do coeficiente de absorção do material que reflete as ondas e do ar.

A Fig. 2.3 mostra o princípio do modelo “Image-Source”, onde há uma fonte de áudio original e um microfone. As paredes, consideradas espelhos, geram as imagens refletidas da fonte, resultando:

- na Fonte-imagem 1, reflexo da fonte original na parede (espelho) 1 através do percurso 1;
- na Fonte-imagem 2, reflexo da fonte original na parede (espelho) 2 através do percurso 2;

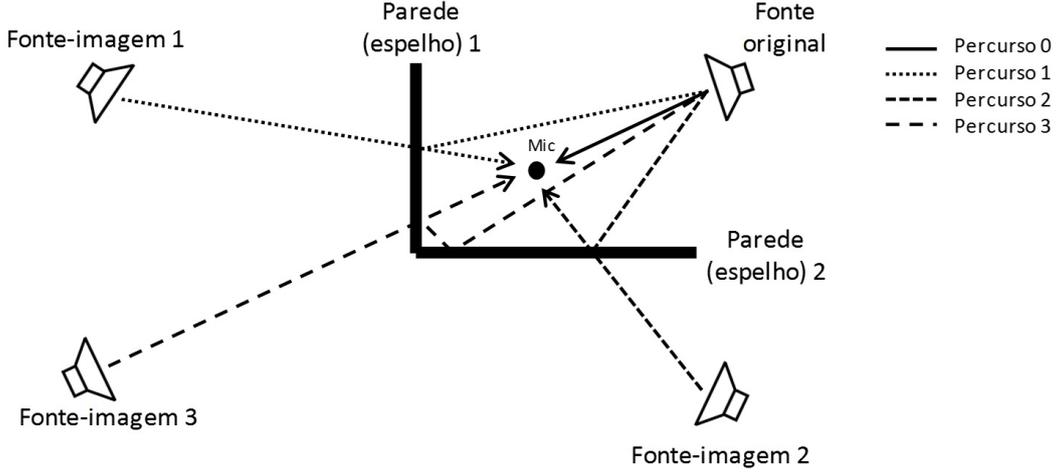


Figura 2.3: Modelo “Image-Source” simplificado.

- na Fonte-imagem 3, reflexo da fonte original inicialmente na parede (espelho) 2 e depois na parede (espelho) 1 através do percurso 3.

A Fonte-imagem 3 é o resultado da reflexão em duas paredes. Logo, partindo do princípio de que podem ocorrer reflexões em diversas paredes, é possível elaborar um conceito mais complexo, estimando a quantidade de percursos, ou “fontes-imagens”, possíveis para um modelo de sala em 3 dimensões, onde há uma quantidade significativamente maior de imagens da fonte.

A complexidade do modelo está justamente em determinar um filtro com coeficientes equivalentes às atenuações que o sinal sofre nos percursos percorridos, considerando os atrasos correspondentes.

No modelo físico utilizado nas simulações desse trabalho [61], denominam-se, no sistema de coordenadas cartesianas (x, y, z) , os vetores das posições da fonte e do microfone como $\rho_{\mathbf{s}}$ e $\rho_{\mathbf{r}}$, respectivamente, sendo as dimensões da sala dadas no vetor \mathbf{r} , o coeficiente de absorção igual a α e o de reflexão igual a β , onde:

$$\rho_{\mathbf{s}} = [x_s, y_s, z_s]^T \quad (2.2)$$

$$\rho_{\mathbf{r}} = [x_r, y_r, z_r]^T \quad (2.3)$$

$$\mathbf{r} = [L_x, L_y, L_z]^T \quad (2.4)$$

$$\alpha = 1 - \beta \quad (2.5)$$

Os coeficientes de reflexão das superfícies são dados por $\beta_{x,i}$, $\beta_{y,i}$ e $\beta_{z,i}$ e o filtro $h(\cdot)$ que representa a RIR é dado por:

$$h(t) = \sum_{u=0}^1 \sum_{l=0}^{\infty} A(\mathbf{u}, \mathbf{l}) \delta(t - \tau(\mathbf{u}, \mathbf{l})) \quad (2.6)$$

onde t caracteriza o tempo, $\delta(\cdot)$ é a função impulso de Dirac e $\mathbf{u}(u, v, w)$ e $\mathbf{l}(l, m, n)$ são parâmetros que controlam a indexação das imagens da fonte em todas as dimensões. O fator de atenuação, $A(\cdot)$, e os atrasos, τ , são obtidos por:

$$A(\mathbf{u}, \mathbf{l}) = \frac{(-\beta_{x,1})^{|l-u|} (-\beta_{x,2})^{|l|} (-\beta_{y,1})^{|m-v|} (-\beta_{y,2})^{|v|} (-\beta_{z,1})^{|n-w|} (-\beta_{z,2})^{|n|}}{4\pi d(\mathbf{u}, \mathbf{l})} \quad (2.7)$$

$$\tau(\mathbf{u}, \mathbf{l}) = d(\mathbf{u}, \mathbf{l}) / c \quad (2.8)$$

onde c é a velocidade do som e $d(\cdot)$ representa a distância do microfone até a respectiva imagem da fonte, dada por:

$$d(\mathbf{u}, \mathbf{l}) = \|\text{diag}(2u-1, 2v-1, 2w-1) \rho_s + \rho_r - \text{diag}(2l, 2m, 2n) \mathbf{r}\| \quad (2.9)$$

sendo $\|\cdot\|$ a normalização euclidiana e $\text{diag}(\cdot)$ a matriz diagonal.

As Eqs. (2.6)-(2.9) indicam o modelo matemático usado para representar a RIR $h(t)$, sendo o modelo simplificado $\hat{h}_p(t)$ usado nas simulações, tal que $\hat{h}_p(t) \approx h(t)$ em relação à função de transferência de potência, dada por:

$$\hat{h}_p(t) = \frac{1}{8\rho\bar{r}} \left\{ \begin{array}{l} \frac{B_z}{\log\left(\frac{B_z}{B_x}\right)} \left(E_i \left(\log \left(\frac{B_z}{B_x} \right) \right) + \log \left(\log \frac{B_z}{B_x} \right) \right) \\ - \left(E_i \left(\log \left(\frac{B_z}{B_y} \right) \right) + \log \left(\log \frac{B_z}{B_y} \right) \right) \rightarrow \text{se } B_x \neq B_y \neq B_z \\ \\ \frac{B_z}{\log\left(\frac{B_z}{B_x}\right)} \left(E_i \left(\log \left(\frac{B_z}{B} \right) \right) + \log \left(\log \frac{B_z}{B} \right) \right) - \gamma \rightarrow \\ \text{se } B_z = B_y \neq B_x \triangleq B \text{ ou } B_z = B_x \neq B_y \triangleq B \\ \\ \frac{B-B_z}{\log\left(\frac{B}{B_z}\right)} \text{ se } B_z \neq B_x = B_y \triangleq B \\ \\ B \text{ se } B_x = B_y = B_z \triangleq B \end{array} \right. \quad (2.10)$$

onde $E_i(\cdot)$ é a integral de primeira ordem da exponencial representada em [63], γ é

a constante de Euler-Mascheroni,

$$\varrho \triangleq \varrho(t) = ct \quad (2.11)$$

e

$$B_x = (\beta_{x,1}\beta_{x,2})^{\frac{\varrho}{L_x}} \quad (2.12)$$

$$B_y = (\beta_{y,1}\beta_{y,2})^{\frac{\varrho}{L_y}} \quad (2.13)$$

$$B_z = (\beta_{z,1}\beta_{z,2})^{\frac{\varrho}{L_z}} \quad (2.14)$$

2.3 Decaimento de Energia da RIR

A representação gráfica da queda de energia da envoltória, $E(t)$, da RIR $h(t)$ é conhecida como EDC (do inglês “Energy Decay Curve”), a qual pode ser obtida, usando a versão normalizada do método de Integração Schroeder [64], [65], através de

$$E(t) = 10 \log_{10} \left(\frac{\int_t^\infty h_p^2(\xi) d\xi}{\int_0^\infty h_p^2(\xi) d\xi} \right) \quad (2.15)$$

em decibéis.

Existem dois pontos importantes na função EDC, o T_{20} e o T_{60} , onde o T_{20} é o intervalo de tempo que $E(t)$ leva para variar de 5 dB até 25 dB e o T_{60} é o tempo que $E(t)$ demora para atingir 60 dB após decair 5 dB. Ambos os parâmetros T_{20} e T_{60} são conhecidos na literatura como tempo de reverberação e são usados para qualificar o ambiente simulado (muito ou pouco reverberante).

A forma discreta da Eq. (2.15) é obtida através da aproximação:

$$\int_t^\infty \widehat{h}_P\{\xi\} d\xi \approx T \sum_{i=0}^\infty \widehat{h}_P(t + iT) \quad (2.16)$$

resultando em:

$$\widehat{E}(t) = 10 \log_{10} \left(\frac{\sum_{i=0}^\infty \widehat{h}_P(t + iT)}{\sum_{i=0}^\infty \widehat{h}_P(t_0 + iT)} \right), \quad t > t_0 \quad (2.17)$$

2.4 Considerações Finais sobre as Misturas

O objetivo deste capítulo foi introduzir os principais conceitos relacionados à classificação das misturas e à teoria sobre modelagem do canal de propagação de ondas sonoras, além de indicar o método que foi utilizado nas simulações e introduzir o conceito de tempo de reverberação, ao qual a eficiência dos sistemas de localização e separação está associado. Mais detalhes a respeito das considerações e análises referentes ao desenvolvimento do modelo podem ser encontrados em [60]-[62].

Capítulo 3

Algoritmos de Separação Cega de Fontes

Este capítulo aborda os principais fundamentos de separação cega de fontes. Inicialmente, são apresentados algoritmos de BSS para misturas instantâneas, desenvolvidos a partir da ICA, sendo descritos o algoritmo ICA baseado no gradiente natural e o FastICA. Em seguida, são apresentadas as técnicas de BSS empregadas em misturas convolutivas, sendo descritos o método de separação no domínio da frequência (de banda estreita), com as principais abordagens para resolver os problemas de permutação e de escalamento, e o método TRINICON (de banda larga). Finalmente, são apresentadas as medidas usualmente empregadas na avaliação da separação.

3.1 Separação de Misturas Instantâneas

Para a separação ser bem sucedida pela ICA, as misturas devem ser instantâneas e determinadas e as fontes devem possuir algumas características, como não-gaussianidade e independência estatística [5]. Os sinais das fontes consideradas neste trabalho são vozes de diferentes pessoas, cujas amostras possuem distribuição supergaussiana e tendem a ser independentes, atendendo às condições para a separação ser bem sucedida, considerando-se ambientes anecóicos e diferenças entre os tempos de chegada dos sinais nos microfones desprezíveis.

Inicialmente será descrita a etapa de pré-processamento, necessária para o bom desempenho da maioria dos métodos de BSS. Em seguida, serão apresentados dois algoritmos adaptativos utilizados para se obter a matriz de separação da ICA.

3.1.1 Pré-processamento

As misturas devem passar por duas etapas de pré-processamento antes de se aplicar a ICA: a centralização e o branqueamento.

Dadas as fontes $s_i(n)$, $i = 1, \dots, N$, o vetor $\mathbf{s}(n)$ contém as amostras das fontes no instante n , ou seja,

$$\mathbf{s}(n) = \begin{bmatrix} s_1(n) \\ s_2(n) \\ \vdots \\ s_N(n) \end{bmatrix} \quad (3.1)$$

Sendo \mathbf{H} a matriz com os coeficientes de mistura, o vetor $\mathbf{x}(n)$ que contém as n -ésimas amostras das misturas $x_j(n)$ é dado por:

$$\mathbf{x}(n) = \mathbf{H}\mathbf{s}(n) \quad (3.2)$$

O processo de centralização, que torna a média de cada mistura $x_j(n)$ igual a zero, é implementado através de:

$$x_{c,j}(n) = x_j(n) - E[x_j(n)] \quad (3.3)$$

onde $E[\cdot]$ corresponde ao valor médio estatístico de cada componente do vetor.

O processo de branqueamento consiste em multiplicar o vetor $\mathbf{x}_c(n)$, com elementos $x_{c,j}(n)$, por uma matriz branqueadora \mathbf{V} , gerando o vetor com elementos descorrelacionados $\mathbf{z}(n)$, dado por:

$$\mathbf{z}(n) = \mathbf{V}\mathbf{x}_c(n) \quad (3.4)$$

onde

$$\mathbf{V} = \mathbf{D}^{-1/2}\mathbf{E}^H \quad (3.5)$$

sendo \mathbf{D} a matriz diagonal com os autovalores da matriz de covariância $\mathbf{C}_x = E[\mathbf{x}(n)\mathbf{x}^H(n)]$ e \mathbf{E} a matriz cujas colunas são os autovetores da matriz de covariância \mathbf{C}_x .

Em alguns casos o processo de branqueamento pode não ser necessário, mas quando aplicado reduz os ciclos de processamento no algoritmo de separação.

3.1.2 Análise de Componentes Independentes

A ICA é uma técnica que, através de algoritmos iterativos [5], busca a matriz de separação \mathbf{W} a ser aplicada ao vetor de mistura $\mathbf{x}(n)$ (ou ao vetor $\mathbf{z}(n)$ descrito acima) para se obter o vetor $\mathbf{y}(n)$, cujas colunas são formadas por sinais com propriedade

de independência estatística, correspondendo às estimativas das fontes.

Há um conjunto de algoritmos baseados na maximização da gaussianidade dos sinais desenvolvidos para se obter a solução da ICA. Entre os principais encontram-se o algoritmo INFOMAX, proposto em [45], e sua versão modificada, apresentada em [66], que utiliza um método de atualização dos coeficientes da matriz de separação baseado no gradiente natural.

ICA baseado no Gradiente Natural

A regra de aprendizado utilizada no algoritmo ICA baseado no gradiente natural, considerando sinais reais, é dada pelos seguintes passos:

$$\mathbf{W} \leftarrow \mathbf{W} + \eta [\mathbf{I} + \Phi(\mathbf{y}) \mathbf{y}^T] \mathbf{W} \quad (3.6)$$

$$\mathbf{W} = \frac{\mathbf{W}}{\|\mathbf{W}\|} \quad (3.7)$$

onde η é o parâmetro de aprendizado, $\phi(\cdot)$ é uma função não-linear utilizada para avaliar a não-gaussianidade dos sinais de saída considerada na separação, e $\|\mathbf{W}\|$ é a norma euclidiana de \mathbf{W} .

A diferença entre os vários algoritmos que usam o gradiente natural está na escolha de $\phi(\cdot)$, pois existem diversas funções com diferentes propriedades matemáticas que podem ser utilizadas de forma que o algoritmo convirja para uma solução apropriada para a separação cega dos sinais. A maximização da não-gaussianidade pode ser obtida através da avaliação de diferentes estatísticas de mais alta ordem dos sinais, tais como a curtose e a negentropia, ou por estimativas de máxima verossimilhança, dentre outros métodos [5].

Algoritmos de Ponto Fixo

Os algoritmos de ponto fixo, introduzidos em [47], também buscam propriedades que levem à separação dos sinais e têm a capacidade de alcançar a convergência de forma mais rápida, em comparação com os algoritmos que usam o gradiente natural. Um dos algoritmos de ponto fixo mais conhecido é o FASTICA, que busca a matriz de separação \mathbf{W} através do procedimento iterativo definido pelas seguintes equações:

$$\mathbf{W} \leftarrow E [\mathbf{x}g(\mathbf{W}^T \mathbf{x})] - E [g'(\mathbf{W}^T \mathbf{x})] \quad (3.8)$$

$$\mathbf{W} = \frac{\mathbf{W}}{\|\mathbf{W}\|} \quad (3.9)$$

onde $g(\cdot)$ é a função não-linear que explora as propriedades estatísticas pertinentes à operação de separação e $g'(\cdot)$ é a sua derivada.

Assim como nos algoritmos baseados no gradiente natural, as diferenças entre os algoritmos da família FASTICA residem na propriedade estatística explorada, que é determinada pela função não-linear utilizada.

Considerações Finais sobre ICA

O algoritmo ICA executa uma etapa importante no processo de separação de misturas convolutivas no domínio da frequência, onde a STFT é aplicada aos sinais gerando valores complexos. Assim, para aplicar o ICA neste caso é necessário utilizar as funções adequadas para sinais complexos, como as apresentadas em [2], [5] e [16].

3.2 Separação de Misturas Convolutivas

Nas misturas convolutivas, os canais aos quais os sinais das fontes são submetidos antes de atingir os sensores podem ser representados por filtros com diversos coeficientes. No caso específico dos canais de propagação acústica, foco desta pesquisa, os filtros são os descritos no Cap. 2.

A Fig. 3.1 mostra um exemplo com duas fontes e dois sensores, onde as misturas $x_1(n)$ e $x_2(n)$ são dadas por:

$$x_1(n) = s_1(n) * h_{11}(n) + s_2(n) * h_{21}(n) \quad (3.10)$$

$$x_2(n) = s_1(n) * h_{12}(n) + s_2(n) * h_{22}(n) \quad (3.11)$$

onde “*” representa a operação de convolução.

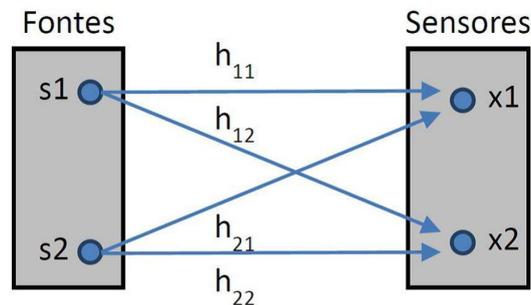


Figura 3.1: Mistura convolutiva com duas fontes e dois sensores.

Para se separar os sinais das fontes $s_i(n)$ a partir dos sinais das misturas $x_i(n)$ no caso convolutivo, é necessário utilizar filtros com comprimentos semelhantes aos

dos filtros de mistura. As estimativas das fontes são então produzidas por

$$\mathbf{y}(n) = \mathbf{W} * \mathbf{x}(n) \quad (3.12)$$

sendo “*” a convolução de cada elemento da matriz \mathbf{W} , dado pela resposta ao impulso de um dos filtros de separação, com o sinal de mistura correspondente. Os coeficientes da matriz de mistura podem ser obtidos por técnicas que operam no domínio da frequência, conhecidas como de banda estreita, ou no domínio do tempo, conhecidas como de banda larga.

3.2.1 Separação no Domínio da Frequência

A convolução de sinais no domínio do tempo corresponde, no domínio da frequência, ao produto de suas transformadas de Fourier. Usando essa propriedade, na separação de misturas convolutivas os sinais são convertidos para o domínio da frequência através da STFT, deixando-os adequados à aplicação da ICA.

A separação de misturas convolutivas no domínio da frequência é obtida através das seguintes etapas:

- I. Aplica-se a STFT às misturas $x_i(n)$ gerando $X_i(f_k, m)$, onde i , m e k são os índices correspondentes às fontes, bloco de dados (ou *frame*) e raia (ou *bin*) de frequência, respectivamente;
- II. Aplica-se o algoritmo ICA em cada raia de frequência para obter a matriz de separação $\mathbf{W}(f_k)$ e a estimativa do sinal separado por faixa de frequência $Y_i(f_k, m)$;
- III. Usa-se o princípio da correlação espectral e/ou DOA para corrigir permutações das estimativas das fontes entre as raias de frequência;
- IV. Usa-se o princípio da distorção mínima para corrigir diferentes escalamentos nas diferentes raias de frequência;
- V. Aplica-se a STFT inversa para obter as estimativas dos sinais das fontes no domínio do tempo.

A Fig. 3.2 mostra um diagrama com todas as etapas do processo de separação de misturas convolutivas.

É importante observar que cada etapa possui características específicas. Na etapa I é fundamental a escolha do salto, comprimento e tipo de janela utilizados na STFT para que o processo de separação seja eficiente. O mesmo princípio vale para a escolha da função adequada no ICA, e para os processos de correção da permutação

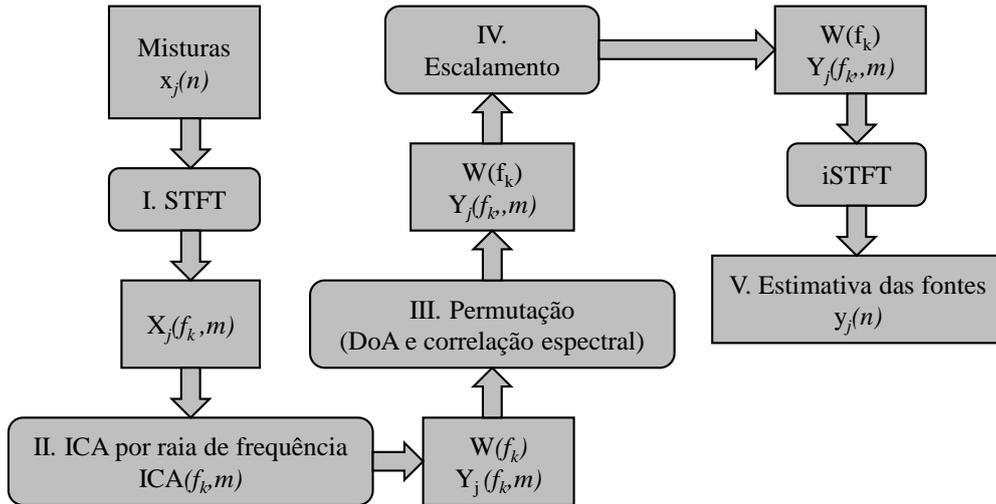


Figura 3.2: Etapas do processo de separação de misturas convolutivas.

e escalamento nas diferentes raias de frequência. Esses parâmetros serão indicados quando utilizados ao longo deste trabalho.

Permutação nas Diferentes Raias de Frequência

O problema da permutação ocorre após o término da etapa II, pois os sinais são separados em cada *bin* de frequência de forma independente, sem garantia de que a ordem das estimativas dos sinais das fontes sejam idênticas em todos os *bins*. Então, utiliza-se o princípio da correlação espectral dos sinais e/ou a identificação da direção de chegada dos sinais nos sensores para reordenar as colunas das matrizes de separação e alinhar as estimativas dos sinais em cada faixa de frequência, conforme descrito a seguir.

Solução através da Estimativa da Direção de Chegada

Uma forma de resolver o problema da permutação é através dos ângulos de chegada dos sinais separados em cada raia de frequência, os quais podem ser estimados da matriz de separação $\mathbf{W}(f_k)$ por diferentes métodos, conforme abordado em [48],[12],[16]. As equações utilizadas para tal procedimento serão apresentadas na Seção 4.6.

Solução através da Correlação entre Raias de Frequência

O sinal de voz de cada falante possui uma característica individual quando se analisa o comportamento da sua envoltória ao longo do tempo, observando-se a tendência de ocorrerem picos de energia simultâneos em diversas raias de frequência. Utilizando este princípio, o emprego da covariância das envoltórias dos sinais nas diversas raias

para ordenar as estimativas das fontes foi proposto em [67]. Para reduzir o número de permutações errôneas, principalmente em raias correspondentes à baixa frequência, foi proposta em [52] a substituição da covariância pelo coeficiente de correlação. Resultados mais confiáveis foram obtidos explorando a esparsidade do sinal de voz no domínio da frequência, conforme sugerido em [15], através do emprego da razão de potência (RP), estabelecida como razão entre a potência de cada sinal separado e a potência total (considerando-se todos os sensores), dada por:

$$\mathbf{RP}_i(f_k, m) = \frac{\|\mathbf{a}_i(f_k)Y_i(f_k, m)\|^2}{\sum_{l=1}^N \|\mathbf{a}_l(f_k)Y_l(f_k, m)\|^2} \quad (3.13)$$

onde $\mathbf{a}_i(f_k)$ é a i -ésima coluna de $\mathbf{W}^{-1}(f_k)$. A razão de potência definida em (3.13) apresenta valor entre 0 e 1, que corresponde ao grau de dominância da i -ésima fonte no ponto (f_k, m) , uma vez que seu denominador contém a soma das energias desta fonte ao longo dos sensores.

Escalamento nas Diferentes Raias de Frequência

O problema do escalamento ocorre porque a separação é feita por raia de frequência de forma independente, não garantindo ganho em amplitude uniforme em todas as faixas de frequência estabelecidas na STFT. Consequentemente, os filtros de separação nas raias geram ganhos diferentes. A solução para este problema foi proposta em [53],[54], usando uma técnica conhecida como Princípio da Distorção Mínima (MDP, do inglês “Minimal Distortion Principle”). Esta técnica propõe a redução máxima da distância entre as estimativas e as misturas. Assim, pelo ponto de vista da BSS, a função de transferência $\mathbf{W}(f)$ deve ser a mais próxima possível de $\mathbf{H}(f)$. Este procedimento é feito através da seguinte normalização:

$$\mathbf{W}(f) \leftarrow \mathbf{\Lambda}(f) \mathbf{W}(f) \quad (3.14)$$

$$\mathbf{\Lambda}(f) = \text{diag}(a_{11}(f), \dots, a_{NN}(f)) \quad (3.15)$$

onde $a_{ij}(f) = [\mathbf{W}^+(f)]_{ij}$ são os elementos da inversa de $\mathbf{W}(f)$ no caso determinado, ou da pseudoinversa de $\mathbf{W}(f)$ no caso subdeterminado. O MDP garante que as distorções nas estimativas são proporcionais às provocadas pela mistura, gerando um resultado melhor do que um modelo livre de restrição em relação ao escalamento.

3.2.2 Separação no Domínio do Tempo

A separação de misturas convolutivas pode ser feita no domínio do tempo, através do algoritmo TRINICON [10],[68],[24]. O termo “Triple-N” se refere a atributos dos sinais de voz, que são Não-brancos (colorido), Não-gaussianos e Não-estacionários.

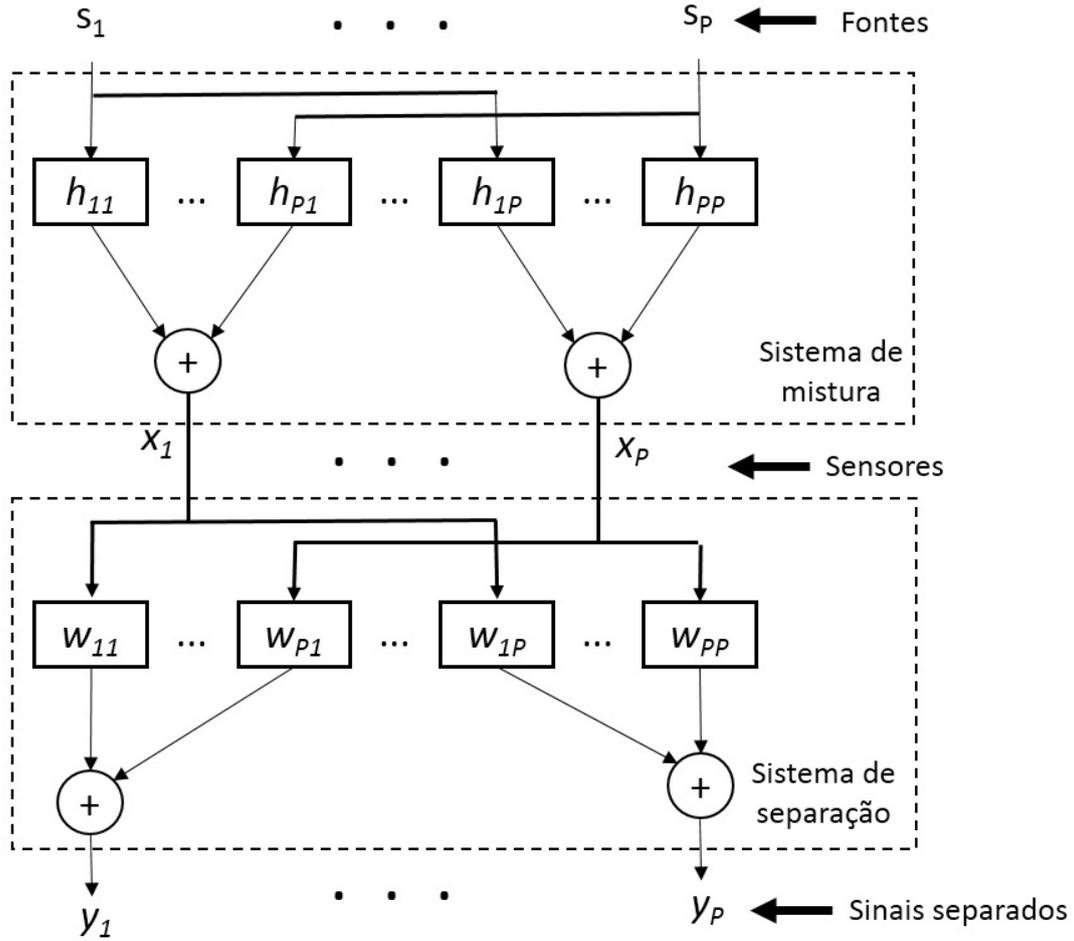


Figura 3.3: Sistema de mistura e separação com os parâmetros utilizados no método TRINICON.

Este método é oriundo da área de pesquisa de desconvolução cega (MCBD, do inglês “Multichannel Blind Deconvolution”) [4],[69], que aborda a separação das fontes como um problema de identificação cega de um filtro de múltiplas entradas e múltiplas saídas (MIMO, do inglês “Multi-Input Multi-Output”).

O processo de adaptação e de separação nesse método ocorre no domínio do tempo, diferentemente do método anterior, cujas matrizes de separação e estimativas das fontes eram obtidas por raia de frequência, sendo necessário aplicar a STFT às misturas antes do processo de separação, e a operação inversa iSTFT nos sinais separados após a BSS.

A Fig. 3.3 ilustra os processos de mistura e separação, onde os sinais das fontes $s_i, i = 1, \dots, P$, são submetidos ao sistema de mistura MIMO \mathbf{H} composto pelos filtros h_{ij} , gerando os sinais das misturas x_j capturados pelos P sensores. Esses sinais são combinados através de filtros w_{ij} , que compõem o sistema de separação, gerando as estimativas das fontes y_i .

Os coeficientes dos filtros que compõem o sistema de separação \mathbf{W} são obtidos através de um algoritmo iterativo. Considerando filtros FIR (do inglês “Finite Im-

pulse Response”) de comprimento L , as estimativas das fontes podem ser escritas na forma matricial como [24],[55]:

$$\mathbf{y}(m, j) = \mathbf{x}(m, j)\mathbf{W}(m) = \sum_{p=1}^P \mathbf{x}_p(m, j)\mathbf{W}_{pq}(m) \quad (3.16)$$

onde m é o índice do bloco (*frame*) de dados, $j = 0, \dots, N - 1$ é o índice de deslocamento dentro do bloco de comprimento N , e

$$\mathbf{x}(m, j) = [\mathbf{x}_1(m, j), \dots, \mathbf{x}_P(m, j)] \quad (3.17)$$

$$\mathbf{y}(m, j) = [\mathbf{y}_1(m, j), \dots, \mathbf{y}_P(m, j)] \quad (3.18)$$

$$\mathbf{W}(m) = \begin{bmatrix} \mathbf{W}_{11}(m) & \cdots & \mathbf{W}_{1P}(m) \\ \vdots & \ddots & \vdots \\ \mathbf{W}_{P1}(m) & \cdots & \mathbf{W}_{PP}(m) \end{bmatrix} \quad (3.19)$$

$$\mathbf{x}_p(m, j) = [x_p(mL + j), \dots, x_p(mL - 2L + 1 + j)] \quad (3.20)$$

$$\mathbf{y}_q(m, j) = [y_q(mL + j), \dots, y_q(mL - D + 1 - j)] \quad (3.21)$$

Na Eq. (3.21), D denota o número de atrasos levado em consideração para explorar o colorimento dos sinais das fontes. $\mathbf{W}_{pq}(m)$ denota a matriz de Sylvester, de dimensões $2L \times D$, que contém os L coeficientes do filtro do p -ésimo sensor à q -ésima saída, dada por:

$$\mathbf{W}_{pq}(m) = \begin{bmatrix} w_{pq,0} & 0 & \cdots & 0 \\ w_{pq,1} & w_{pq,0} & \ddots & \vdots \\ \vdots & w_{pq,1} & \ddots & 0 \\ w_{pq,L-1} & \vdots & \ddots & w_{pq,0} \\ 0 & w_{pq,L-1} & \ddots & w_{pq,1} \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & w_{pq,L-1} \\ 0 & \cdots & 0 & 0 \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \end{bmatrix} \quad (3.22)$$

A forma geral do algoritmo TRINICON baseado no gradiente natural [70], com a função custo desenvolvida a partir do princípio da minimização da informação mútua apresentada em [68], é dada por:

$$\mathbf{W}(m) = \mathbf{W}(m - 1) - \mu \Delta \mathbf{W}(m) \quad (3.23)$$

onde

$$\Delta \mathbf{W}(m) = \sum_{i=0}^{\infty} \beta(i, m) \mathbf{W}(i) \left\{ \widehat{\mathbf{R}}_{yy}(i) - \widehat{\mathbf{R}}_{ss}(i) \right\} \widehat{\mathbf{R}}_{ss}^{-1}(i) \quad (3.24)$$

$\beta(i, m)$ é uma função janela normalizada tal que $\sum_{i=1}^m \beta(i, m) = 1$ [55]; $\widehat{\mathbf{R}}_{yy}$ é a matriz de correlação, de dimensão $PD \times PD$, estimada em um bloco de dados de comprimento N , contendo as submatrizes de correlação e correlação cruzada entre as saídas do sistema de separação para os atrasos $-D + 1, \dots, D - 1$; e $\widehat{\mathbf{R}}_{ss}(i) = \text{bdiag}(\widehat{\mathbf{R}}_{yy}(i))$ é a matriz diagonal em blocos obtida zerando todas as submatrizes de $\widehat{\mathbf{R}}_{yy}(i)$ que não se encontram na diagonal principal.

Para o caso com duas fontes e duas misturas temos:

$$\Delta \mathbf{W} = \sum_{i=0}^m \beta(i, m) \mathbf{W}(i) \begin{bmatrix} 0 & \tilde{\mathbf{R}}_{y_1 y_2}(i) \mathbf{R}_{y_2 y_2}^{-1}(i) \\ \tilde{\mathbf{R}}_{y_1 y_2}(i) \mathbf{R}_{y_1 y_1}^{-1}(i) & 0 \end{bmatrix} \quad (3.25)$$

onde, no caso mais geral,

$$\tilde{\mathbf{R}}_{y_p y_q}(i) = \frac{1}{N} \sum_{j=0}^{N-1} \phi_D \left(\mathbf{y}_q(i, j) \mathbf{R}_{y_q y_q}^{-1}(i) \mathbf{y}_q^H(i, j) \right) \mathbf{y}_p^H(i, j) \mathbf{y}_q(i, j) \quad (3.26)$$

sendo

$$\phi_D(s) = -\frac{f_D'(s)}{f_D(s)} \quad (3.27)$$

uma função não linear que pode ser derivada analiticamente a partir da função densidade de probabilidade f_D .

3.3 Avaliação do Sistema de Separação

As principais medidas utilizadas na avaliação do desempenho dos métodos de separação cega de fontes são a relação fonte-interferência (SIR, do inglês “Source to Interferences Ratio”), a relação fonte-distorção (SDR, do inglês “Source to Distortion Ratio”) e a relação fonte-artefatos (SAR, do inglês “Source to Artifacts Ratio”), definidas em [71]. A aplicação dessas medidas requer o conhecimento dos sinais das fontes para realizar a comparação com os sinais separados, sendo que os dados do sistema de mistura não são necessários. A utilização dessas medidas se baseia no princípio de que as misturas podem ser modeladas como:

$$\mathbf{x}(n) = \mathbf{H} * \mathbf{s}(n) + \nu(n) \quad (3.28)$$

onde \mathbf{H} é a matriz cujos elementos são as respostas impulsivas dos filtros de mistura apresentados na Cap. 2 e $\nu(n)$ é o vetor que coleta os erros de modelagem e ruídos.

Dada a j -ésima fonte s_j e a sua estimativa y_j , a primeira etapa para o cálculo

da SIR, da SDR e da SAR, consiste na decomposição de y_j em quatro componentes, conforme:

$$y_j = s_{alvo} + e_{interf} + e_{ruído} + e_{artef} \quad (3.29)$$

onde s_{alvo} corresponde a uma versão modificada de s_j por uma distorção permitida, e_{interf} a erros devidos a interferências, $e_{ruído}$ a erros devido a ruídos, e e_{artef} a erros devido a artefatos.

Sendo $P_{\{\mathbf{v}_1, \dots, \mathbf{v}_Q\}} \{y_j\}$ a projeção ortogonal de y_j no subespaço gerado pelos vetores $\mathbf{v}_1, \dots, \mathbf{v}_Q$, as componentes da Eq. (3.29) são calculadas da seguinte forma:

$$s_{alvo} = P_{\{(s_j^\tau)_{0 \leq \tau \leq L-1}\}} \{y_j\} \quad (3.30)$$

$$e_{interf} = P_{\{(s_{j'}^\tau)_{1 \leq j' \leq N, 0 \leq \tau \leq L-1}\}} \{y_j\} - s_{alvo} \quad (3.31)$$

$$e_{ruído} = P_{\{(s_{j'}^\tau, \nu_{j'}^\tau)_{1 \leq j' \leq N, 1 \leq i \leq N, 0 \leq \tau \leq L-1}\}} \{y_j\} - P_{\{(s_{j'}^\tau)_{1 \leq j' \leq N, 0 \leq \tau \leq L-1}\}} \{y_j\} \quad (3.32)$$

$$e_{artef} = y_j - P_{(s_{j'}^\tau, \nu_{j'}^\tau)_{1 \leq j' \leq N, 1 \leq i \leq N, 0 \leq \tau \leq L-1}} \{y_j\} \quad (3.33)$$

onde L é o atraso máximo permitido.

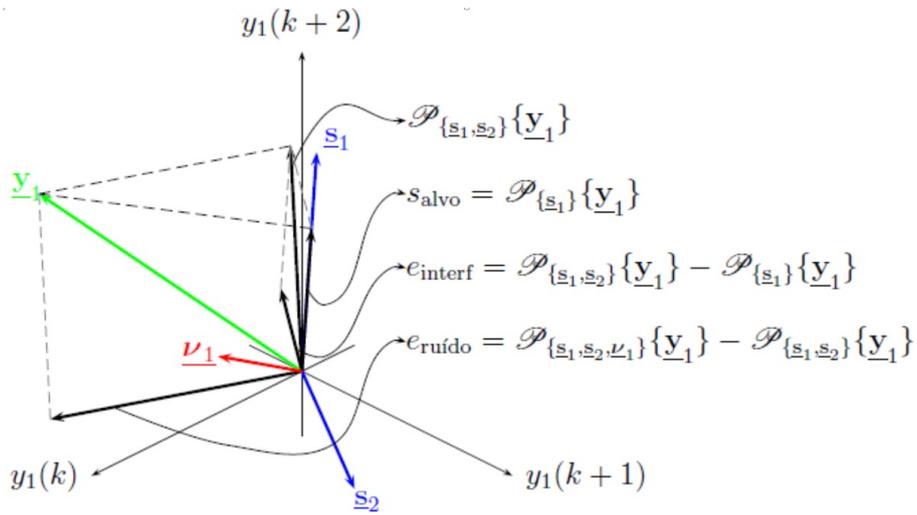
A Fig. 3.4 mostra a decomposição das estimativas das fontes (para o caso 2×2), considerando vetores com três amostras consecutivas \mathbf{y}_j em termos de projeções em subespaços gerados por $\mathbf{s}_1, \mathbf{s}_2$ e ν_j .

Após obtidas as componentes da Eq. (3.29), obtêm-se as medidas de avaliação através das equações:

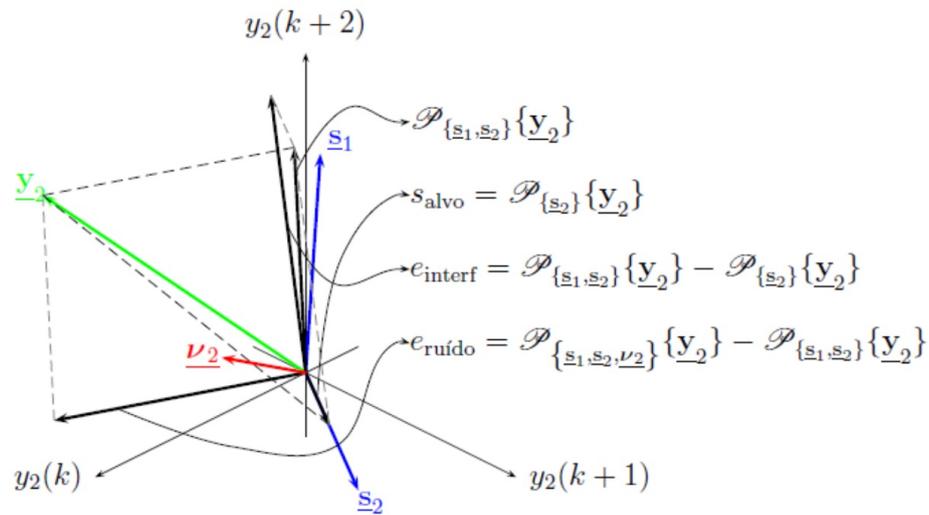
$$\text{SIR} = 10 \log_{10} \frac{\|s_{alvo}\|^2}{\|e_{interf}\|^2} \quad (3.34)$$

$$\text{SDR} = 10 \log_{10} \frac{\|s_{alvo}\|^2}{\|e_{interf} + e_{ruído} + e_{artef}\|^2} \quad (3.35)$$

$$\text{SAR} = 10 \log_{10} \frac{\|s_{alvo} + e_{interf} + e_{ruído}\|^2}{\|e_{artef}\|^2} \quad (3.36)$$



(a)



(b)

Figura 3.4: Ilustração da decomposição, num contexto de misturas instantâneas, do vetor (a) \underline{y}_1 ; (b) \underline{y}_2 , supondo três amostras consecutivas das estimativas das fontes.

Capítulo 4

Algoritmos de Estimação de Direção de Chegada

A localização e o rastreamento de fontes acústicas em ambientes com reverberação são tarefas desafiadoras e a base de diversas aplicações de áudio, conforme visto no Cap. 1. Um exemplo natural típico de sistema de localização de fontes acústicas é o sistema auditivo humano, onde os ouvidos recebem o som e o cérebro “processa” a informação, fazendo a estimação da direção da origem do som. O sistema proposto nesta tese procura fazer a mesma operação através de algoritmos.

Neste trabalho, consideramos que a localização das fontes num ambiente corresponde a estimar suas posições no plano horizontal onde encontram-se tanto os microfones quanto as fontes. São utilizados dois arranjos lineares de microfones, sendo obtidas estimativas das DOAs das fontes nos dois arranjos, conforme ilustrado na Fig. 4.1. A posição da fonte é determinada pelo encontro das retas geradas usando as direções de chegada calculadas, ϕ_1 e ϕ_2 .

Os principais algoritmos de estimação da DOA serão apresentados neste capítulo. Inicialmente, descreveremos o algoritmo GCC-PHAT, apropriado para o caso de haver uma única fonte ativa no ambiente. Em seguida, apresentaremos três métodos propostos para obter as DOAs de múltiplas fontes: os métodos do Campo Distante e SCT, que empregam a técnica de separação cega de fontes no domínio da frequência, e o método TRINICON, que executa a separação no domínio do tempo. Finalmente, propomos alterações nos métodos de BSS com o objetivo de melhorar a acurácia dos métodos de estimação das DOAs quando considerado o seu uso com trechos curtos dos sinais, além de reduzir a complexidade computacional dos métodos.

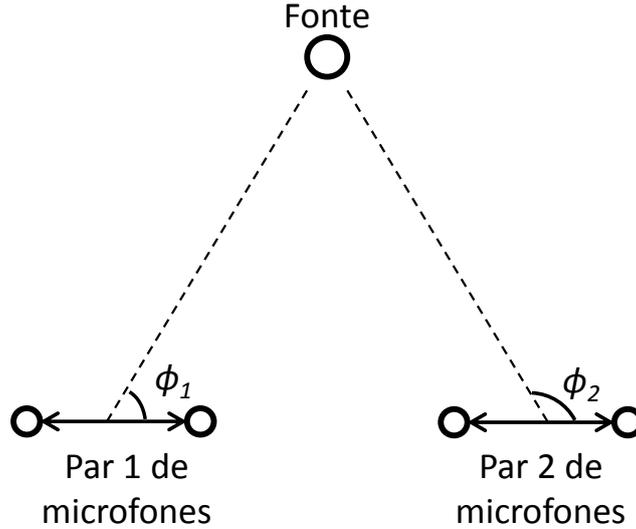


Figura 4.1: Sistema para estimação da posição com 2 pares de microfones.

4.1 Estimativa da DOA com Uma Fonte Ativa

As funções de correlação cruzada generalizada (GCC, do inglês “Generalized Cross-Correlation”) foram propostas para estimação da direção de chegada do sinal de uma fonte observado em um par de microfones [9],[33]. Essas funções são obtidas pela transformada inversa de Fourier da densidade espectral de potência cruzada desses sinais, $S_{x_1x_2}(\omega)$, multiplicada por uma função peso, $\psi(e^{j\omega})$, ou seja:

$$R_{x_1x_2}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \psi(e^{j\omega}) S_{x_1x_2}(\omega) e^{j\omega n} d\omega \quad (4.1)$$

onde

$$S_{x_1x_2}(\omega) = X_1(e^{j\omega})X_2^*(e^{j\omega}) \quad (4.2)$$

sendo $X_j(e^{j\omega})$ a transformada de Fourier de $x_j(n)$ e “*” o operador complexo conjugado.

A quantidade de amostras de atraso entre os sinais nos dois microfones é dada pelo valor de n que maximiza $R_{x_1x_2}(n)$, sendo conhecida como diferença de tempo de chegada (TDOA, do inglês “Time Difference of Arrival”) dos sinais nos microfones¹.

A função peso ψ é utilizada para evitar o aparecimento de picos espúrios em $R_{x_1x_2}(n)$, devido à reverberação e a ruídos. Escolhendo-se a função peso igual a 1, estaremos utilizando a correlação cruzada clássica, apropriada apenas em casos ideais (ambientes não-reverberantes e sem ruídos). O algoritmo denominado SCOT-GCC utiliza a função custo indicada na Eq. (4.3) abaixo. Já o algoritmo GCC-

¹Para aumentar a resolução da estimativa da TDOA, pode-se interpolar os sinais das misturas $x_i(n)$ antes de se calcular a função correlação cruzada.

PHAT emprega a função custo indicada na Eq. (4.4). Utilizando-se um desses dois algoritmos, obtêm-se bons resultados em diversas situações práticas, como em salas com nível de reverberação moderado e cabines de carros.

$$\psi(e^{j\omega}) = \frac{1}{\sqrt{S_{x_1x_1}(\omega) S_{x_2x_2}(\omega)}} \quad (4.3)$$

$$\psi(e^{j\omega}) = \frac{1}{|S_{x_1x_2}(\omega)|} \quad (4.4)$$

Um dos empregos desses algoritmos é na determinação da DOA, a partir da TDOA, para o projeto de filtros espaciais (*beamformers*) empregados em diversos sistemas em diferentes áreas. Assumindo-se a condição de campo distante, em que a distância dos microfones é pequena quando comparada à distância entre os microfones e a fonte, pode-se considerar que a frente de onda que chega aos dois microfones é plana, conforme mostra a Fig. 4.2. Assim, o ângulo ϕ correspondente à DOA pode

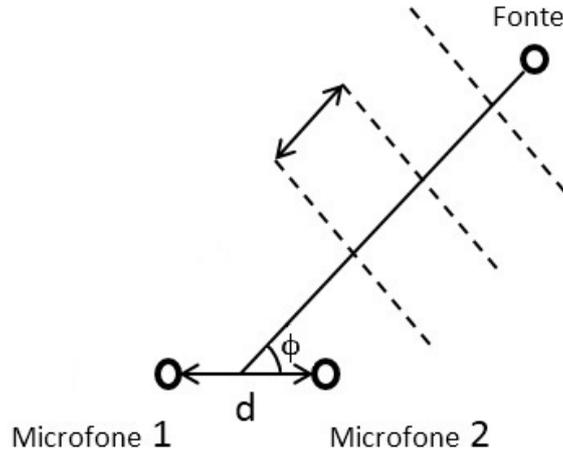


Figura 4.2: Frente de onda na condição de campo distante.

ser obtido a partir da diferença do tempo de chegada das ondas nos microfones, Δt , pela relação

$$\phi = \arccos \frac{v\Delta t}{d} \quad (4.5)$$

onde v é a velocidade de propagação da onda e d é a distância entre os microfones.

A Fig. 4.3 ilustra o desempenho do algoritmo GCC-PHAT para um trecho de sinal de voz, mostrado, junto com a sua versão atrasada de 14 amostras, na Fig. 4.3(a). Na Fig. 4.3(b) encontra-se a função correlação cruzada dos dois sinais, enquanto que na Fig. 4.3(c) encontra-se a função correlação cruzada generalizada do algoritmo GCC-PHAT. Pode-se observar que $R_{x_1x_2}(n)$ se aproxima da função impulso, com máximo em n igual a 14, quando a função peso da Eq. (4.4) é utilizada.

Variações no método GCC-PHAT foram propostas para estimar a TDOA nos

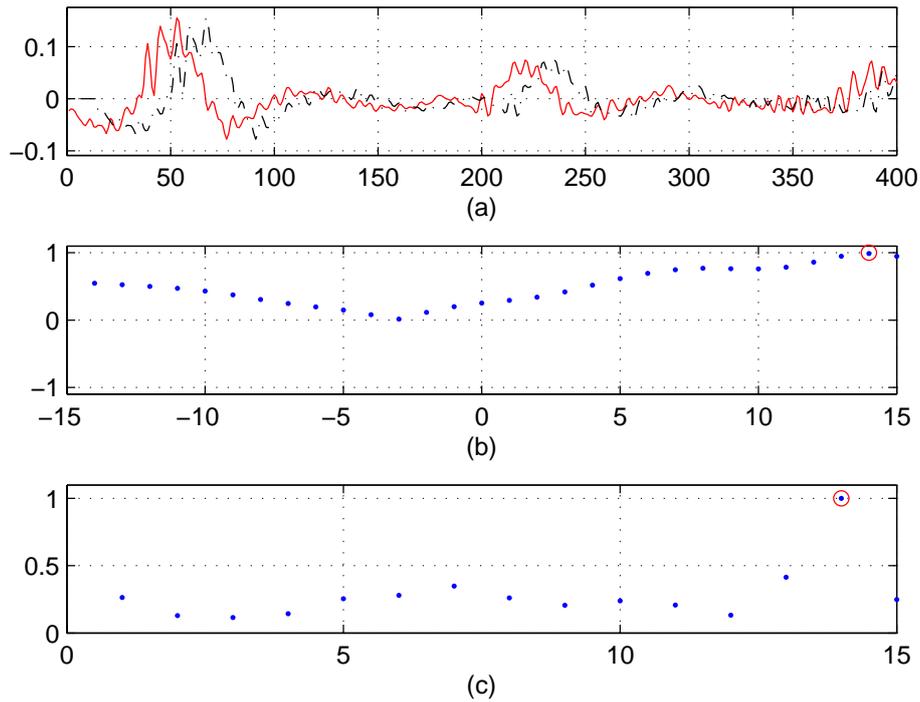


Figura 4.3: Exemplo ilustrativo do desempenho do algoritmo GCC-PHAT. (a) trecho de sinal de voz e sua versão atrasada de 14 amostras; (b) correlação cruzada dos sinais; (c) GCC-PHAT.

casos de fontes móveis e múltiplas fontes, as quais não recorrem à separação dos sinais [26],[30],[39],[72]. No caso com múltiplas fontes, há a tendência de ocorrerem picos da função correlação cruzada generalizada em valores equivalentes aos atrasos dos sinais das diferentes fontes nos dois microfones. Assim, pode-se estimar a direção de cada uma delas através da detecção dos seus picos. Outra estratégia é trabalhar com blocos de amostras assumindo esparsidade das fontes no tempo. Assim, haverá momentos em que uma das fontes estará mais forte, causando um pico na função correlação cruzada no instante correspondente à diferença de tempo de chegada do seu sinal nos microfones. A ocorrência dos picos pode ser avaliada estatisticamente e usada para auxiliar na tarefa de obtenção da direção de chegada no caso com múltiplas fontes e de fontes móveis.

Como o método GCC-PHAT é aplicado aos sinais das misturas, seu desempenho é degradado à medida que a quantidade de fontes ativas aumenta. A intensidade da reverberação também afeta a acurácia do algoritmo. Portanto, o método GCC-PHAT será utilizado neste trabalho apenas para estimação da DOA em casos em que há apenas uma fonte ativa.

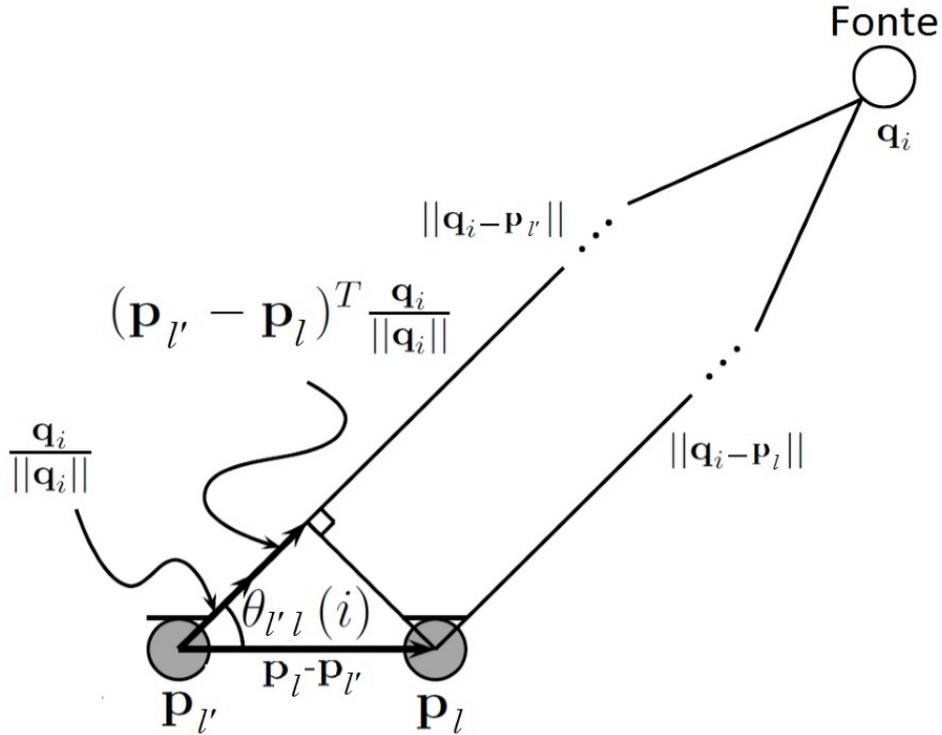


Figura 4.4: Modelo de campo distante.

4.2 Estimativa da DOA com Múltiplas Fontes Ativas

Apresentaremos nesta seção técnicas de estimação das DOAs de múltiplas fontes que utilizam a matriz de separação \mathbf{W} obtida por algoritmos de BSS de misturas convolutivas.

4.2.1 Método do Campo Distante

Uma forma de encontrar as direções de chegada para múltiplas fontes, proposta em [48], utiliza, além da matriz de separação obtida por BSS, informações geométricas, conforme ilustrado na Fig. 4.4. É adotado o modelo de campo distante, que considera a distância entre os sensores (com posições especificadas pelos vetores \mathbf{p}_l e $\mathbf{p}_{l'}$ na Fig. 4.4) muito pequena quando comparada às distâncias entre a fonte (com posição especificada por \mathbf{q}_i) e os sensores. Assim, é admitido que os percursos diretos pelos quais o sinal passa até chegar aos sensores tendem a ser paralelos e que as atenuações sofridas pelos sinais são praticamente iguais. A única diferença significativa é na variação de tempo, Δt , em que os sinais atingem os sensores, influenciando na diferença de fase dos sinais. Esse parâmetro depende do ângulo de incidência do sinal, $\theta_{l'l}(i)$, da distância entre os sensores e da frequência do sinal.

A relação entre as respostas ao impulso dos filtros correspondentes aos dois ca-

minhos, obtida com as aproximações acima, é dada por:

$$h_{l'i}(t) = h_{li}(t + \Delta t) \quad (4.6)$$

Analisando a diferença de fase dos filtros no domínio da frequência, obtém-se:

$$\frac{H_{li}(k)}{H_{l'i}(k)} \approx e^{j \frac{2\pi f_k (\mathbf{p}_{l'} - \mathbf{p}_l)^T \mathbf{q}_i}{v \|\mathbf{q}_i\|}} = e^{j \frac{2\pi f_k d_{l'} \cos(\theta_{l'}(i))}{v}} \quad (4.7)$$

e, portanto,

$$\theta_{l'}(i) = \arccos \left(\frac{\arg \left(\frac{H_{li}(k)}{H_{l'i}(k)} \right)}{2\pi f_k v^{-1} d_{l'}} \right) \quad (4.8)$$

onde v é a velocidade de propagação do som no ar e $d_{l'}$ é a distância entre os sensores.

A Eq. (4.8) pode ser usada para estimar a DOA, uma vez que $H_{li}(k)$ e $H_{l'i}(k)$ podem ser obtidos através da inversa da matriz $\mathbf{W}(k)$, calculada por ICA no domínio da frequência. Dessa forma, encontram-se estimativas do ângulo de chegada para todas as raias de frequência e, em seguida, avalia-se o número de ocorrências de cada valor obtido, através de um histograma. O valor correspondente ao pico do histograma será a estimativa final de $\theta_{l'}(i)$.

Esse método apresenta alguns problemas. Inicialmente é possível verificar que as aproximações acima levam a erros nas estimativas, uma vez que o canal de propagação das ondas acústicas não é formado apenas por um percurso direto, conforme visto na Seção 2.1, e as matrizes $\mathbf{W}(k)$ são obtidas através da ICA, a qual não é precisa em todos os casos. Outra restrição desse método é que ele só deve ser utilizado para frequências que satisfazem:

$$f < \left| \frac{v}{2d_{l'}} \right| \quad (4.9)$$

devido ao “aliasing” espacial.

4.2.2 Método SCT

Em [11] e [12], é proposta a utilização da transformada SCT (do inglês “State Coherence Transform”), que utiliza a coerência da fase da matriz de separação obtida pela ICA e possibilita gerar “picos prominentes” na envoltória nos pontos correspondentes aos TDOAs.

A proposta de estimativa da TDOA parte do princípio de que, em condições

ideais, a matriz de mistura é modelada por:

$$\mathbf{H}(k) = \begin{bmatrix} |H_{11}(k)| e^{-j\varphi_{11}(k)} & |H_{12}(k)| e^{-j\varphi_{12}(k)} \\ |H_{21}(k)| e^{-j\varphi_{21}(k)} & |H_{22}(k)| e^{-j\varphi_{22}(k)} \end{bmatrix} \quad (4.10)$$

onde $\varphi_{iq}(k) = 2\pi f_k T_{iq}$, sendo T_{iq} o tempo de propagação da q -ésima fonte ao i -ésimo microfone e f_k a frequência em Hz associada ao k -ésimo *bin*.

A matriz $\mathbf{H}(k)$ é obtida pela inversa da matriz de separação, ou seja, $\mathbf{H}(k) = \mathbf{W}^{-1}(k)$. Com a condição de campo distante, tem-se:

$$r_1(k) = \frac{H_{12}(k)}{H_{22}(k)} \approx e^{-j2\pi f_k \Delta t_1} \quad (4.11)$$

$$r_2(k) = \frac{H_{11}(k)}{H_{21}(k)} \approx e^{-j2\pi f_k \Delta t_2} \quad (4.12)$$

onde Δt_1 e Δt_2 são os TDOAs das fontes. Cada desvio de fase depende da frequência e da TDOA, assim pode ser considerado como um estado associado a cada fonte. Em condições reais de propagação, o sinal sofre distorção devido à reverberação, refletindo nos resultados de $r_i(k)$.

Assumindo a condição de campo distante e um ambiente anecóico, a propagação da fonte entre microfones no domínio da frequência, para TDOA de τ , pode ser representada por:

$$c(k, \tau) = e^{-j2\pi f_k \tau} \quad (4.13)$$

A TDOA da i -ésima fonte pode ser estimada através da minimização da distância euclidiana entre o modelo ideal e $r_i(k)$, conforme indicado em:

$$\overline{\Delta t}_i = \underset{\tau}{\operatorname{argmin}} \sum_k \|c(k, \tau) - \bar{r}_i(k)\| \quad (4.14)$$

onde $\bar{r}_i(k) = \frac{r_i(k)}{\|r_i(k)\|}$.

Devido à permutação entre raias, não é possível saber a qual fonte pertence cada TDOA. Portanto, não é possível estabelecer a TDOA de uma fonte utilizando diretamente a Eq. (4.14). Porém, $\varphi_{iq}(k)$ da Eq. (4.10) mostra que podem ser feitas as estimativas de múltiplas TDOAs com uma transformada adequada, a qual utiliza conjuntamente os estados associados a todas as fontes, e assim o sistema se torna invariante à permutação. A transformada proposta é a SCT [11], dada por:

$$SCT(\tau) = \sum_k \sum_{i=1}^N \left[1 - g \left(\frac{\|c(k, \tau) - \bar{r}_i(k)\|}{2} \right) \right] \quad (4.15)$$

onde N é o número de estados observados e $g(\cdot)$ é uma função não-linear. Em [11],[12] é mostrado que, para $N = 2$ e escolhendo $g(x) = x$, a SCT será maximizada por valores de τ correspondentes às máximas verossimilhanças dos TDOAs das fontes.

4.2.2.1 Aplicação da SCT no Caso Subdeterminado

A SCT é apropriada apenas para o caso em que o número de fontes é menor ou igual ao número de sensores. Para o caso subdeterminado, em que o número de fontes é maior que o número de sensores, é utilizada a cSCT (do inglês “Cumulated SCT”).

Aplicando BSS no caso subdeterminado, a ICA fornecerá estimativas das duas fontes dominantes, de acordo com a esparsidade das fontes. Portanto, torna-se conveniente dividir o sinal em blocos no domínio do tempo, de modo que a ICA possa atuar nas diversas frequências e blocos. Com isso, faz-se necessária a extensão do modelo da SCT, incluindo a informação dos blocos de tempo e gerando:

$$cSCT(\tau) = \sum_b \sum_k \sum_{i=1}^N \left[1 - g \left(\frac{\|c(k, \tau) - \bar{r}_i^b(k)\|}{2} \right) \right] \quad (4.16)$$

onde $\bar{r}_i^b(k)$ é o estado normalizado obtido pelo k -ésimo bin de frequência no bloco b .

4.2.3 Método TRINICON

A estimativa da TDOA usando o algoritmo TRINICON, descrito na Seção 3.3, é obtida através dos filtros do sistema de separação \mathbf{W} , utilizando o princípio de que é possível identificar a TDOA a partir da diferença de tempo entre os picos dos filtros w_{pq} . Assim, as estimativas das TDOAs, para o caso de duas fontes e dois sensores, são obtidas por [10]:

$$\hat{\tau}_1 = \arg \max_n |w_{12}(n)| - \arg \max_n |w_{22}(n)| \quad (4.17)$$

$$\hat{\tau}_2 = \arg \max_n |w_{11}(n)| - \arg \max_n |w_{21}(n)| \quad (4.18)$$

Pode-se observar das equações acima que a exata estimativa de todos os coeficientes dos filtros de separação \mathbf{W} não é necessária para obter localização satisfatória, sendo necessário apenas identificar os coeficientes correspondentes aos caminhos diretos de propagação.

4.3 Modificações nos Métodos de BSS para Utilização na Estimação das DOAs

Nesta seção, são propostas modificações nos algoritmos de separação cega de fontes de modo a torná-los mais eficientes para a aplicação de estimação das DOAs. Os algoritmos de BSS convencionais foram desenvolvidos visando à separação dos sinais das fontes, ou seja, à minimização da interferência dos sinais das demais fontes na estimativa do sinal de uma dada fonte.

Devido à presença de reverberação em recintos fechados, é necessária a utilização de filtros longos para modelar os caminhos percorridos pelos sinais das fontes até os microfones, bem como para desfazer os efeitos das misturas dos sinais. Consequentemente, a obtenção dos coeficientes dos filtros de separação requer um número elevado de amostras dos sinais, de modo que se obtenha um bom resultado para o problema da separação. Como neste trabalho o foco principal não é a separação das fontes, mas as estimativas das DOAs a partir dos parâmetros do sistema de separação, apresentaremos alterações nos algoritmos de BSS com o objetivo de produzir estimativas mais acuradas das DOAs em um período de tempo mais curto.

4.3.1 Algoritmo BSS no Domínio da Frequência

O espectrograma de um sinal de voz típico é mostrado na Fig. 4.5. Nessa figura, observa-se que os sinais de fala têm potência significativa ao longo de um vasto conjunto de frequências, e que exibem componentes de frequências que apresentam maior potência quando comparadas às de frequências adjacentes. Na Fig. 4.5, tais frequências aparecem como faixas horizontais no espectrograma e correspondem à frequência fundamental, harmônicos e formantes. Uma vez que as componentes das misturas correspondentes a essas frequências terão potência significativa, é esperado que sejam as mais apropriadas para a execução das estimativas das DOAs pelos métodos que utilizam os resultados do algoritmo BSS no domínio da frequência. As componentes de frequência de baixa energia terão maior tendência a apresentarem efeitos de ruídos do ambiente e de medição.

Propomos, então, utilizar apenas as componentes de maior potência dos sinais de mistura nos métodos do Campo Distante e SCT. Assim, após executarmos as DFTs de tamanho K do m -ésimo *frame* dos sinais de mistura, para obter seus coeficientes no domínio de frequências $X_i(f_k, m)$, selecionamos os N_p coeficientes que apresentam os maiores valores de $\min\{|X_i(f_k, m)|\}$, ou seja, definimos um subconjunto de amostras das misturas no domínio da frequência correspondentes aos índices

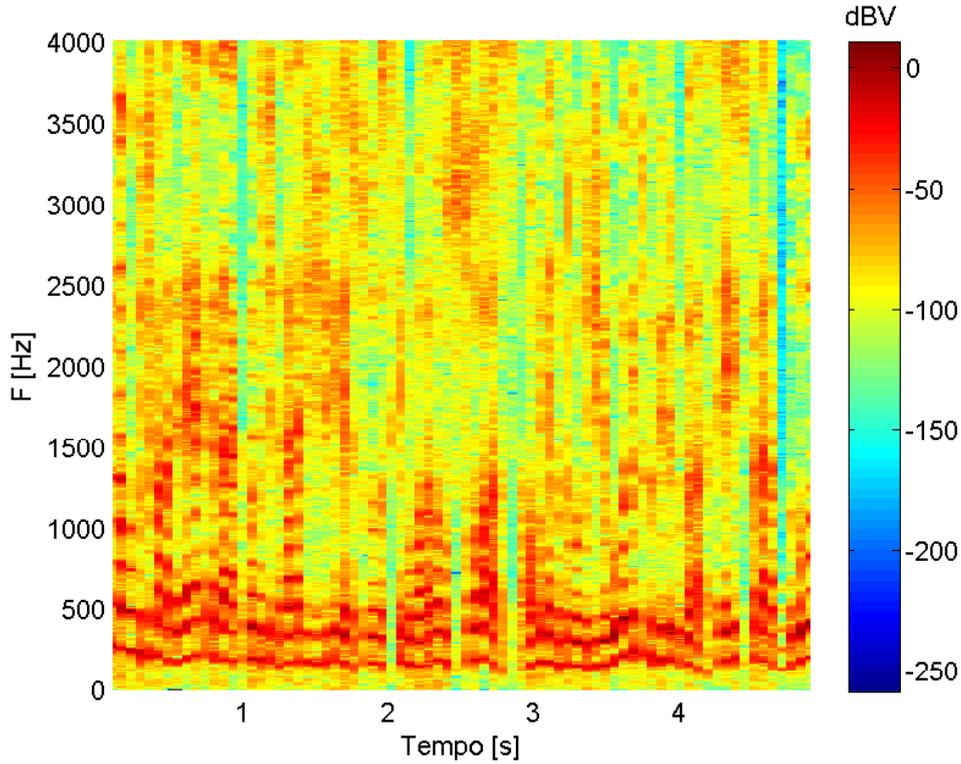


Figura 4.5: Exemplo de espectrograma de sinal de voz.

$\{k_{i_1}, \dots, k_{i_{N_p}}\}$ das frequências com maior energia:

$$\tilde{X}_i(f_l, m) = X_i(f_{k_l}, m), \quad l = 1, \dots, N_p \quad e \quad i = 1, \dots, N \quad (4.19)$$

Utilizamos, então, apenas as N_p amostras de $X_i(f_k, m)$ no algoritmo BSS descrito na Seção 3.2.1. Com essa alteração, além de se obter uma redução significativa da complexidade computacional dos algoritmos quando $N_p \ll K$, estimativas das DOAs com maior acurácia são desenvolvidas em um menor intervalo de tempo. No Cap. 5, será investigada a escolha adequada do parâmetro N_p . Outra possível abordagem, ao invés de fixar N_p , seria definir um limiar para a potência dos sinais em cada *bin* de frequência, selecionando apenas os coeficientes das misturas no domínio da frequência com energia acima desse limiar e rejeitando as demais.

4.3.2 Algoritmo TRINICON

Se considerarmos apenas um intervalo curto de tempo, o sinal de voz não contém informação suficiente para excitar todos os modos naturais dos filtros que compõem o sistema de separação das fontes. Isso se deve ao fato de que filtros de ordens altas são necessários para a obtenção de bons resultados, principalmente em ambientes com alta reverberação, e de que o sinal de voz é predominantemente formado por com-

ponentes em torno da frequência fundamental, harmônicos e formantes, conforme ilustrado na Fig. 4.5. No exemplo de separação de duas fontes em um ambiente com 150 ms de tempo de reverberação apresentado em [55], foram necessários 10 s dos sinais de fala para obter SIR em torno 20 dB, sendo utilizados filtros de comprimento $L = 1024$. Esse elevado tempo de atuação do algoritmo não se mostra apropriado para as aplicações consideradas nesta tese.

O algoritmo de estimação da DOA baseado no algoritmo TRINICON utiliza apenas os valores máximos dos coeficientes de cada filtro do sistema de separação. Portanto, não é necessária a obtenção dos valores ótimos de todos os coeficientes desses filtros, mas apenas o índice correto do coeficiente de maior valor de cada um deles.

Como o objetivo desse trabalho é a localização de múltiplas fontes, e não a separação dos seus sinais, propomos alterar o algoritmo de forma a obter estimativas mais acuradas das DOAs quando um trecho pequeno do sinal (poucos segundos) é utilizado. Para isso, empregamos o mesmo princípio de distribuição da energia de atualização entre coeficientes aplicado no algoritmo de filtragem adaptativa PNLMS (do inglês “Proportionate Normalized Least Mean Square”). Nesse algoritmo, os coeficientes com maiores valores (em módulo) são atualizados mais rapidamente, ou seja, um passo de adaptação proporcional ao valor do coeficiente é usado na equação de atualização. Essa técnica foi desenvolvida para aumentar a velocidade de convergência do algoritmo NLMS quando utilizado na modelagem de sistemas com respostas ao impulso esparsas. A convergência para sistemas não esparsos é prejudicada.

Com a alteração proposta, o termo de atualização das matrizes de coeficientes dos filtros de separação em banda larga para o caso com duas fontes e duas misturas é dado por:

$$\Delta \mathbf{W} = \sum_{i=0}^{\infty} \beta(i, m) \begin{bmatrix} \Lambda_{11} \mathbf{W}_{12} \tilde{\mathbf{R}}_{y_1 y_2}(i) & \Lambda_{12} \mathbf{W}_{11} \tilde{\mathbf{R}}_{y_1 y_2}(i) \\ \Lambda_{21} \mathbf{W}_{22} \tilde{\mathbf{R}}_{y_1 y_2}(i) & \Lambda_{22} \mathbf{W}_{21} \tilde{\mathbf{R}}_{y_1 y_2}(i) \end{bmatrix} \quad (4.20)$$

onde

$$\Lambda_{kl} = \frac{\Gamma_{kl}}{\mathbf{x}_l^T(i) \Gamma_{kl} \mathbf{x}_l(i) + \delta} \quad (4.21)$$

$$\Gamma_{kl} = \text{diag}\{g_{kl}(0), \dots, g_{kl}(L-1)\} \quad (4.22)$$

$$g_{kl}(i) = \frac{\gamma_{kl}(i)}{\frac{1}{L} \sum_{j=0}^{L-1} \gamma_{kl}(j)} \quad (4.23)$$

$$\gamma_{kl}(i) = \max\{\gamma_{kl}^{\min}, |w_{kl}(i)|\} \quad (4.24)$$

$$\gamma_{kl}^{\min} = \rho \max\{\delta_p, |w_{kl}(0)|, \dots, |w_{kl}(L-1)|\} \quad (4.25)$$

Valores típicos dos parâmetros são $\delta_p = \delta = 0.01$ e $\rho = 0.01$. Deve-se observar que, durante o processo de atualização dos coeficientes, é necessário garantir a estrutura *Sylvester* da matriz $\mathbf{W}(m+1)$ (ver Seção 3.2.2). Portanto, na prática, atualizam-se apenas os L primeiros elementos da primeira coluna de cada matriz \mathbf{W}_{kl} (ver Eq. (3.22)).

No Cap. 7 serão avaliados os efeitos das modificações propostas na estimação da DOA para ambientes com baixa e alta reverberações.

Capítulo 5

Sistema de Localização de Fontes e Determinação da Região de Localização Robusta

Neste capítulo serão apresentados o sistema de localização de múltiplas fontes proposto e as condições para torná-lo robusto.

O diagrama da Fig. 5.1 apresenta uma visão geral do problema de localização e da solução proposta para o caso de duas fontes. O sistema de localização é composto por dois arranjos de microfones, que capturam os sinais das fontes misturados e modificados pelo “Sistema de mistura”, o qual pode ser modelado por um conjunto de filtros FIR correspondentes aos canais acústicos de cada par fonte-microfone, conforme descrito no Cap. 2. Neste trabalho, assume-se que cada conjunto de microfones contém ao menos um número de elementos igual ao de fontes ativas, de modo a utilizar um dos algoritmos BSS desenvolvido para o caso determinado descrito no Cap. 3. A partir dos parâmetros do algoritmo BSS empregado, determinam-se as direções de chegada de cada fonte em relação a cada conjunto de microfones. Finalmente, as posições das fontes são obtidas por meio do encontro das retas traçadas a partir das estimativas das DOAs dos sinais nos dois conjuntos de sensores, conforme mostrado na Fig. 5.2. Apesar de considerarmos nos experimentos apenas o caso de duas fontes simultaneamente ativas e de arranjos com dois microfones, o sistema pode ser utilizado para um maior número de fontes, aumentando-se o número de microfones de cada arranjo.

Nesta tese, nos concentraremos na aplicação do sistema a ambientes interativos inteligentes, como salas de aula à distância e teleconferências, nos quais os locutores estão todos voltados para a parede da sala onde se encontram a tela e os microfones. Nesse caso, ilustrado na Fig. 5.2, apenas a distância entre os conjuntos de microfones

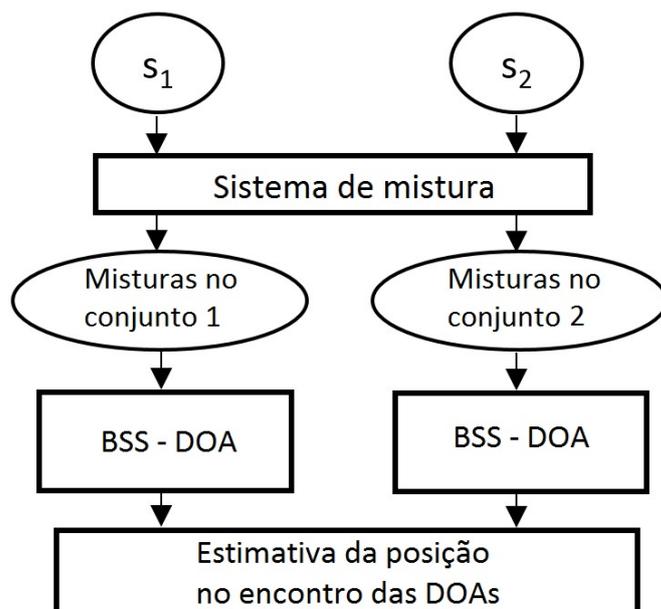


Figura 5.1: Sistema de localização baseado em BSS.

é variada¹.

As condições para o sistema ser robusto dependem da acurácia nas estimativas das DOAs, do erro máximo admitido nas estimativas das posições, do posicionamento dos dois conjuntos de microfones e da região de atuação do sistema de localização. As relações entre esses quatro fatores e a robustez do sistema serão analisadas neste capítulo, com o objetivo de identificar a região em que a localização será bem sucedida, escolher de forma otimizada a distância entre os dois arranjos de microfones e verificar a acurácia mínima nas estimativas das DOAs necessária para que a localização seja bem sucedida.

5.1 Fundamentos Teóricos para Estimar a Região de Localização Robusta

O objetivo desta seção é propor um modelo determinístico para obter a região apropriada para a localização das fontes e auxiliar na escolha da posição dos dois conjuntos de microfones/sensores, a partir da acurácia nas estimativas das DOAs e do erro máximo admitido nas estimativas das posições.

A Fig. 5.2 mostra os elementos presentes no sistema de localização para cada fonte: a fonte na posição P_F ; os dois conjuntos de sensores, M_A e M_B ; as duas retas correspondentes às DOAs reais nos conjuntos de sensores M_A e M_B ; as duas retas correspondentes às DOAs estimadas com maiores erros para cada conjunto de

¹O método e os resultados serão apresentados para uma sala de comprimento e largura iguais a 10 metros, sem perda de generalidade.

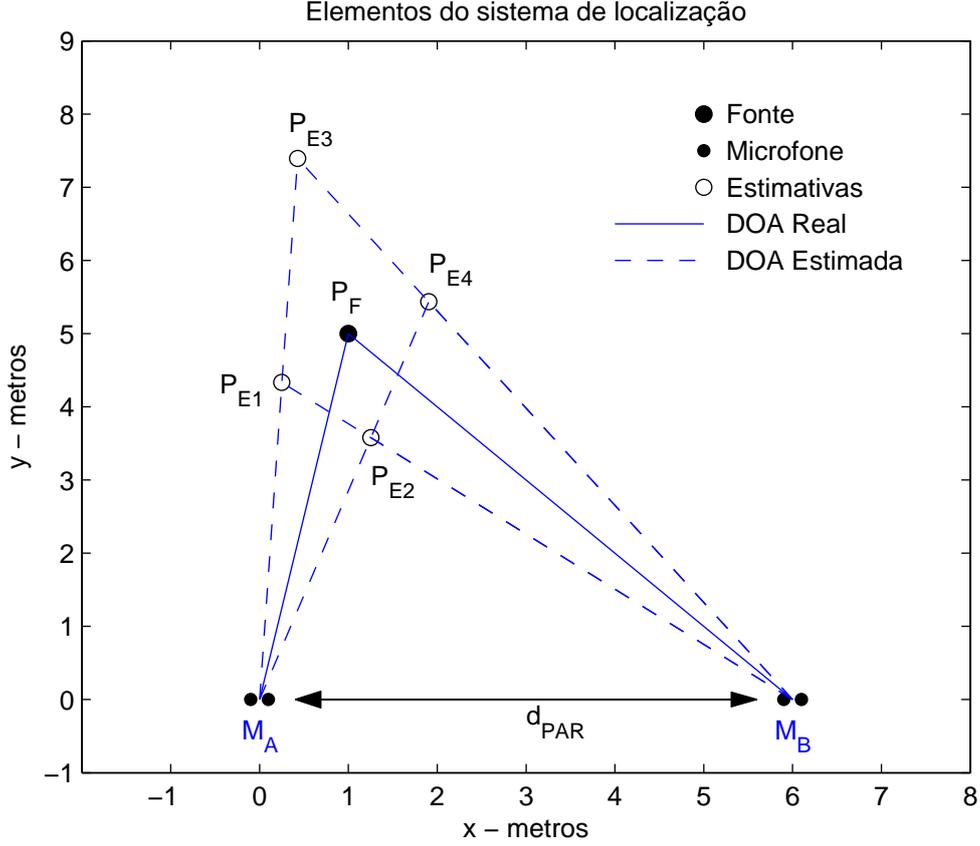


Figura 5.2: Elementos do sistema de localização.

microfones; as quatro estimativas da posição P_F , dadas por P_{E_i} , para $i = 1, \dots, 4$; a distância entre os conjuntos de sensores d_{PAR} .

Sendo os ângulos θ_{AR} e θ_{BR} correspondentes às DOAs reais nos conjuntos de sensores M_A e M_B , e admitindo-se um erro máximo $\Delta\theta$ em suas estimativas, temos as retas correspondentes às DOAs estimadas com maiores erros, caracterizadas pelos ângulos $\theta_{A1} = \theta_{AR} + \Delta\theta$, $\theta_{A2} = \theta_{AR} - \Delta\theta$, $\theta_{B1} = \theta_{BR} + \Delta\theta$, $\theta_{B2} = \theta_{BR} - \Delta\theta$, sendo as posições P_{E_n} , para $n = 1, \dots, 4$, determinadas pelo encontro dessas retas. Os seus coeficientes angulares são, portanto, $m_{AR} = \tan(\theta_{AR})$, $m_{BR} = \tan(\theta_{BR})$, $m_{A1} = \tan(\theta_{A1})$, $m_{A2} = \tan(\theta_{A2})$, $m_{B1} = \tan(\theta_{B1})$ e $m_{B2} = \tan(\theta_{B2})$. Os erros das posições estimadas P_{E_n} em relação à posição real P_F são dados por:

$$D_{E_n} = |\mathbf{P}_{E_n} - \mathbf{P}_F| \quad (5.1)$$

sendo $|\cdot|$ a norma euclidiana, e \mathbf{P}_{E_n} e \mathbf{P}_F os vetores de coordenadas dos respectivos pontos. Esses vetores podem ser expressos como:

$$\mathbf{P}_{E_1} = [P_x(m_{A1}, m_{B1}), P_y(m_{A1}, m_{B1})]^T \quad (5.2)$$

$$\mathbf{P}_{E_2} = [P_x(m_{A_1}, m_{B_2}), P_y(m_{A_1}, m_{B_2})]^T \quad (5.3)$$

$$\mathbf{P}_{E_3} = [P_x(m_{A_2}, m_{B_1}), P_y(m_{A_2}, m_{B_1})]^T \quad (5.4)$$

$$\mathbf{P}_{E_4} = [P_x(m_{A_2}, m_{B_2}), P_y(m_{A_2}, m_{B_2})]^T \quad (5.5)$$

$$\mathbf{P}_F = [P_x(m_{A_R}, m_{B_R}), P_y(m_{A_R}, m_{B_R})]^T \quad (5.6)$$

Considerando M_A na posição $(0, 0)$ e M_B na posição $(d_{PAR}, 0)$, as posições estimadas e a posição real das fontes são obtidas através de:

$$P_x(m_A, m_B) = \frac{-m_B d_{PAR}}{m_A - m_B} \quad (5.7)$$

$$P_y(m_A, m_B) = m_A \cdot \frac{-m_B d_{PAR}}{m_A - m_B} \quad (5.8)$$

sendo m_A e m_B os coeficientes angulares das respectivas retas.

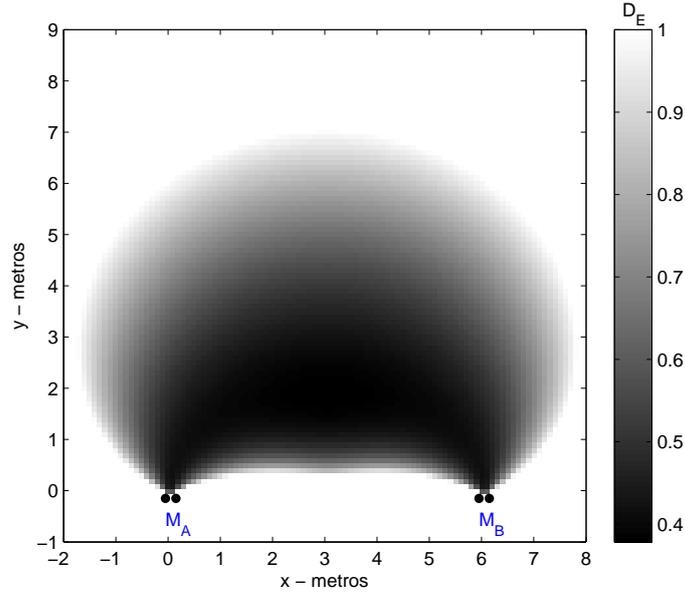
A partir das equações acima pode-se calcular o valor médio do erro, D_E , para cada posição no plano (x, y) , através de:

$$D_E = 0.25 \sum_{n=1}^4 D_{E_n} \quad (5.9)$$

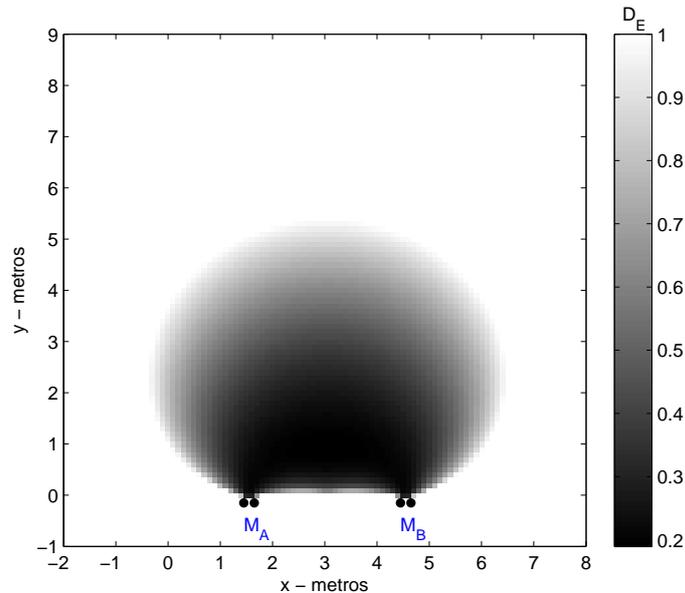
Assim, é possível identificar as posições onde D_E é menor que o erro máximo admitido, D_{MAX} , sendo essas consideradas apropriadas na região pesquisada.

A Fig. 5.3 mostra a região estabelecida como apropriada no plano (x, y) numa sala de comprimento e largura iguais a 10 m, onde calculamos o D_E em cada posição, para erro máximo admitido nas estimativas das DOAs, $\Delta\theta$, igual a 4° e distância entre os pares de microfones d_{PAR} igual a 6 m no Caso A (Fig. 5.3(a)) e a 3 m no Caso B (Fig. 5.3(b)). As posições onde D_E é menor que D_{MAX} formam a região apropriada para a localização, sendo estas representadas num mapa de cores em tons de cinza, e as posições onde D_E é maior que D_{MAX} representadas em branco.

Analisando a Fig. 5.3(a), verifica-se que as posições com y maior ou igual a 7 m são consideradas inapropriadas para localização e que a área apropriada representa 54.16% da área total da sala no Caso A, ou 64.16% se excluídas as posições com y entre -1 e 0. Da Fig. 5.3(b), observa-se que as posições com y maior que 5 são consideradas inapropriadas e que a área apropriada representa 30.10% da área total da sala no Caso B, ou 40.10% se excluídas as posições com y entre -1 e 0. Portanto, conclui-se que, ao diminuir a distância entre os pares, houve uma redução



(a)



(b)

Figura 5.3: Estimativa da região de localização apropriada para (a) $d_{PAR} = 6$ m e (b) $d_{PAR} = 3$ m, sendo $\Delta\theta = 4^\circ$ e $D_{MAX} = 1$ m.

das posições consideradas apropriadas para se fazer a localização.

Podemos tecer alguns comentários sobre os parâmetros d_{PAR} , $\Delta\theta$ e D_{MAX} , que influenciam na determinação da região apropriada de localização:

- d_{PAR} - deve ser escolhido de forma a otimizar o desempenho do sistema de localização: se muito grande, as estimativas das posições, representadas por P_{E_n} na Fig. 5.2, podem ficar muito distantes da P_F , gerando grandes erros de localização; se muito pequeno, as DOAs tendem a ficar paralelas e, consequen-

temente, pequenos erros nas estimativas das DOAs podem provocar grandes erros nas estimativas das posições;

- $\Delta\theta$ - simula o erro introduzido pelo método utilizado para estimar as DOAs, sendo desejável que este seja o menor possível para aumentar a região de localização robusta; porém, pode admitir maiores valores, viabilizando o emprego do sistema em ambientes com maior reverberação, se utilizado com d_{PAR} otimizado;
- D_{MAX} - está associado ao erro máximo admitido pela aplicação.

A Fig. 5.4(a) mostra a variação da área percentual considerada apropriada, formada pelas posições em que $D_E < D_{MAX}$, em tons de cinza onde o mais claro representa 100% e o mais escuro 0%, para erros máximos nas estimativas das DOAs ($\Delta\theta$) entre 1° e 10° e para distância entre os conjuntos de microfones (d_{PAR}) entre 1 e 10 m, considerando o erro máximo admitido na localização (D_{MAX}) igual a 1 m. A Fig. 5.4(b) mostra o comportamento do valor médio da área percentual considerada apropriada ao variar d_{PAR} , para $D_{MAX} = 1$ m, considerando erros nas estimativas das DOAs entre² 1° e 7° , ou seja,

$$R_m(d_{PAR}) = \frac{1}{7} \sum_{D_{MAX}=1}^1 \sum_{\Delta\theta=1^\circ}^{7^\circ} A_{aprop}(d_{PAR}, \Delta\theta, D_{MAX}) \quad (5.10)$$

onde $A_{aprop}(d_{PAR}, \Delta\theta, D_{MAX})$ é a área percentual considerada apropriada dada por

$$A_{aprop}(d_{PAR}, \Delta\theta, D_{MAX}) = 100 \cdot \frac{A_{D_E < D_{MAX}}}{A_{TOTAL}} [\%] \quad (5.11)$$

sendo $A_{D_E < D_{MAX}}$ a área da região formada pelas posições onde $D_E < D_{MAX}$ e A_{TOTAL} a área total pesquisada.

As Figs. 5.4(c) e 5.4(d) mostram os gráficos correspondentes para D_{MAX} igual a 0.5 m.

Analisando a influência do erro de estimação das DOAs nas áreas consideradas apropriadas para localização A_{aprop} nas Figs. 5.4(a) e 5.4(c), observa-se que quanto menor $\Delta\theta$, maior é a região, correspondendo a tons de cinza mais claros.

Analisando o comportamento da área apropriada com a variação d_{PAR} na Fig. 5.4(a), verifica-se que para $\Delta\theta$ entre 1° e 4° , o aumento de d_{PAR} permite o aumento de A_{prop} ; para 5° , o aumento da distância até 7 m aprimora a região de performance satisfatória do sistema, sendo A_{prop} constante para d_{PAR} entre 7 m e 9 m, e diminuindo para d_{PAR} igual a 10; para 6° , o sistema apresenta a melhor performance para

²Esses valores de $\Delta\theta$ foram escolhidos de acordo com os resultados experimentais (que serão exibidos nos Capítulos 6 e 7), os quais apresentaram quantidade significativa de erros nesse intervalo.

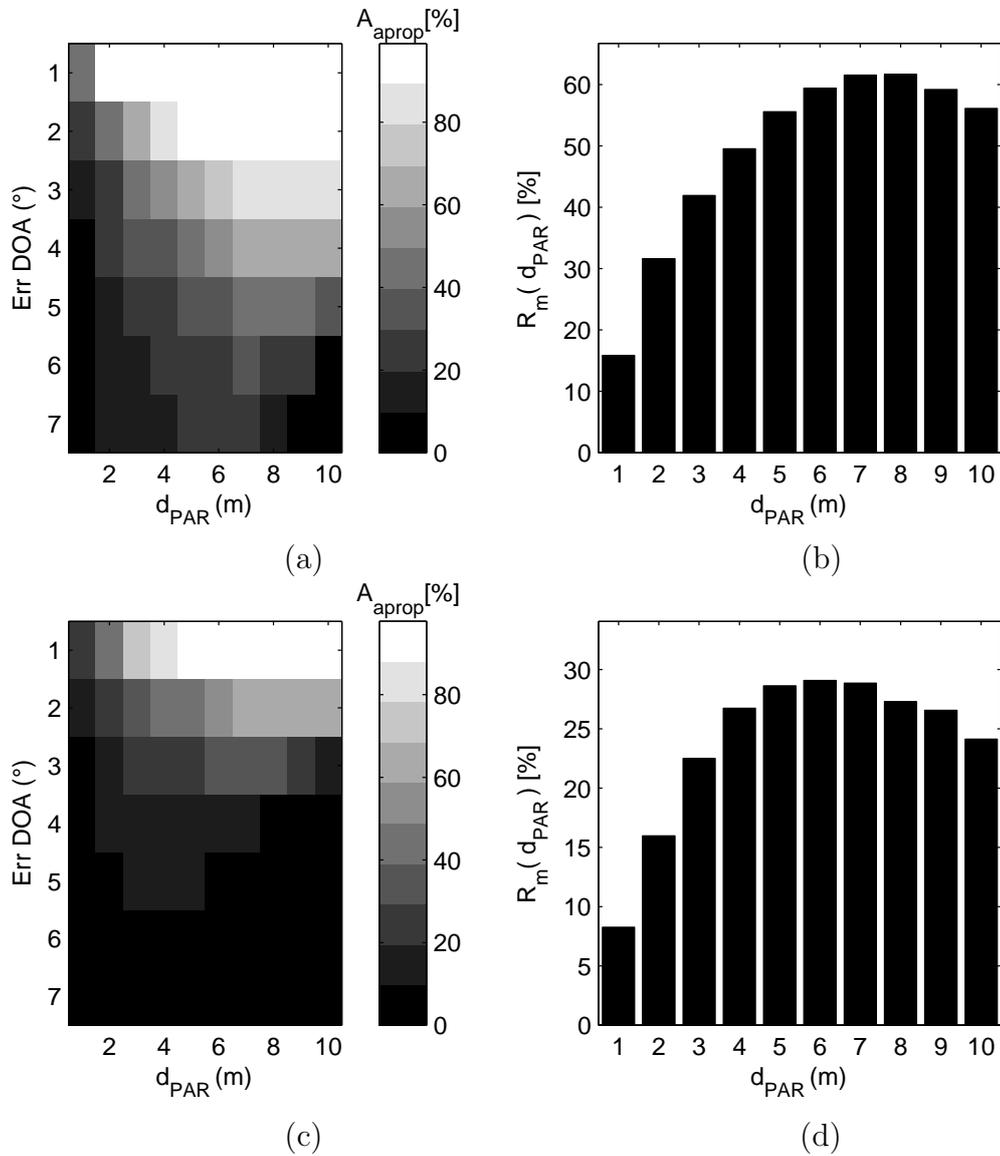


Figura 5.4: Área apropriada percentual, em escala de cinza, para diferentes valores de d_{PAR} e $\Delta\theta$, com (a) $D_{MAX} = 1$ m e (c) $D_{MAX} = 0.5$ m. Valores médios das áreas apropriadas, R_m , para diferentes valores de d_{PAR} , com (b) $D_{MAX} = 1$ m e (d) $D_{MAX} = 0.5$ m.

d_{PAR} igual a 7 m; para valores de $\Delta\theta$ superiores a 6° , percebe-se um aumento progressivo da região escura, que representa a queda de rendimento, sendo que para $\Delta\theta$ igual a 10° há predominância de tons fortemente escuros, correspondendo a regiões apropriadas inferiores a 10%, indicando a inviabilidade da utilização do sistema de localização.

Analisando o valores médios da região apropriada na Fig. 5.4(b), os melhores desempenhos do sistema ocorrem para d_{PAR} igual a 7 m e 8 m, com o valor médio da área percentual apropriada em torno de 60% da total.

A Fig. 5.4(c) mostra que para D_{MAX} igual a 0.5 m os tons claros ocorrem apenas para $\Delta\theta < 3^\circ$, indicando a necessidade do erro máximo nas estimativas das DOAs ser pequeno para o sistema de localização ser viabilizado.

Os resultados da Fig. 5.4(d) mostram que para D_{MAX} igual a 0.5 m a escolha de valores de d_{PAR} em torno de 6 m permitem que o sistema seja utilizado de forma satisfatória numa maior região da sala.

5.2 Estimação da Região Apropriada em Ambientes Simulados

Nesta seção, a análise de desempenho do sistema de localização apresentada na seção anterior será aplicada em ambientes simulados. Assim, será possível identificar a região apropriada para a localização e estabelecer a distância ideal entre os pares de microfones, dados a acurácia das DOAs e o erro máximo admitido nas estimativas das posições.

Serão considerados dois casos: em um ambiente simulado de baixa reverberação, com T_{60} igual a 200 ms, considerando os erros obtidos por algoritmos de estimação da DOA para localização de uma fonte; em um ambiente simulado de reverberação alta, com T_{60} igual a 800 ms, considerando os erros de estimação da DOA para localização de uma fonte e obtidos por algoritmos BSS para localização de duas fontes.

5.2.1 Ambiente Simulado

Nesta seção serão apresentadas as características do ambiente simulado, como posicionamento das fontes e microfones, dimensões da sala e outras propriedades necessárias para a análise da região apropriada.

O ambiente de simulação empregado com T_{60} igual a 200 ms usa as mesmas condições experimentais que [2], [17] e [73], sendo equivalente a uma sala projetada com paredes revestidas por materiais com propriedades que diminuem a intensidade da reverberação.

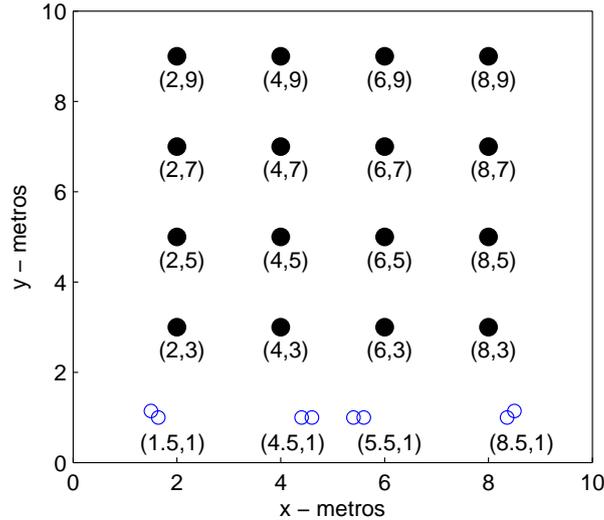


Figura 5.5: Posicionamento dos microfones e fontes no ambiente dos testes simulados.

O ambiente com T_{60} igual a 800 ms caracteriza um recinto com intensidade de reverberação elevada para algoritmos de BSS, sendo equivalente a uma sala sem tratamento acústico para diminuir a reverberação, sendo nessas condições os resultados dos métodos utilizados para estimar as DOAs bem menos acurados.

As posições das fontes e dos pares de microfones utilizadas nas simulações estão mostradas na Fig. 5.5, a qual representa uma sala com comprimento e largura iguais a 10 m. Cada fonte estará posicionada em um dos pontos marcados com o círculo preto, que têm coordenadas iguais a $x = 2, 4, 6$ e 8 , e $y = 3, 5, 7$, e 9 . Cada conjunto de microfones estará numa das posições indicadas por dois círculos, com $y = 1$.

A distância entre os microfones em cada par é igual a 20 cm e os pares nas posições (1.5,1) e (8.5,1) possuem as respectivas inclinações de -45° e 45° , em relação ao eixo horizontal, sendo estas inclinações as que forneceram menores erros de estimativas das DOAs. Os testes de localização foram realizados considerando dois casos: utilizando as estimativas das DOAs nos conjuntos de microfones das posições (1.5,1) e (8.5,1), denominado Caso A; e utilizando as estimativas das DOAs nos conjuntos de microfones das posições (4.5,1) e (5.5,1), denominado Caso B.

5.2.2 Região Adequada para uma Fonte em Ambiente com Baixa Reverberação

Nesta seção, avaliamos o desempenho do sistema de localização de uma fonte no ambiente simulado com baixa reverberação, sendo T_{60} igual a 200 ms. Os fundamentos teóricos apresentados na seção anterior utilizam o erro nas estimativas das DOAs, $\Delta\theta$, para calcular a região robusta de localização. Neste caso, o sistema de

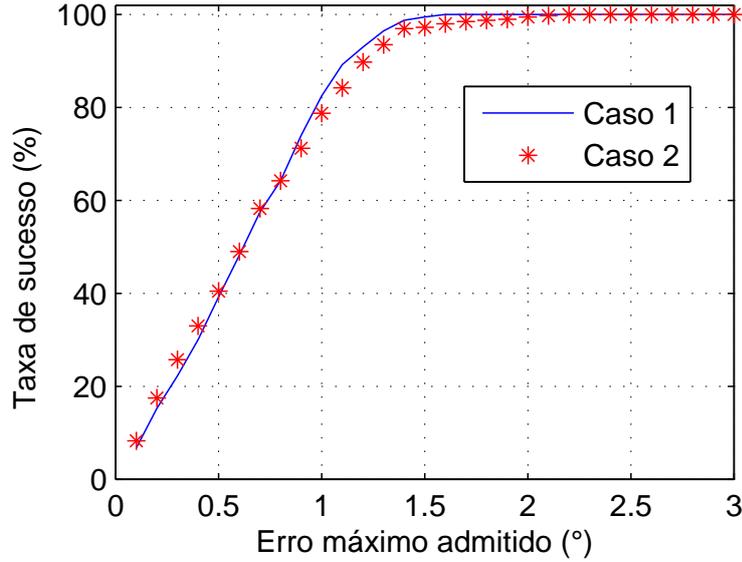


Figura 5.6: Taxa de acerto dado o erro máximo admitido.

localização está submetido a um ambiente de baixa reverberação, sem as complexidades introduzidas pelas misturas, permitindo estimativas das DOAs e posições com menores erros, conforme será apresentado na sequência do texto.

A Fig. 5.6 apresenta a taxa de sucesso obtida na estimação da DOA no algoritmo GCC-PHAT e da Eq. (4.5) para valores de erro máximo admitido entre 0.1° e 3° . No Caso 1 o par de microfones utilizado para obter a DOA encontra-se na posição (1.5,1), enquanto que no Caso 2 o par está na posição (4.5,1). O caso com o par na posição (8.5,1) é geometricamente simétrico ao Caso 1, apresentando resultados semelhantes; a mesma observação se aplica ao par na posição (5.5,1), sendo esta simétrica em relação à posição (4.5,1). Da Fig. 5.6, observa-se que nos dois casos considerados a quase totalidade das estimativas das DOAs apresenta erros menores que $\Delta\theta = 1.4^\circ$, sendo estes distribuídos quase uniformemente entre 0.1° e 1.4° . Não se observam diferenças significativas nos valores máximos das DOAs para as diferentes posições dos pares. Os resultados mostrados na Fig. 5.6 nos permitem determinar a faixa de valores de $\Delta\theta$ a ser utilizada na determinação da região apropriada para localização.

A Fig. 5.7 mostra os valores médios percentuais da área apropriada em relação à área total, para $\Delta\theta$ entre 0.1° e 1.4° , considerando diferentes valores de d_{PAR} e D_{MAX} . Esses resultados auxiliam na escolha da distância dos microfones d_{PAR} , dado o erro de localização tolerável D_{MAX} . Variou-se d_{PAR} entre 1 m e 10 m, com passo de 1 m, e D_{MAX} entre 0.2 m e 0.5 m, com passo de 0.1 m. Dessa figura, conclui-se que, para $D_{MAX} = 0.2$ m, o melhor valor de d_{PAR} é 7 m, resultando no valor médio da área apropriada igual a 61%. Já para D_{MAX} entre 0.3 m e 0.5 m, os melhores valores de d_{PAR} são 8 m e 9 m, não havendo praticamente diferença entre as áreas

médias apropriadas para esses dois valores. As áreas médias apropriadas percentuais (em relação a área total) com $d_{PAR} = 7$ m são aproximadamente 81%, 91% e 97% para $D_{MAX} = 0.3$ m, 0.4 m e 0.5 m, respectivamente.

Os comportamentos das variáveis envolvidas - região apropriada, d_{PAR} , $\Delta\theta$ e D_{MAX} - são analisados através das Figs. 5.6 a 5.9. A Fig. 5.6 indica a acurácia da DOA disponível para o ambiente simulado com T_{60} igual a 200 ms, a Fig. 5.7 auxilia na escolha da distância dos microfones, a Fig. 5.8 auxilia na escolha do D_{MAX} e a Fig. 5.9 mostra a região considerada apropriada para a localização.

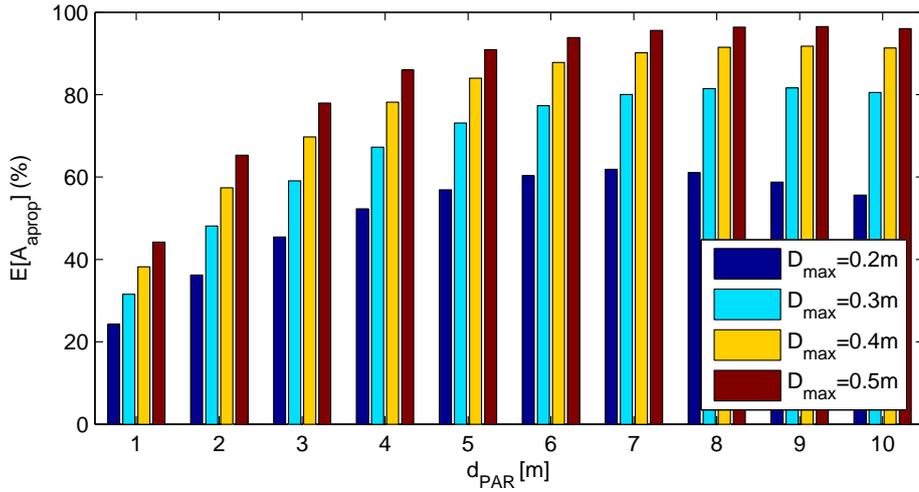


Figura 5.7: Valor médio da área apropriada ao variar $\Delta\theta$ entre 0.1° e 1.5° , para d_{PAR} entre 1 m e 10 m e D_{MAX} entre 0.2 m e 0.5 m.

A Fig. 5.8 mostra os valores médios percentuais da área apropriada para D_{MAX} entre 0.1 m e 0.65 m, sendo a média obtida com $\Delta\theta$ entre 0.1° e 1.4° e d_{PAR} entre 5 m e 9 m. Enquanto para D_{MAX} igual a 0.2 m a área apropriada é de 63.5%, para D_{MAX} igual a 0.5 m esse valor aumenta para 96%.

A partir dos resultados da Fig. 5.8, escolhemos como valor de referência $D_{MAX} = 0.25$ m para o erro máximo admitido na localização de uma fonte no ambiente simulado com $T_{60} = 200$ ms, uma vez que a área considerada apropriada para a localização com esse erro é relativamente grande (72% da área total) e que esse valor de erro de localização pode ser considerado pequeno quando comparado às dimensões da sala, e, portanto, tolerável em grande parte das aplicações.

A Fig. 5.9 mostra a área apropriada, em escala de cinza, para $D_{MAX} \leq 0.25$ m e erro máximo das estimativas das DOAs $\Delta\theta$ igual a 0.5° , 1° e 1.5° , sendo a distância entre pares $d_{PAR} = 7$ m. Essa configuração dos arranjos de microfones foi denominada Caso A.

Analisando as regiões apropriadas da Fig. 5.9 e considerando as taxas de sucesso da Fig. 5.6, tiramos as seguintes conclusões para o ambiente simulado com baixa reverberação: 40% das estimativas das DOAs apresentam $\Delta\theta < 0.5^\circ$, sendo a área

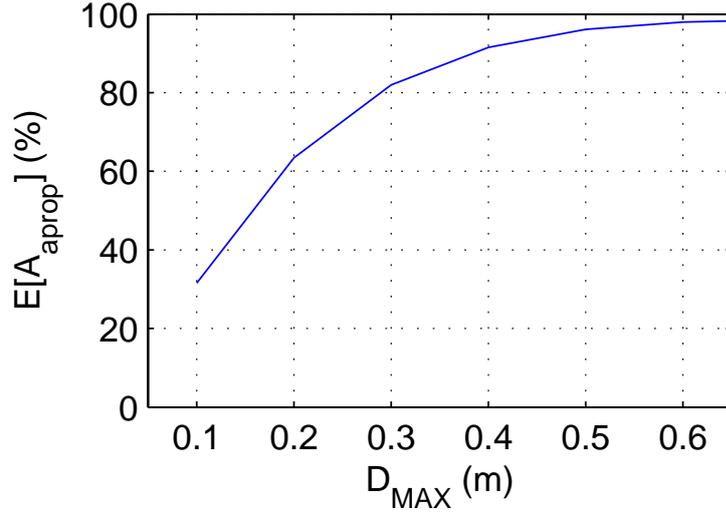


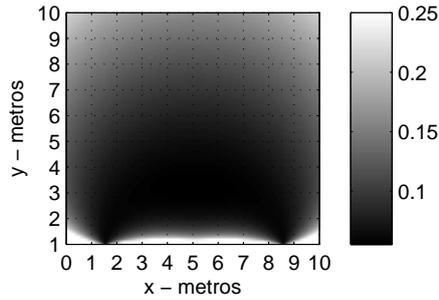
Figura 5.8: Valores médios da área apropriada para diferentes erros de localização D_{MAX} .

apropriada para esse erro máximo das DOAs igual a 98% da área pesquisada; 80% das estimativas das DOAs apresentam $\Delta\theta < 1^\circ$, sendo a área apropriada igual a 61% da área pesquisada, estando excluídas dessa região as posições das fontes com $y = 9$ m; 100% das estimativas das DOAs apresentam $\Delta\theta < 1.5^\circ$, sendo a área apropriada para esse erro máximo das DOAs igual a 32% da área pesquisada, estando excluídas as posições com $y = 7$ m e 9 m.

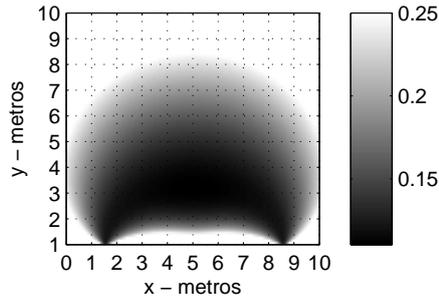
Assim, para d_{PAR} igual a 7 m, pode-se esperar desempenho robusto do sistema nas posições com $y = 3$ m e 5 m, sendo tolerados erros de até 1.5° nas estimativas das DOAs. Para bom desempenho na localização de fontes nas posições com $y = 7$ m, os erros nas estimativas das DOAs devem ser menores que 1° , o que ocorre em cerca de 80% das estimativas. Já nas posições com $y = 9$ m, são necessários erros menores que 0.7° , o que ocorre em aproximadamente 60% das estimativas.

A Fig. 5.10 mostra a área apropriada, em escala de cinza, para erro máximo tolerável $D_{MAX} \leq 0.25$ m e erro máximo das estimativas das DOAs $\Delta\theta$ igual a 0.5° , 1° e 1.5° , sendo a distância entre arrays $d_{PAR} = 1$ m. Essa configuração dos arranjos de microfones foi denominada Caso B.

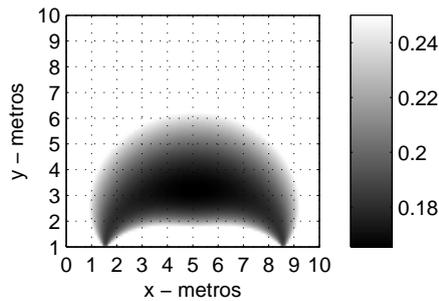
Comparando os resultados mostrados nas Fig. 5.9 e 5.10, verifica-se uma grande redução das áreas apropriadas causada pela diminuição da d_{PAR} . O Caso B, com $d_{PAR} = 1$ m, exige uma melhor acurácia nas estimativas das DOAs, uma vez que com $\Delta\theta = 1^\circ$ a área apropriada representa apenas 12% da região pesquisada, enquanto para o Caso A, com $d_{PAR} = 7$ m, a área apropriada era de 61%. Portanto, espera-se uma baixa performance do sistema de localização quando $d_{PAR} = 1$ m.



(a)



(b)



(c)

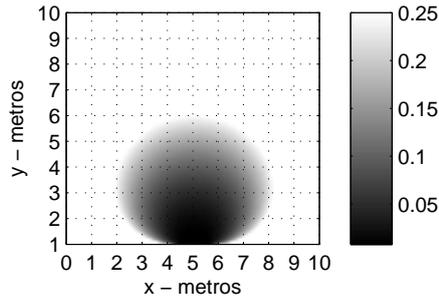
Figura 5.9: Região apropriada para $d_{PAR} = 7$ m (Caso A) com $D_{MAX} \leq 0.25$ m, e $\Delta\theta$ igual a (a) 0.5° , (b) 1° e (c) 1.5° .

5.2.3 Região Apropriada em Ambiente com Alta Reverberação

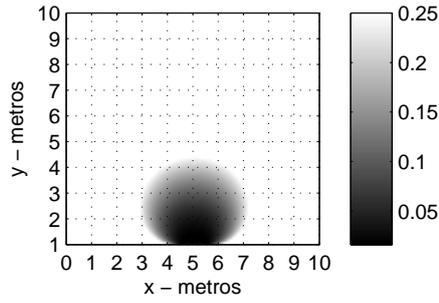
Nesta seção, avaliamos o desempenho do sistema de localização no ambiente simulado com alta reverberação, sendo T_{60} igual a 800 ms. As estimativas das DOAs, obtidas pelo algoritmo GCC-PHAT para uma fonte e pelos algoritmos BSS para duas fontes, apresentaram erros máximos entre³ 0° e 7° .

A Fig. 5.11 mostra os valores médios percentuais da área apropriada em relação à área total, para $\Delta\theta$ entre 1° e 7° , considerando diferentes valores de d_{PAR} e D_{MAX} . Esses resultados auxiliam na escolha da distância dos microfones d_{PAR} , dado o erro de localização tolerável D_{MAX} . Variou-se d_{PAR} entre 1 m e 10 m, com passo de 1 m, e D_{MAX} entre 0.6 m e 1.4 m, com passo de 0.2 m. Dessa figura, conclui-se que para

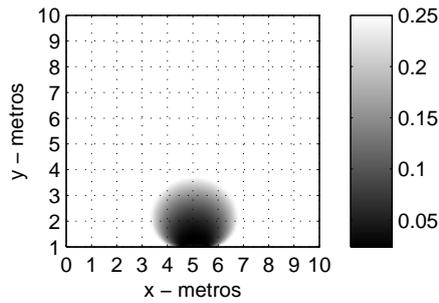
³Esses resultados serão apresentados no Cap. 7.



(a)



(b)



(c)

Figura 5.10: Região apropriada para $d_{PAR} = 1$ m (Caso B) com $D_{MAX} \leq 0.25$ m, e $\Delta\theta$ igual a (a) 0.5° , (b) 1° e (c) 1.5° .

$D_{MAX} = 0.6$ m, o melhor valor de d_{PAR} é 5 m, resultando no valor médio da área apropriada igual a 44%. Já para $D_{MAX} = 0.8$ m, os melhores valores de d_{PAR} são 6 m e 7 m, não havendo praticamente diferença entre as áreas médias apropriadas para esses dois valores, sendo essa igual a 56%. Para $D_{MAX} = 1$ m, os melhores valores de d_{PAR} são 7 m e 8 m, com área apropriada média igual a 56%. Para $D_{MAX} = 1.2$ e 1.4 m, as maiores áreas foram obtidas com d_{PAR} igual a 8 m e 9 m, sendo iguais a 75% e 82%, respectivamente.

A Fig. 5.12 mostra os valores médios percentuais da área apropriada para D_{MAX} entre 0.4 m e 1.6 m, sendo a média calculada variando-se $\Delta\theta$ entre 1° e 7° e d_{PAR} entre 6 m e 9 m. Enquanto para D_{MAX} igual a 0.4 m a área apropriada é de 34% da área total, para D_{MAX} igual a 1.4 m esse valor aumenta para 84%.

A partir dos resultados da Fig. 5.12, escolhemos como valor de referência $D_{MAX} = 1.2$ m para o erro máximo admitido na localização de uma ou duas fontes

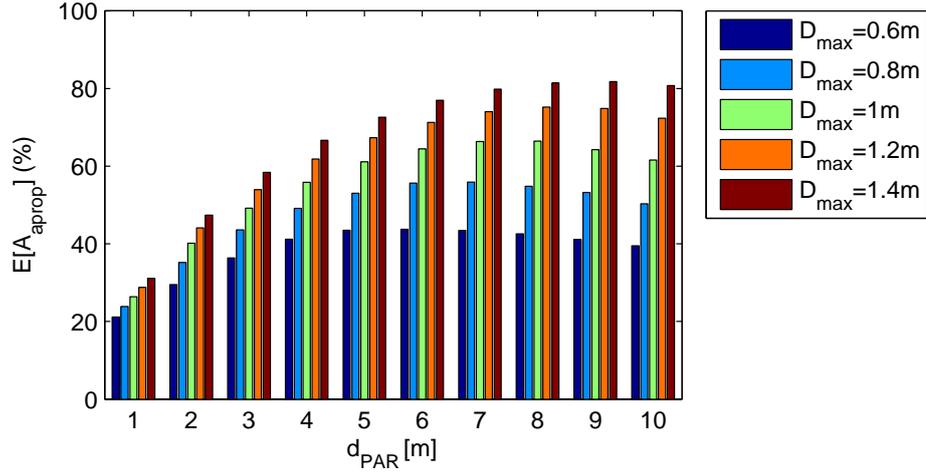


Figura 5.11: Valor médio da área apropriada ao variar $\Delta\theta$ entre 1° e 7° , para d_{PAR} entre 1 m e 10 m e D_{MAX} entre 0.6 m e 1.4 m.

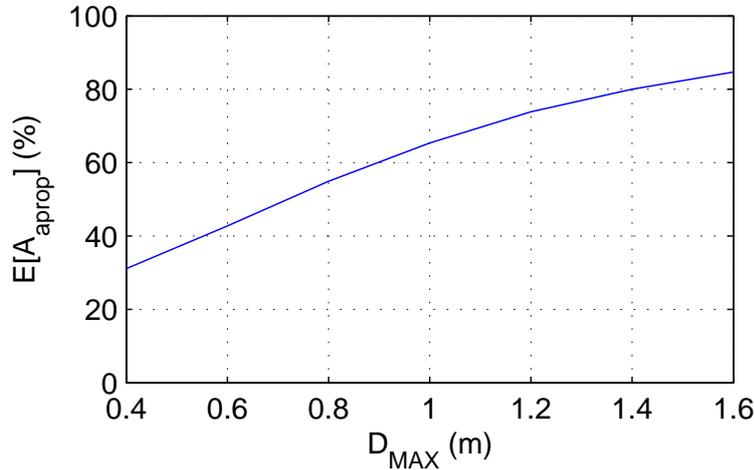
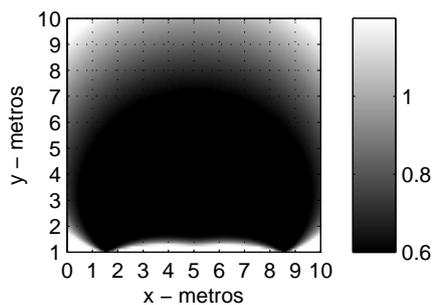


Figura 5.12: Valores médios da área apropriada para diferentes erros de localização D_{MAX} .

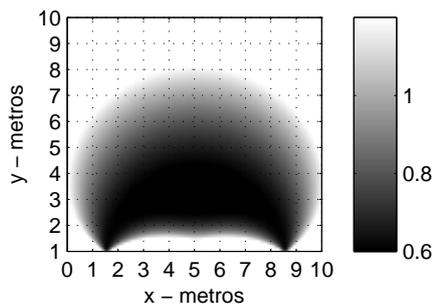
no ambiente simulado com $T_{60} = 800$ m, uma vez que a área considerada apropriada para a localização com esse erro é relativamente grande (75% da área total) e que esse valor de erro de localização pode ser tolerável quando comparado às dimensões da sala para grande parte das aplicações, principalmente considerando que apenas dois pares de microfones estão sendo usados.

A Fig. 5.13 mostra a área apropriada, em escala de cinza para $D_{MAX} \leq 1.2$ m e erros máximos das estimativas das DOAs $\Delta\theta$ iguais a 3° , 5° e 7° , sendo a distância entre arrays $d_{PAR} = 7$ m. Essa configuração dos arranjos de microfones foi denominada Caso A.

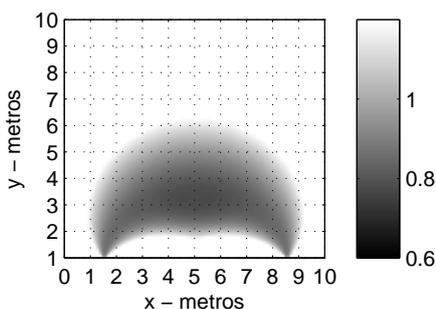
Analisando as regiões apropriadas da Fig. 5.13 e as posições das fontes apresentadas na Fig. 5.5, tiramos as seguintes conclusões para o ambiente simulado com alta reverberação: a área apropriada para $\Delta\theta = 3^\circ$ corresponde a 93% da área da



(a)



(b)



(c)

Figura 5.13: Região apropriada para $d_{PAR} = 7$ m (Caso A) com $D_{MAX} \leq 1.2$ m, e $\Delta\theta$ igual a (a) 3° , (b) 5° e (c) 7° .

região pesquisada; para $\Delta\theta = 5^\circ$, esse percentual cai para 56%, sendo excluídas dessa região as posições das fontes com $y = 9$ m; para $\Delta\theta = 7^\circ$, a área apropriada percentual é 32%, sendo excluídas as posições com $y = 7$ m e 9 m.

Assim, para d_{PAR} igual a 7 m, pode-se esperar desempenho robusto do sistema nas posições com $y = 3$ m e 5 m, sendo tolerados erros de até 7° nas estimativas das DOAs. Para bom desempenho na localização de fontes nas posições com $y = 7$ m, os erros nas estimativas das DOAs devem ser menores que 5° . Já nas posições com $y = 9$ m, são necessários erros menores que 3° .

A Fig. 5.14 mostra a área apropriada, em escala de cinza, para erro máximo tolerável $D_{MAX} \leq 1.2$ m e erro máximo das estimativas das DOAs $\Delta\theta$ igual a 3° , 5° e 7° , sendo a distância entre arrays $d_{PAR} = 1$ m. Essa configuração dos arranjos de microfones foi denominada Caso B.

Comparando os resultados mostrados nas Fig. 5.13 e 5.14, verifica-se uma grande

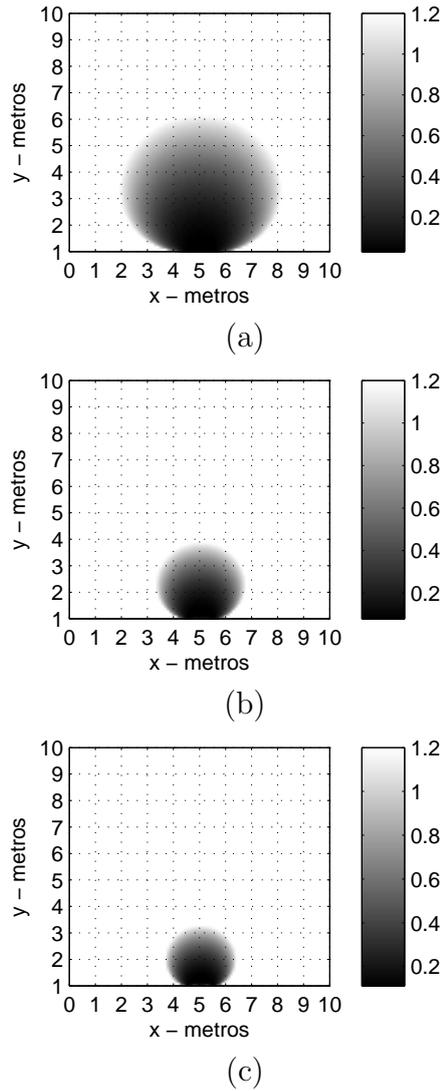


Figura 5.14: Região apropriada para $d_{PAR} = 1$ m (Caso B) com $D_{MAX} \leq 1.2$ m, e $\Delta\theta$ igual a (a) 3° , (b) 5° e (c) 7° .

redução das áreas apropriadas causada pela diminuição da d_{PAR} . O Caso B, com $d_{PAR} = 1$ m, exige uma melhor acurácia nas estimativas das DOAs, uma vez que com $\Delta\theta = 5^\circ$ a área apropriada representa apenas 8% da região pesquisada, enquanto para o Caso A, com $d_{PAR} = 7$ m, a área apropriada era de 56%. Portanto, espera-se um desempenho ruim do sistema de localização quando $d_{PAR} = 1$ m.

5.3 Estimação da Região Apropriada em Ambiente Real

Nesta seção, a análise de desempenho do sistema de localização apresentada na Seção 5.1 será aplicada em um ambiente real. Assim, será possível identificar a região apropriada para a localização e estabelecer a distância ideal entre os pares de

microfones, dados a acurácia das estimativas das DOAs obtidas com sinais gravados e o erro máximo admitido nas estimativas das posições.

5.3.1 Descrição do Ambiente Real

Os testes em ambiente real foram realizados em uma sala de aula⁴, mostrada na Fig. 5.15. Suas dimensões são 8 m de largura e 14.9 m de comprimento, e o tempo de reverberação medido é igual a $T_{60} = 900$ ms, caracterizando um ambiente desafiador para os algoritmos de BSS e para os sistemas de localização.



Figura 5.15: Fotografias da sala de aula utilizada nos testes.

As posições das fontes e dos pares de microfones utilizados nos testes são mostradas na Fig. 5.16. Cada fonte será posicionada em um dos pontos marcados com o círculo preto, e cada par de microfones estará numa das posições indicadas por dois círculos⁵.

Foram emitidos sucessivamente, através de caixas de som, 10 sinais de voz distintos, com duração de 10 segundos cada, sendo os sinais dos oito microfones gravados simultaneamente⁶. As misturas foram compostas somando-se os sinais gravados emitidos de 2 posições diferentes.

Os testes de localização foram realizados em dois casos: utilizando os pares

⁴Sala D105 do Centro de Tecnologia da UFRJ.

⁵Os microfones utilizados foram dos modelos DONNER DR 3100 (unidirecional) e Behringer ECM8000 (omnidirecional).

⁶Os sinais foram aquiritados com taxa de amostragem de 44.1 kHz, sendo em seguida subamostrados para 16 KHz de modo a reduzir a complexidade computacional do processamento.

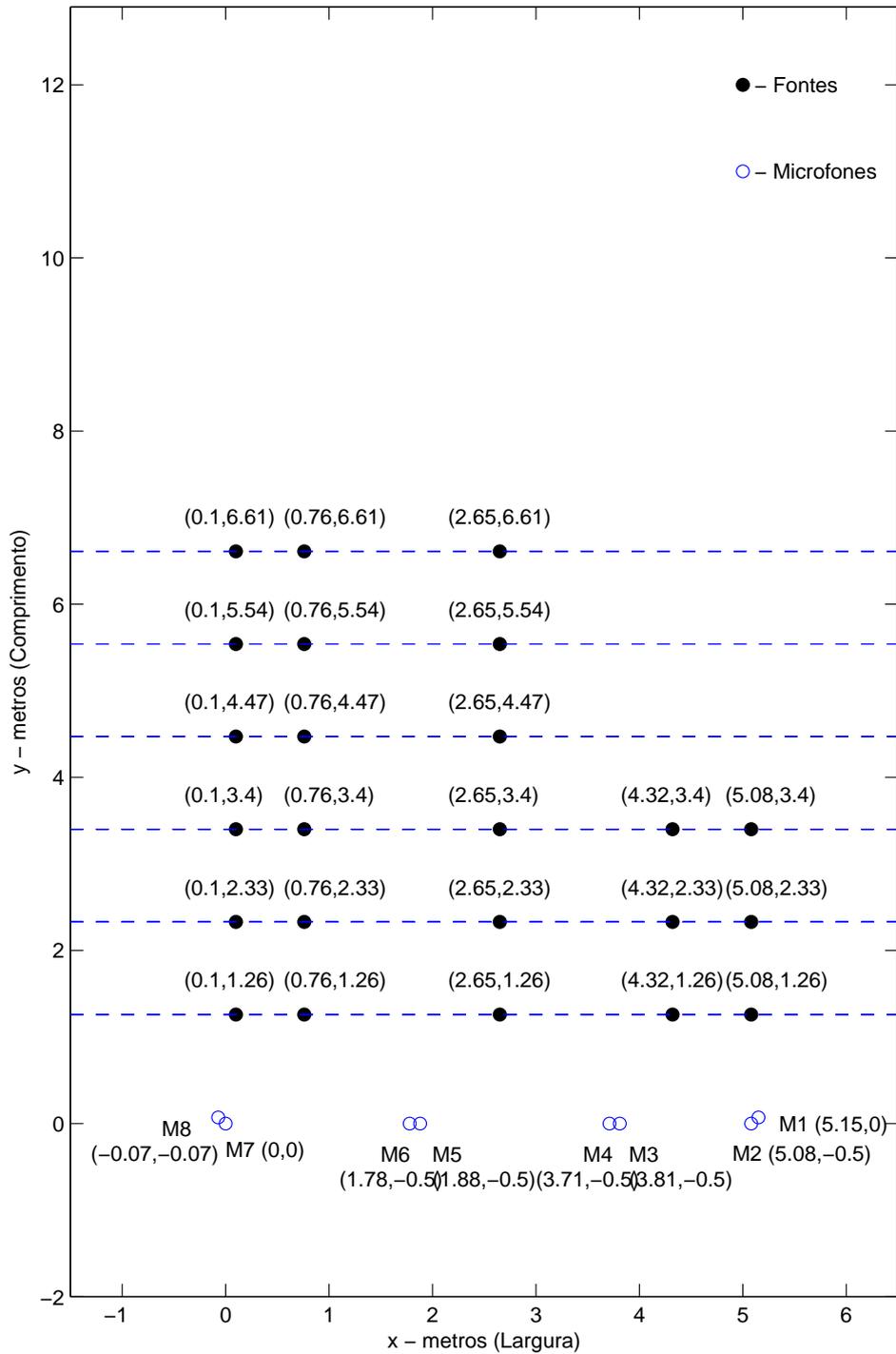


Figura 5.16: Posicionamento dos microfones e fontes nos testes em ambiente real.

(M1,M2) e (M7,M8), denominado Caso A; e utilizando os pares (M3,M4) e (M5,M6),

denominado Caso B. A distância entre os microfones de cada par foi igual a 10 cm.

5.3.2 Região Adequada para o Ambiente Real

As estimativas das DOAs, obtidas pelo algoritmo GCC-PHAT para uma fonte e pelos algoritmos BSS para duas fontes, apresentaram erros máximos entre⁷ 1° e 9° .

A Fig. 5.17 mostra os valores médios percentuais da área adequada em relação à área total, calculados variando-se $\Delta\theta$ entre 1° e 9° , considerando diferentes valores de d_{PAR} e D_{MAX} . Variou-se d_{PAR} entre 1 m e 7 m, com passo de 1 m, e D_{MAX} entre 0.6 m e 1.4 m, com passo de 0.2 m. Dessa figura, conclui-se que os melhores valores para d_{PAR} estão entre 5 m e 6 m. Para $D_{MAX} = 0.6$ m e 0.8 m, o melhor valor de d_{PAR} é 5 m, resultando em valores médios da área adequada iguais a 48% e 61%, respectivamente. Para $D_{MAX} = 1$ m, 1.2 m e 1.4 m, o melhor valor de d_{PAR} é 6 m, com áreas apropriadas médias iguais a 67%, 79% e 85%, respectivamente.

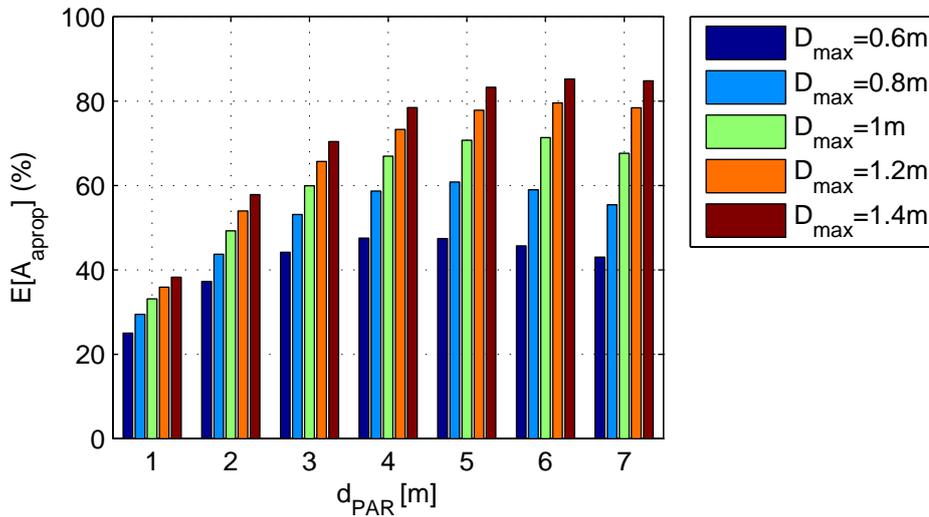


Figura 5.17: Valor médio da área adequada ao variar $\Delta\theta$ entre 1° e 9° .

A Fig. 5.18 mostra os valores médios percentuais da área adequada para D_{MAX} entre 0.4 m e 1.6 m, sendo a média obtida com $\Delta\theta$ entre 1° e 9° e d_{PAR} entre 4 m e 6 m. Enquanto para D_{MAX} igual a 0.6 m a área adequada é de 52% da área total, para D_{MAX} igual a 1.2 m esse valor aumenta para 82%.

A partir dos resultados da Fig. 5.18, escolhemos como valor de referência $D_{MAX} = 1.2$ m para o erro máximo admitido na localização, uma vez que a área considerada adequada para a localização com esse erro foi de 84% da área total e que esse valor de erro de localização pode ser tolerável quando comparado às dimensões da sala, considerando a aplicação de ensino a distância.

⁷Esses resultados serão apresentados nos Caps. 6 e 7.

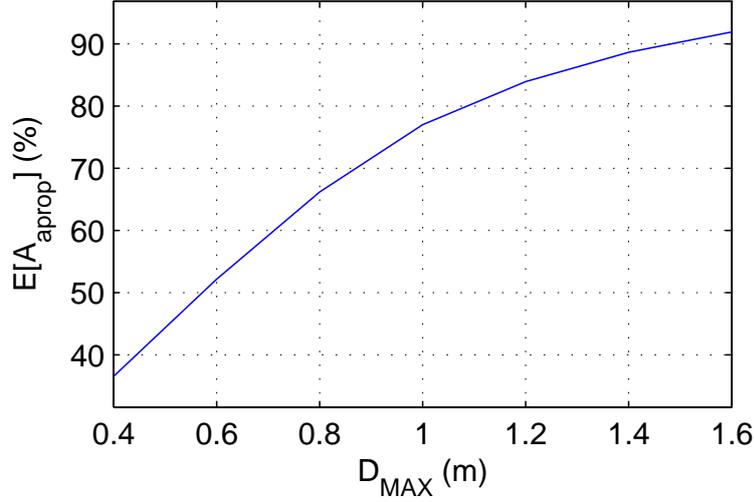


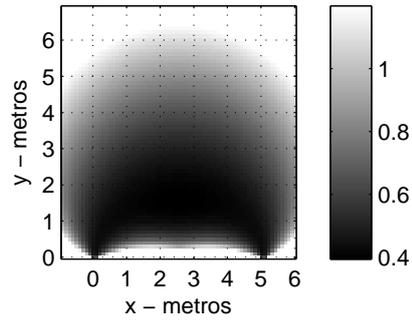
Figura 5.18: Valores médios da área apropriada para diferentes erros de localização D_{MAX} .

A Fig. 5.19 mostra a área apropriada, em escala de cinza, para $D_{MAX} \leq 1.2$ m e erros máximos das estimativas das DOAs $\Delta\theta$ iguais a 5° , 7° , 9° e 11° , para o Caso A, em que a distância entre arrays $d_{PAR} = 5$ m.

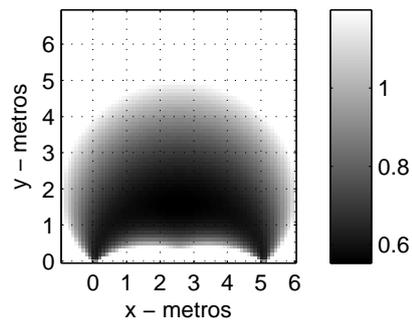
Analisando as regiões apropriadas da Fig. 5.19 e as posições das fontes apresentadas na Fig. 5.16, tiramos as seguintes conclusões para o ambiente real: a área apropriada para $\Delta\theta = 5^\circ$ correspondente a 81% da área da região pesquisada, sendo excluídas dessa região as posições das fontes com $y = 6.61$ m; para $\Delta\theta = 7^\circ$, esse percentual cai para 56%, sendo excluídas dessa região as posições das fontes com $y \geq 5.54$ m; para $\Delta\theta = 9^\circ$, a área apropriada percentual é de 37%, sendo excluídas as posições com $y \geq 3.8$ m; para $\Delta\theta = 11^\circ$, a área apropriada percentual é de 24%, estando apenas as posições mais próximas dos microfones, com $y = 1.26$ m, na região apropriada. Assim, o bom desempenho do sistema de localização depende da acurácia das estimativas das DOAs, sendo que os erros de localização aumentam à medida que a distância das fontes aos microfones aumenta.

A Fig. 5.20 mostra a área apropriada, em escala de cinza, para erro máximo tolerável $D_{MAX} \leq 1.2$ m e erro máximo das estimativas das DOAs $\Delta\theta$ igual a 2° , 5° e 7° , considerando o Caso B, no qual a distância entre arrays é $d_{PAR} = 2$ m.

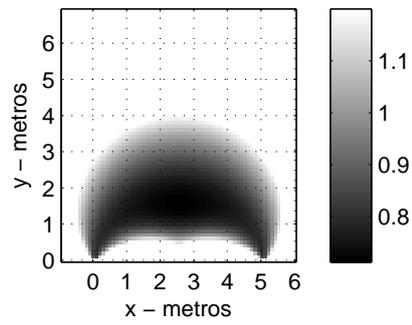
Os resultados da Fig. 5.20, quando comparados aos da Fig. 5.19, mostram a significativa redução da área apropriada causada pela diminuição de d_{PAR} . O Caso B, com $d_{PAR} = 2$ m, exige uma melhor acurácia nas estimativas das DOAs, uma vez que com $\Delta\theta = 5^\circ$ a área apropriada representa apenas 37% da região pesquisada, enquanto para o Caso A, com $d_{PAR} = 5$ m, a área apropriada era de 81%. Portanto, espera-se um desempenho ruim do sistema de localização quando $d_{PAR} = 2$ m. Os resultados de localização com os dados experimentais serão apresentados no Cap. 7.



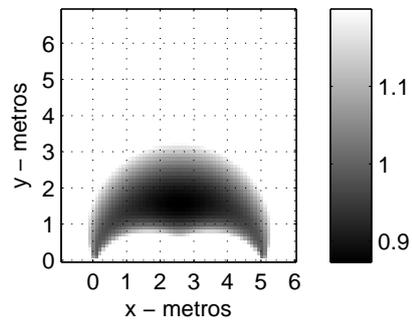
(a)



(b)

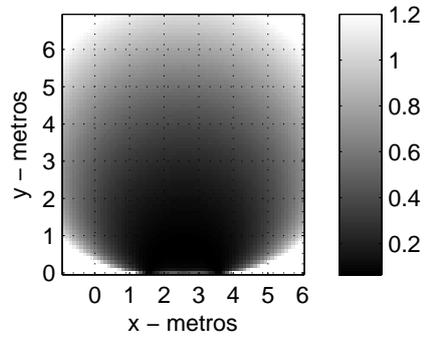


(c)

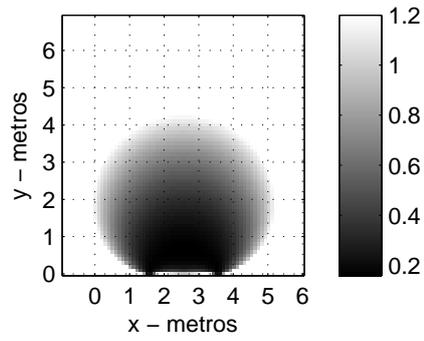


(d)

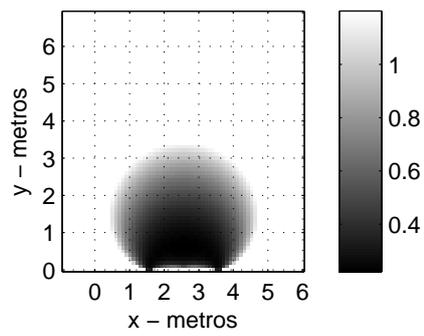
Figura 5.19: Área considerada apropriada para a localização com $D_{MAX} = 1.2$ m, $d_{PAR} = 5$ m (Caso A) e $\Delta\theta$ igual a (a) 5° , (b) 7° , (c) 9° e (d) 11° .



(a)



(b)



(c)

Figura 5.20: Região apropriada para $d_{PAR} = 2$ m (Caso B) com $D_{MAX} \leq 1.2$ m, e $\Delta\theta$ igual a (a) 2° , (b) 5° e (c) 7° .

Capítulo 6

Resultados da Localização para Uma Fonte Ativa

Neste capítulo, avaliamos os resultados obtidos com o sistema de localização quando apenas uma fonte encontra-se ativa no intervalo de tempo analisado. Consideramos os ambientes simulados e real apresentados no Cap. 5. As estimativas das DOAs são obtidas através do algoritmo GCC-PHAT descrito no Cap. 4. O objetivo principal destes experimentos é avaliar a eficiência do procedimento de estimação da região apropriada e da escolha de d_{PAR} apresentado no Cap. 5 para uma única fonte ativa, sem a presença dos erros nas estimativas das DOAs introduzidas pelos processos de misturas e de BSS. O cenário mais complexo de múltiplas fontes será avaliado no Cap. 7.

O efeito da reverberação nos sinais empregados em ambientes simulados foi obtido através do método das imagens descrito na Seção 2.3; as características dos ambientes simulados e real foram apresentadas nas Seções 5.2.1 e 5.3.1; o algoritmo GCC-PHAT foi descrito na Seção 4.1.

6.1 Localização de uma Fonte em Ambiente Simulado com Baixa Reverberação

Nesta seção, avaliaremos os resultados experimentais no ambiente simulado com T_{60} igual a 200 ms. Conforme abordado na Seção 5.2, o procedimento de otimização do sistema de localização utiliza a acurácia do erro nas estimativas das DOAs para encontrar a melhor distância entre os dois conjuntos de microfones e a região apropriada de funcionamento do sistema. Portanto, inicialmente investigaremos o erro nas estimativas das DOAs para todas as possíveis posições da fonte e do conjunto de microfones¹. Em seguida, avaliaremos os erros das estimativas das posições das

¹Essas posições estão mostradas na Fig. 5.5.

fontes considerando os arranjos de microfones nas posições (1.5,1) e (8.5,1) e nas posições (4.5,1) e (5.5,1).

A Fig. 6.1 apresenta a taxa de sucesso nas estimativas das DOAs com erro máximo admitido $\Delta\theta = 1^\circ$ para cada posição da fonte. Esse valor de $\Delta\theta$ foi escolhido por resultar em erros de localização $D_{MAX} < 0.25$ m, de acordo com a análise da Seção 5.2.2. A Fig. 6.1(a) mostra os percentuais de estimativas com erros menores que 1° obtidos com o par de microfones na posição (1.5,1), denominado de Caso 1, enquanto que a Fig. 6.1(b) mostra os percentuais correspondentes com o par de microfones na posição (4.5,1), denominado de Caso 2. Devido à simetria da posição (8.5,1) em relação à (1.5,1), consideraremos que as estimativas das DOAs para o par de microfones na posição (8.5,1) apresentam erros percentuais espelhados em relação ao Caso 1. A mesma consideração pode ser feita em relação às DOAs do par de microfones na posição (5.5,1), que é simétrica à posição (4.5,1) e, portanto, com erros espelhados em relação ao Caso 2.

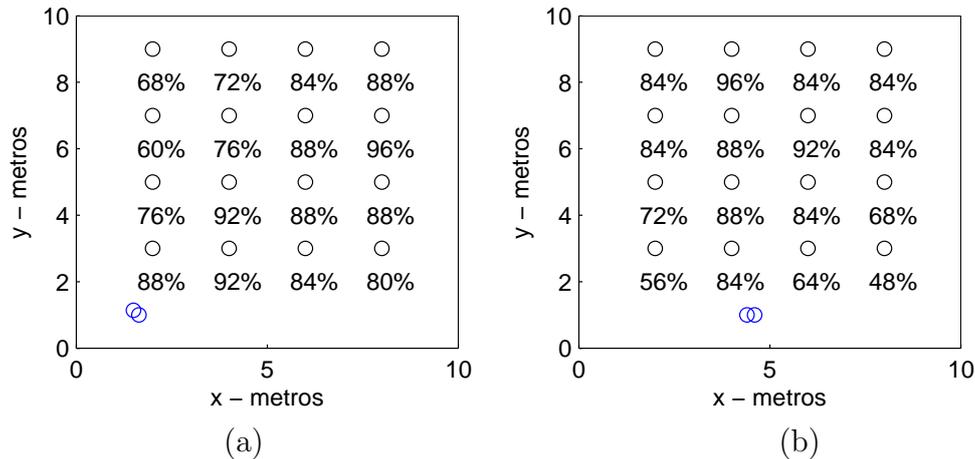


Figura 6.1: Taxa de sucesso nas estimativas das DOAs em cada possível posição da fonte considerando $\Delta\theta = 1^\circ$ para (a) Caso 1 e (b) Caso 2.

Analisando as taxas de sucesso das estimativas das DOAs para as posições mostradas na Fig. 6.1, verifica-se que na maioria dos casos elas são maiores que 60%, sendo os piores resultados encontrados nas posições (2,3) e (8,3) do Caso 2, cujas respectivas taxas de sucesso são iguais a 56 % e 48 %. Os erros maiores nesses casos estão associados ao fato de os ângulos de suas DOAs reais estarem próximos de 0° e 180° , sendo as estimativas dessas direções mais propensas a erros. Esses grandes erros foram evitados no Caso 1, inclinando-se o conjunto de microfones de 45° , de forma a evitar direções de chegada com ângulos próximos de 0° e 180° .

A Fig. 6.2 mostra a variação do valor médio da taxa de sucesso para erro máximo nas estimativas das DOAs entre 0.1° e 3° . Nessa figura, observa-se que não há diferenças significativas entre os Casos 1 e 2, e que para o erro máximo nas

estimativas das DOAs igual a 1° a taxa de sucesso é 80%, e para 1.4° é praticamente 100 %.

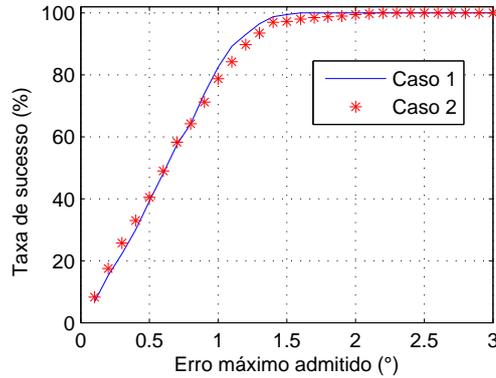


Figura 6.2: Taxa de sucesso média ao variar o erro máximo das estimativas das DOAs $\Delta\theta$ admitido.

A Fig. 6.3 apresenta a taxa de sucesso da localização de uma fonte com erro máximo admitido $D_{MAX} = 0.25$ m para cada possível posição da fonte. A Fig. 6.3(a) mostra os percentuais de estimativas das posições com erros menores que 25 cm obtidas com os dois conjuntos de microfones nas posições (1.5,1) e (8.5,1), denominado de Caso A, enquanto que a Fig. 6.3(b) mostra os percentuais correspondentes com os conjuntos de microfones nas posições (4.5,1) e (5.5,1), denominado de Caso B.

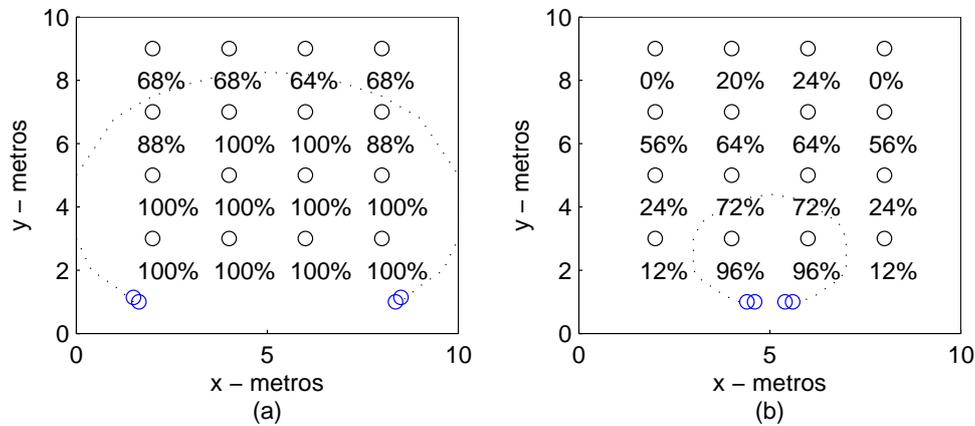


Figura 6.3: Taxa de sucesso da localização em cada possível posição da fonte considerando erro máximo admitido $D_{MAX} = 25$ cm para (a) Caso A e (b) Caso (B).

Analisando os resultados das posições mais próximas dos microfones, com $y = 3$ m, observa-se que as posições (4,3) e (6,3) estão dentro das regiões apropriadas das Figs. 5.9 e 5.10 (mostradas em linhas tracejadas na Fig. 6.3) e apresentam taxas de sucesso altas, iguais a 100% no Caso A e 96% no Caso B. As posições (2,3) e (8,3) estão dentro das regiões apropriadas da Fig. 5.9 e fora das regiões apropriadas da Fig. 5.10, justificando as respectivas taxas de sucesso iguais a 100% e 12% para os Casos A e B.

As posições com $y = 5$ m estão dentro das regiões apropriadas do Caso A na Fig. 5.9 e apresentam taxas de sucesso igual a 100%. No Caso B as posições (4,5) e (6,5) estão na região apropriada apenas da Fig. 5.10(a), correspondente a estimativas das DOAs com menores erros, e apresentam taxas de sucesso igual a 72%. Já as posições (2,5) e (8,5) estão fora das regiões apropriadas do Caso B, apresentando taxas de sucesso de apenas 24%.

Os resultados para as posições (4,7) e (6,7) mostram taxas de sucesso iguais a 100% para o Caso A e 64% para o Caso B. Uma vez que no Caso A essas posições estão dentro das regiões apropriadas, observa-se que os resultados experimentais estão de acordo com os teóricos. No Caso B essas posições estão fora das regiões apropriadas da Fig. 5.10 e suas taxas de sucesso moderadas indicam que o modelo proposto é conservador.

Os resultados para as posições (2,7) e (8,7) mostram o Caso A com melhor performance, com taxas de sucesso igual a 88%, enquanto que as taxas correspondentes para o Caso B foram 56%. Essas posições estão dentro das regiões apropriadas do Caso A e fora daquelas do Caso B.

Os resultados para as posições mais afastadas dos microfones, com $y = 9$ m, mostram o Caso A com taxa de sucesso média igual a 66% e o Caso B com 12%. Os resultados ruins para o Caso B estão de acordo com a análise da região apropriada, que considerou essas posições como inapropriadas.

A Fig. 6.4 mostra o comportamento do valor médio da taxa de sucesso ao variar o erro máximo admitido nas estimativas das posições, considerando D_{MAX} entre 0.1 e 3 m, para os Casos A e B. O valor desse parâmetro depende da exigência da aplicação, sendo limitado pela dificuldade em obter a acurácia necessária nas estimativas das DOAs.

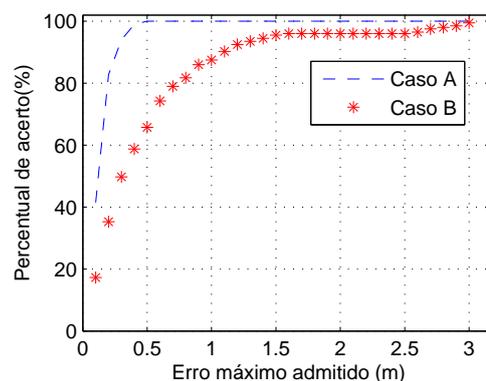


Figura 6.4: Taxa de sucesso média da localização ao variar o erro máximo de localização admitido.

Os resultados da Fig. 6.4 mostram que o Caso A possui desempenho significativamente superior ao Caso B. Para erro máximo de localização de 0.5 m, o Caso A

apresenta taxa de sucesso igual a 100%, enquanto a taxa de sucesso correspondente no Caso B é igual a 65%. O Caso B atinge taxas próximas a 100% apenas quando o erro máximo admitido é de 1.5 m. Esses resultados estão de acordo com a análise teórica apresentada no Cap. 5, da qual foi estabelecido que a distância entre os pares igual a 7 m (d_{PAR} otimizada) permitiria melhor desempenho, dadas as propriedades deste experimento.

6.2 Localização de uma Fonte em Ambiente Simulado com Alta Reverberação

Nesta seção avaliaremos os resultados experimentais no ambiente simulado com T_{60} igual a 800 ms. Seguiremos a metodologia abordada na seção anterior, sendo inicialmente investigada a acurácia das estimativas das DOAs para todas as possíveis posições da fonte e do conjunto de microfones². Em seguida, avaliaremos os erros das estimativas das posições das fontes considerando duas diferentes posições para os arranjos de microfones. Adotaremos a mesma notação da seção anterior: os Casos 1 e 2, utilizados na avaliação dos erros das estimativas das DOAs, correspondem às posições do par de microfones em (1.5,1) e (4.5,1), respectivamente. Os Casos A e B, utilizados na avaliação dos erros de localização, correspondem às posições dos pares de microfones iguais a (1.5,1),(8.5,1) e (4.5,1),(5.5,1), respectivamente.

A Fig. 6.1 apresenta a taxa de sucesso nas estimativas das DOAs com erro máximo admitido $\Delta\theta = 3^\circ$ para o Caso 1 e $\Delta\theta = 1^\circ$ para o Caso 2, para cada posição da fonte. Esses valores de $\Delta\theta$ foram escolhidos por resultarem em $D_{MAX} < 1.2$ m, de acordo com a análise da Seção 5.2.3.

Os resultados da Fig. 6.1(a) para o Caso 1 indicam condições favoráveis para o sistema de localização com fontes nas posições com $y = 3$ m, 5 m e 7 m, com taxas de sucesso das estimativas das DOAs entre 72% e 100%. As posições com $y = 9$ m apresentam taxas de sucesso variando entre 40% e 64%, caracterizando uma região de pior desempenho do sistema de localização.

Os resultados da Fig. 6.1(b), para o Caso 2, mostram baixas taxas de sucesso para praticamente todas as posições, exceto para aquelas próximas aos microfones, (4,3) e (6,3). Assim, o posicionamento do conjunto de microfones do Caso 2 pode ser considerado inapropriado.

A Fig. 6.6 apresenta a variação do valor médio da taxa de sucesso para erro máximo nas estimativas das DOAs entre 0.1° e 10° . Nessa figura, observa-se que há diferenças significativas entre os Casos 1 e 2. Para o erro máximo nas estimativas das DOAs igual a 2° , a taxa de sucesso no Caso 1 é de 80%, podendo-se utilizar

²Essas posições estão mostradas na Fig. 5.5.

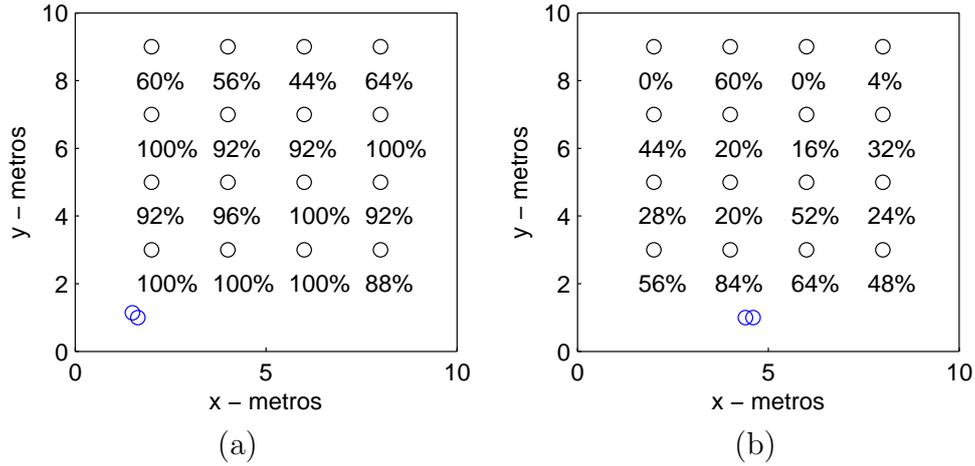


Figura 6.5: Taxa de sucesso nas estimativas das DOAs em cada possível posição da fonte para $T_{60} = 800$ ms, considerando (a) Caso 1 com $\Delta\theta = 3^\circ$ e (b) Caso 2 para $\Delta\theta = 1^\circ$.

os resultados da região apropriada encontrados na Seção 5.2.3 para prever o desempenho do sistema de localização. Para o Caso 2, apenas 55% das estimativas das DOAs obtiveram $\Delta\theta \leq 2^\circ$, não sendo apropriado utilizar o resultado da região robusta da Seção 5.2.3.

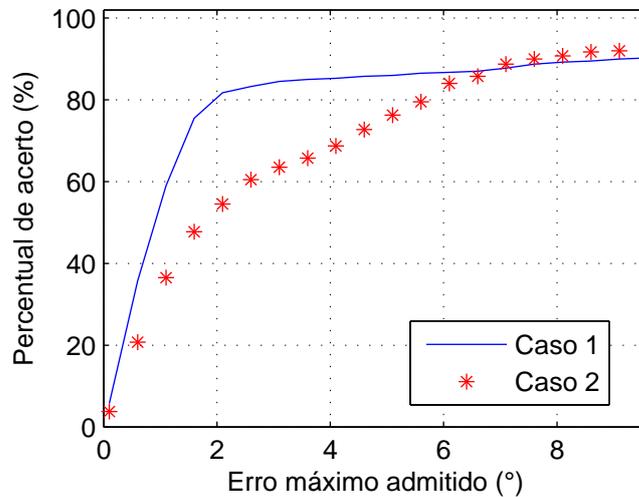


Figura 6.6: Taxa de sucesso média ao variar o erro máximo das estimativas das DOAs $\Delta\theta$ admitido para $T_{60} = 800$ ms.

A Fig. 6.7 apresenta a taxa de sucesso da localização de uma fonte com erro máximo admitido $D_{MAX} = 1.2$ m para cada possível posição da fonte. A Fig. 6.7(a) mostra os percentuais para o Caso A, enquanto que a Fig. 6.7(b) mostra os percentuais correspondentes para o Caso B.

Os resultados do Caso A para as posições com $y = 3$ m, 5 m e 7 m mostraram boas taxas de sucesso, variando entre 80% e 100%, conforme esperado, uma vez que

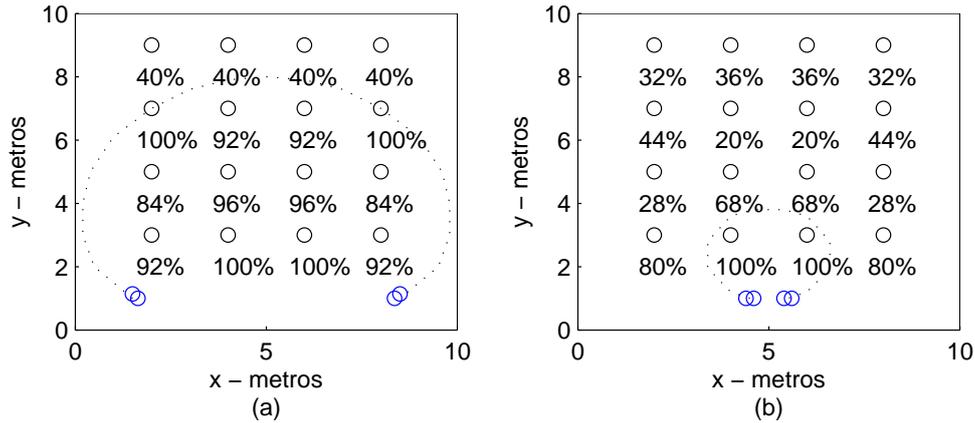


Figura 6.7: Taxa de sucesso da localização em cada possível posição da fonte, com $T_{60} = 800$ ms, considerando erro máximo admitido $D_{MAX} = 1.2$ m para (a) Caso A e (b) Caso (B).

essas posições estão dentro da região apropriada estabelecida na Fig. 5.13 na Seção 5.2.3 (mostradas em linhas tracejadas na Fig. 6.7). As posições com $y = 9$ m estão fora dessa região e apresentaram baixas taxas de sucesso, iguais ou inferiores a 40%.

Os resultados do Caso B para as posições com $y = 7$ m e 9 m mostraram baixas taxas de sucesso, conforme esperado devido ao desempenho ruim dos algoritmos de estimação das DOAs em relação ao exigido. As posições (3,4), (3,6), (5,4) e (5,6) toleram erros de estimação das DOAs maiores, pois elas estão na região apropriada da Fig. 5.14, admitindo $\Delta\theta = 2^\circ$ para taxas de sucesso na Fig. 6.6 iguais a 57%. As posições (2,3) e (8,3), apesar de estarem fora da região apropriada do Caso B, apresentaram taxas de sucesso igual a 80%, superando as expectativas. Provavelmente essas posições foram favorecidas pela pequena distância entre os microfones e a fonte, associado ao fato do modelo proposto dar relativamente grande peso no cálculo do D_{ER} para a posição com menor diferença nas DOAs estimadas e mais distante da posição da fonte.

A Fig. 6.8 apresenta o comportamento do valor médio da taxa de sucesso ao variar o erro máximo admitido nas estimativas das posições, considerando D_{MAX} entre 0.1 e 3 m, para os Casos A e B.

Os resultados da Fig. 6.8 mostram que o Caso A possui desempenho significativamente superior ao Caso B, conforme esperado pela análise do Cap. 5. O valor de d_{PAR} no Caso A corresponde ao valor ótimo desse parâmetro segundo a análise, admitindo uma região significativamente maior para localização robusta do que a do Caso B. Para erro máximo de localização de 0.5 m, o Caso A apresenta taxa de sucesso de cerca de 70%, enquanto a taxa de sucesso no Caso B é de apenas 30%.

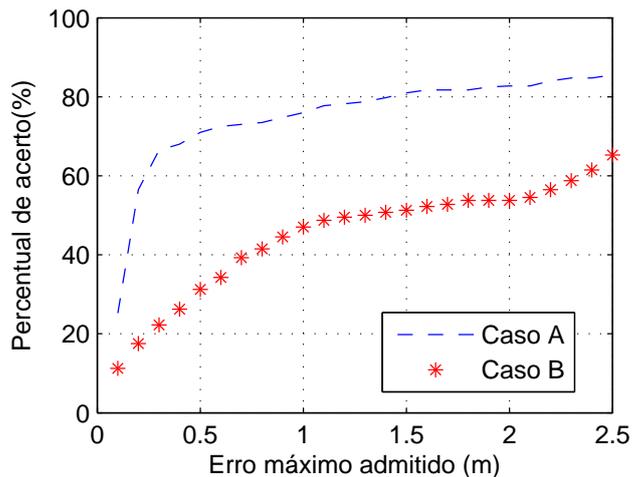


Figura 6.8: Taxa de sucesso média da localização ao variar o erro máximo de localização admitido para $T_{60} = 800$ ms.

6.3 Influência da Distância entre Arranjos e da Reverberação na Localização

Nesta seção avaliaremos, através de experimentos em ambientes simulados, a influência do tempo de reverberação T_{60} e da distância entre os conjuntos de microfones d_{PAR} na acurácia do sistema de localização, e compararemos os resultados dos experimentos com os obtidos nas Seções 5.2.2 e 5.2.3.

As posições das fontes e dimensões da sala desses experimentos são iguais às utilizadas anteriormente. Foram acrescentadas as posições dos arranjos de microfones $d_{PAR} = 3$ m e 5 m, com os respectivos pares nas posições (3.5,1);(6.5,1) e (2.5,1);(7.5,1). Além de $T_{60} = 200$ e 800 ms, foram simulados ambientes com $T_{60} = 400$ e 600 ms.

A Fig. 6.9 mostra os valores da taxa média de acerto obtidos nos experimentos ao se variar D_{MAX} , para os quatro valores de T_{60} e de d_{PAR} especificados acima, considerando a presença de uma única fonte e utilizando o algoritmo GCC-PHAT para estimar as DOAs.

Os resultados da Fig. 6.9 mostram que para todos os tempos de reverberação testados, as melhores taxas de acerto foram obtidas com $d_{PAR} = 5$ m e 7 m. Para $T_{60} = 800$ ms, há uma pequena degradação nessas taxas quando d_{PAR} é reduzido para 5 m. No entanto, à medida que os arranjos de microfones são aproximados, a performance do sistema piora significativamente, sendo esse efeito mais acentuado para ambientes com maiores tempos de reverberação. Com $d_{PAR} = 1$ m, a taxa de sucesso supera 80% apenas em ambientes com baixa reverberação ($T_{60} = 200$ e 400 ms) para erro máximo admitido em torno de 1 m. Já para $d_{PAR} = 5$ m e 7 m, essa mesma taxa é obtida com erros de localização abaixo de 40 cm para

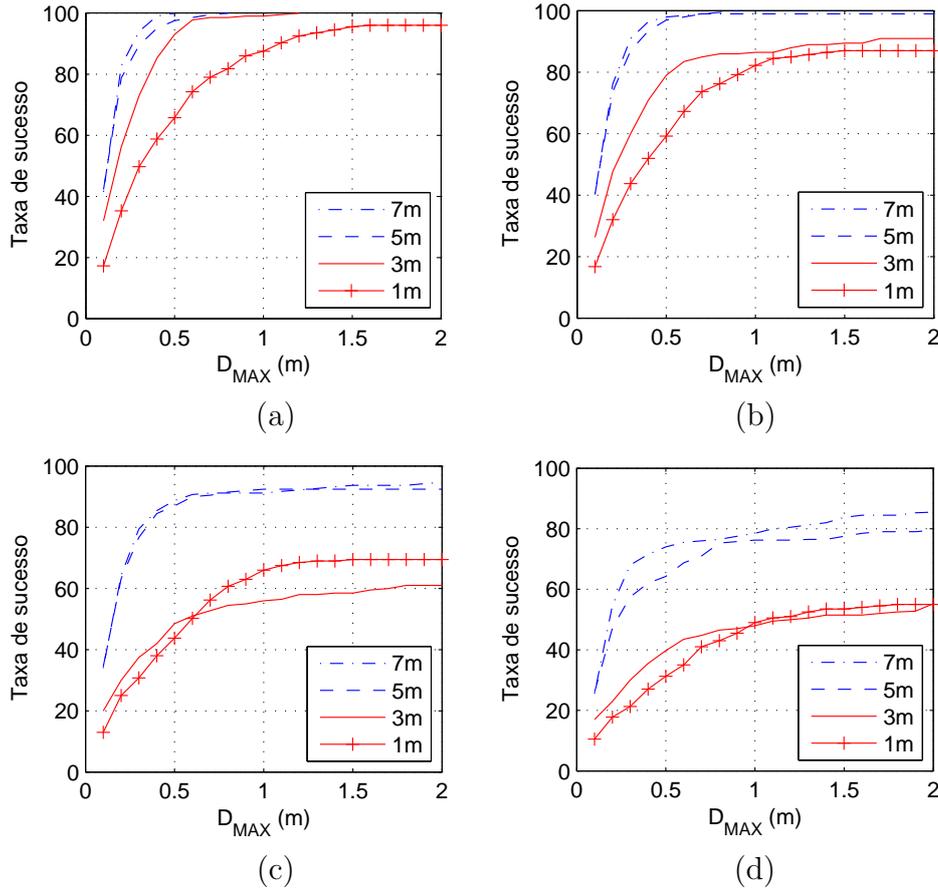


Figura 6.9: Taxa de sucesso média na localização de uma fonte ao variar D_{MAX} para o ambiente simulado com T_{60} igual a (a) 200, (b) 400, (c) 600 e (d) 800 ms.

ambientes com baixa e média reverberação, e em torno de 1 m para ambientes com alta reverberação. Esses resultados estão de acordo com a análise apresentada na Seção 5.2.2, a qual determinou $d_{PAR} = 7$ m como o valor correspondente à melhor performance do sistema, e indicou uma leve deterioração para $d_{PAR} = 5$ m, e uma deterioração significativa de performance para $d_{PAR} = 1$ m.

Avaliando os resultados da Fig. 6.9 frente à variação do T_{60} , é possível concluir que o emprego de d_{PAR} otimizado torna o sistema de localização mais robusto, viabilizando o seu emprego mesmo em ambientes com alta reverberação. Para ambientes com tratamento acústico para reduzir a reverberação, pode-se esperar, de acordo com as Figs. 6.9(a) e 6.9(b), taxas de sucesso em torno de 100% para erro máximo admitido igual 50 cm usando-se $d_{PAR} = 5$ ou 7 m.

6.4 Localização de Uma Fonte no Ambiente Real

Nesta seção, avaliamos os resultados da localização de uma fonte no ambiente real, descrito no Cap. 5, com tempo de reverberação T_{60} medido igual a 900 ms. Seguindo

a metodologia utilizada para os ambientes simulados, inicialmente investigaremos a acurácia das estimativas das DOAs para todas as possíveis posições da fonte e do conjunto de microfones, mostradas na Fig. 5.16. Em seguida, avaliaremos os erros das estimativas das posições das fontes considerando os pares de arranjos ([M1,M2] e [M7,M8]) e ([M3,M4] e [M5,M6]) da Fig. 5.16.

A Fig. 6.10 apresenta as taxas de sucesso nas estimativas das DOAs em cada par de microfones³ para todas as possíveis posições da fonte. As Figs. 6.10(a) e 6.10(b) apresentam as taxas de sucesso com os arranjos [M1,M2] (denominado de Caso 1) e [M7,M8] (denominado de Caso 2), respectivamente, para $\Delta\theta = 7^\circ$, enquanto que as Figs. 6.10(c) e 6.10(d) apresentam as taxas de sucesso com os arranjos [M3,M4] (denominado de Caso 3) e [M5,M6] (denominado de Caso 4), respectivamente, para $\Delta\theta = 2^\circ$. Esses valores de $\Delta\theta$ foram escolhidos por resultarem em erros de localização $D_{MAX} < 1.2$ m, de acordo com a análise da Seção 5.3. A falta de simetria dos resultados obtidos nos Casos 1 e 2, e nos Casos 3 e 4, pode ser explicada pelos diferentes materiais da sala, que contém janelas de vidro ao longo de uma das paredes laterais, e pela presença de um pilar, reflexões e difração nas mesas de fórmicas e madeiras.

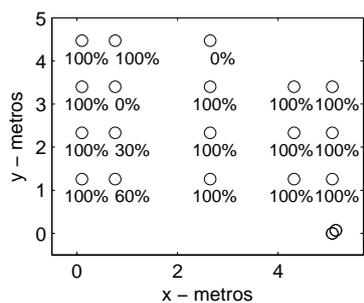
Analisando as taxas de sucesso das estimativas das DOAs para as posições mostradas na Fig. 6.5, verificam-se no Caso 1 taxas de sucesso inferiores a 70% nas posições (0.8,1.26), (0.8,2.35), (0.8,3.40) e (2.6,4.3). Quando combinadas às estimativas das DOAs do Caso 2, essas estimativas ruins podem causar mau desempenho do sistema de localização para fontes nessas posições. Nas outras posições, as taxas de sucesso foram de 100%, indicando potencial bom desempenho do sistema.

No Caso 2, verificam-se taxas de sucesso inferiores a 70% nas posições (0.1,1.26), (4.3,1.26), (4.3,2.35), (0.8,4.4), (2.6,4.4). Nos Casos 3 e 4, as taxas de sucesso são predominantemente baixas, com médias iguais a 23% e 15%, respectivamente. A julgar por estes rendimentos, os Casos 3 e 4 podem ser considerados inapropriados pela análise da região apropriada apresentada no Cap. 5.

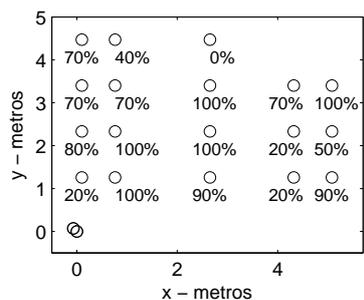
Portanto, da Fig. 6.10 podemos concluir que o posicionamento dos microfones nas extremidades da região de localização, como nos Casos 1 e 2, e com inclinações respectivas de $+45^\circ$ e -45° , resultaram nas melhores taxas de sucesso, com médias respectivas de 83% e 66%. Esses posicionamentos dos conjuntos de microfones também proporcionaram os melhores rendimentos no ambiente simulado com $T_{60} = 800$ ms.

A Fig. 6.11 apresenta as taxas de sucesso médias variando-se o erro máximo admitido nas estimativas das DOAs entre 0° e 15° . As médias foram feitas descartando-se os resultados com $y = 4.47$ m, ou seja, considerando apenas as posições com

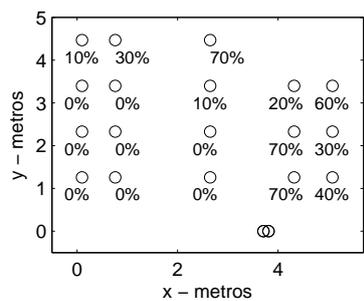
³Ao contrário do ambiente simulado, no ambiente real não há simetria das posições das fontes em relação aos arranjos de microfones e, portanto, as DOAs nos 4 arranjos foram estimadas.



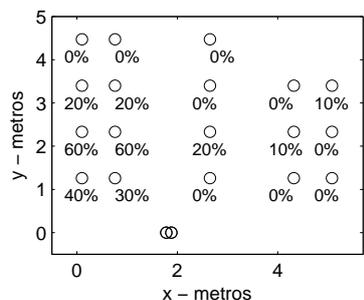
(a)



(b)



(c)



(d)

Figura 6.10: Taxa de sucesso nas estimativas das DOAs em cada possível posição da fonte considerando $\Delta\theta = 7^\circ$ para (a) Caso 1 e (b) Caso 2, e $\Delta\theta = 2^\circ$ para (c) Caso 3 e (d) Caso 4.

simetria, de modo a tornar a comparação entre os quatro casos mais justa.

Analisando os resultados da Fig. 6.11, verifica-se para o Caso 1 que a taxa média de acertos é de 97% para $\Delta\theta \geq 7.5^\circ$, enquanto que para o Caso 2 essa taxa é obtida para $\Delta\theta \geq 11^\circ$. Já os Casos 3 e 4 não alcançam a taxa de sucesso de 80%. Os respectivos erros máximos admitidos nos Casos 1 e 2 para obter 80% de sucesso são

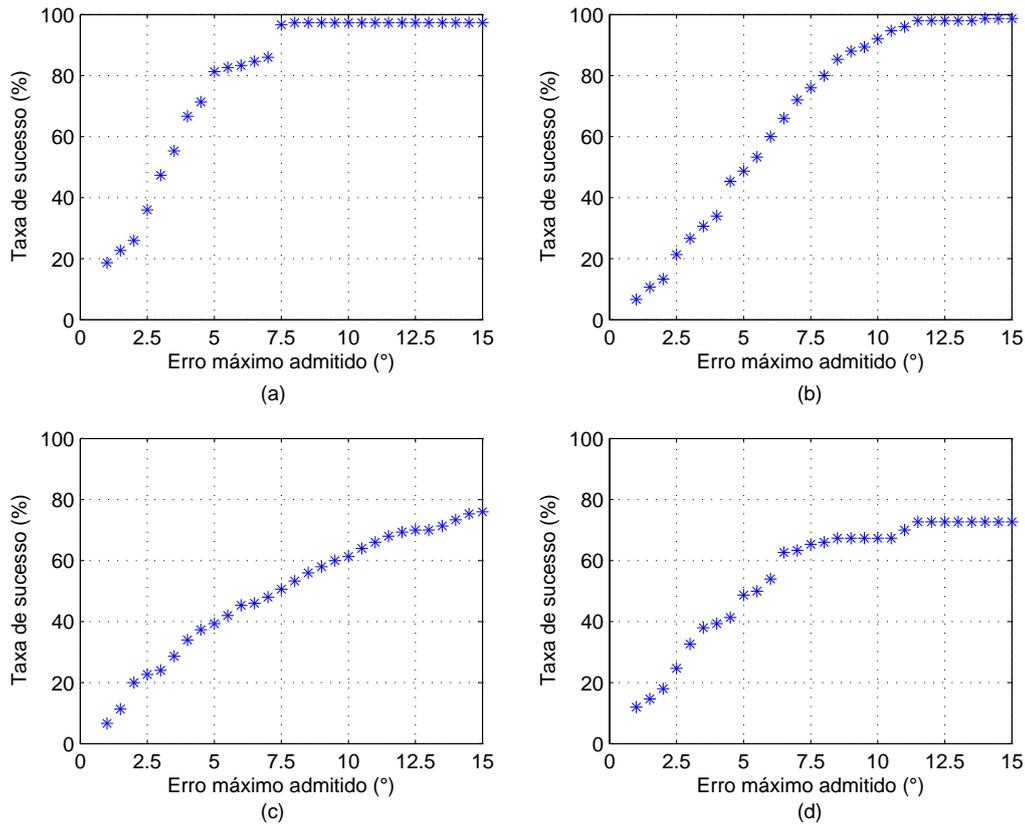


Figura 6.11: Taxa de sucesso média nas estimativas das DOAs ao variar o erro máximo admitido para (a) Caso 1, (b) Caso 2, (c) Caso 3 e (d) Caso 4.

iguais a 5° e 7.5° . Observa-se também que os Casos 1 e 4 apresentaram performances predominantemente melhores quando comparados aos geometricamente simétricos Casos 2 e 3. Esse resultado pode ser explicado pela falta de simetria dos materiais das paredes da sala. Finalmente, conclui-se que, devido às baixas taxas de sucesso nos Casos 3 e 4, a combinação das suas DOAs não deve resultar na localização robusta da fonte.

A Fig. 6.12 apresenta a taxa de sucesso da localização de uma fonte com erro máximo admitido $D_{MAX} = 1.2$ m para cada possível posição da fonte. A Fig. 6.12(a) mostra os percentuais de estimativas das posições com erros menores que 1.2 m obtidas com os dois conjuntos de microfones [M1,M2] e [M7,M8], denominado de Caso A, enquanto que a Fig. 6.12(b) mostra os percentuais correspondentes com os conjuntos de microfones [M3,M4] e [M5,M6], denominado de Caso B.

Analisando-se os resultados do Caso A, verifica-se que a maior parte das posições, as com $y \leq 4$ m, estão dentro das regiões apropriadas das Figs. 5.19(a) e 5.19(b) (mostradas em tracejado na Fig. 6.12), correspondentes a $\Delta\theta = 5^\circ$ e 7° , respectivamente, e apresentam taxas de sucesso predominantemente altas. A taxa de sucesso na posição (0.8,1.25) é igual a apenas 60%, ocasionada pela baixa acurácia da DOA

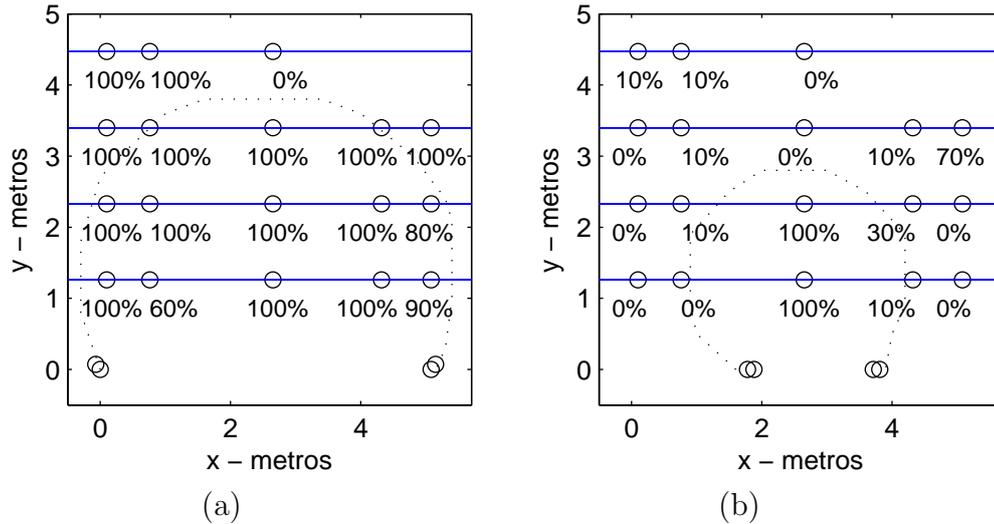


Figura 6.12: Taxa de sucesso da localização em cada possível posição da fonte considerando $D_{MAX} = 1.2$ m para (a) Caso A e (b) Caso B.

associada à distância entre a fonte e o conjunto de microfones do Caso 1. As posições com $y = 2.35$ m e com $y = 3.4$ m possuem taxa de sucesso média de 96% e 100%, respectivamente, estando de acordo com as expectativas estabelecidas na Fig. 5.19. A taxa de sucesso de 0% na posição (2.6,4.47) pode ser explicada observando-se da Fig. 5.19 que a taxa de sucesso para $\Delta\theta = 5^\circ$, exigida para que esta posição esteja na região apropriada, é pequena. As posições com $y = 5.54$ m e 6.61 m estão fora das regiões apropriadas da Fig. 5.19 e apresentaram baixas taxas de sucesso, como esperado.

Analisando os resultados do Caso B, verifica-se que as posições (1.26,2.6) e (2.35,2.6) estão dentro das regiões apropriadas das Figs. 5.20(a)-(c) e apresentam taxas de sucesso iguais a 100%. A posição (5.1,3.4) está dentro da região apropriada apenas da Fig. 5.20(a), mostrando que requer erros menores nas estimativas das DOAs, e apresentou taxa de sucesso igual a 70%, superando as expectativas. As demais posições estão fora da região apropriada da Fig. 5.20(c), compatível com os erros nas estimativas das DOAs, e apresentaram taxas de sucesso baixas.

A Tabela 6.1 apresenta os valores médios da taxa de sucesso, considerando todas as posições com um mesmo valor de y . Observa-se desta tabela que no Caso A as posições com $y = 1.26$ m, 2.36 m e 3.4 m apresentaram excelentes resultados, o que não foi observado no Caso B. A taxa percentual média de sucesso foi de 60% para o Caso A e de 17% para o Caso B, confirmando a análise teórica do Cap. 5.

A Fig. 6.13 apresenta a variação do valor médio da taxa de sucesso para erro máximo admitido nas estimativas das posições D_{MAX} entre 0 e 2.5 m, considerando apenas as posições com $y = 1.26$ m, 2.36 m e 3.40 m, para os Casos A e B.

Os resultados da Fig. 6.13 mostram que o Caso A possui desempenho signifi-

Percentual de erros menor que 1.2 m							
Posições	Média						Média
	y = 1.26	y = 2.36	y = 3.40	y = 4.47	y = 5.54	y = 6.61	
Caso A	90%	96%	100%	67%	10%	0%	60%
Caso B	22%	28%	18%	7%	17%	10%	17%

Tabela 6.1: Taxa de sucesso médio das estimativas das posições no ambiente real com 1 fonte.

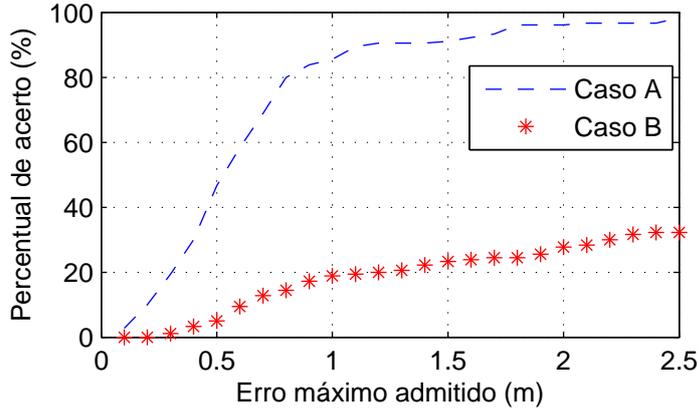


Figura 6.13: Taxa de sucesso média da localização ao variar D_{MAX} para o ambiente real.

cativamente superior ao Caso B. Para $D_{MAX} = 50$ cm, o Caso A apresenta taxa de sucesso de 50%, enquanto que para o Caso B essa taxa é de apenas 5%; já para $D_{MAX} = 1$ m, o Caso A possui um percentual de acerto de 86%, enquanto que o Caso B possui 19%.

Esses resultados estão de acordo com as análises teóricas do Cap. 5, onde foi estabelecido que o valor d_{PAR} otimizado, correspondente ao Caso A, permitiria melhor desempenho, dadas as propriedades deste experimento. A utilização da distância inapropriada entre os pares de microfones no Caso B inviabiliza a localização da fonte.

6.5 Considerações Finais sobre a Localização de Uma Fonte

O sistema de localização para uma fonte apresentou excelente performance no ambiente simulado com $T_{60} = 200$ ms, em consequência da quantidade significativa de estimativas das DOAs (obtidas pelo método GCC-PHAT) apresentarem erros de até 1° , permitindo que a realização da localização com baixos erros nas estimativas das posições ocorresse em uma grande região do recinto.

A utilização do sistema em um ambiente com $T_{60} = 800$ ms provocou uma

redução significativa na acurácia da localização. Porém, posicionando-se adequadamente os arranjos de microfones, obteve-se taxa de sucesso de cerca de 70% nas estimativas da posição da fonte para erro máximo admitido igual a 0.5 m. As DOAs apresentaram erro máximo de 2° em aproximadamente 80% dos valores estimados.

No ambiente real em que os testes foram realizados, com tempo de reverberação medido $T_{60} \approx 900$ ms, obteve-se taxa de sucesso média igual a 90% para erro máximo de localização admitido igual a 1.2 m, utilizando-se as posições otimizadas dos pares de microfones.

O procedimento proposto para determinar a região apropriada para localização da fonte foi validado com experimentos simulados e medidos. A análise abordada na Seção 5.3 permitiu a escolha da distância entre os microfones de forma otimizada, melhorando significativamente o desempenho do sistema de localização em todos os testes realizados.

Capítulo 7

Resultados da Localização de Duas Fontes Ativas

Neste capítulo, avaliaremos os resultados obtidos com o sistema de localização quando duas fontes encontram-se ativas no período analisado¹. Consideraremos os ambientes simulados e real apresentados no Cap. 5. As estimativas das DOAs são obtidas através dos algoritmos de separação cega de fontes descritos no Cap. 4. O objetivo principal destes experimentos é avaliar o desempenho do sistema de localização proposto na presença de múltiplas fontes e a eficiência do procedimento de estimação da região apropriada e da escolha de d_{PAR} apresentado no Cap. 5, com a presença das imprecisões introduzidas pelos processos de mistura e de separação das fontes.

As características dos ambientes simulados e real foram apresentadas nas Seções 5.2.1 e 5.3.1 e os algoritmos para estimação das DOAs a partir dos parâmetros fornecidos pelos métodos de BSS foram descritos nas Seções 4.2 a 4.5. A frequência de amostragem utilizada foi igual a 16 kHz. O algoritmo BSS no domínio da frequência utilizou a STFT com janelas de Hanning de 2048 pontos com saltos de 1024 amostras, e FFTs de tamanho 2048.

Os testes foram realizados com duas fontes ativas simultaneamente, utilizando todas as combinações entre as posições pre-determinadas, sendo as DOAs das fontes em cada conjunto de microfones estimadas e, em seguida, as posições determinadas pelo encontro das retas correspondentes às DOAs estimadas.

7.1 Método do Campo Distante

Nesta seção, avaliaremos o desempenho do sistema de localização através de resultados experimentais em ambiente simulado com as DOAs estimadas pelo algoritmo

¹O método de localização proposto pode ser facilmente estendido para mais de duas fontes ativas, requerendo o aumento correspondente do número de microfones de cada arranjo.

do campo distante apresentado na Seção 4.2.1.

7.1.1 Ambiente Simulado com Baixa Reverberação

Nesta seção, apresentaremos os resultados experimentais no ambiente simulado com $T_{60} = 200$ ms. Seguiremos a metodologia abordada na Seção 6.1, sendo inicialmente investigada a acurácia das estimativas das DOAs para todas as possíveis posições das fontes e do conjunto de microfones² e, em seguida, serão avaliados os erros das estimativas das posições das fontes considerando duas diferentes posições para os arranjos de microfones. Adotaremos a notação do Cap. 6: os Casos 1 e 2, utilizados na avaliação do erro das estimativas das DOAs, correspondem às posições do par de microfones em (1.5,1) e (4.5,1), respectivamente; já os Casos A e B, utilizados na avaliação dos erros de localização, correspondem às posições dos arranjos de microfones iguais a (1.5,1),(8.5,1) e (4.5,1),(5.5,1), respectivamente.

A Fig. 7.1 apresenta a taxa de sucesso média nas estimativas das DOAs com erro máximo admitido $\Delta\theta = 5^\circ$ para o Caso 1 e $\Delta\theta = 2^\circ$ para o Caso 2, para uma das fontes na posição indicada pelo valor percentual, sendo a média calculada colocando-se a segunda fonte em todas as outras posições e obtendo-se a taxa de sucesso correspondente. Esses valores de $\Delta\theta$ foram escolhidos por resultar em erros de localização $D_{MAX} < 1.2$ m, de acordo com as análises das Figs. 5.13 e 5.14 da Seção 5.2.3. O sinal de voz utilizado para cada fonte foi formado pela concatenação de gravações de uma voz masculina e uma feminina, sendo usados locutores diferentes para cada fonte.

Na Fig. 7.1(a) observa-se que o método do Campo Distante apresentou bons resultados no Caso 1, com valor médio da taxa de acerto igual a 78%, criando boa perspectiva de sucesso na etapa de localização, como mostram as propriedades na Seção 5.2.3. A posição (2,5) apresenta taxa de sucesso igual a 47%. Porém, na localização da fonte nessa posição, a estimativa da DOA será combinada com a estimativa da DOA no par de microfones simetricamente posicionado, cuja taxa de sucesso é de 87%, igual à da posição simétrica (8,5) no Caso 1. Essa posição admitirá erros nas estimativas da DOA de até 7° , como indicado na Fig. 5.13, apresentando uma boa perspectiva de sucesso na localização.

As posições (2,7) e (8,7) estão no limite entre a região apropriada e a não apropriada na Fig. 5.13(b). Uma vez que a taxa de sucesso é igual a 50% na posição (8,7) e 80% na posição simétrica (2,7), espera-se um valor próximo do máximo tolerável na localização.

As posições (4,7) e (6,7) estão na região apropriada da Fig. 5.13(b), com as respectivas taxas de acerto iguais a 87% e 90%, criando boas expectativas em relação

²Essas posições estão mostradas na Fig. 5.5.

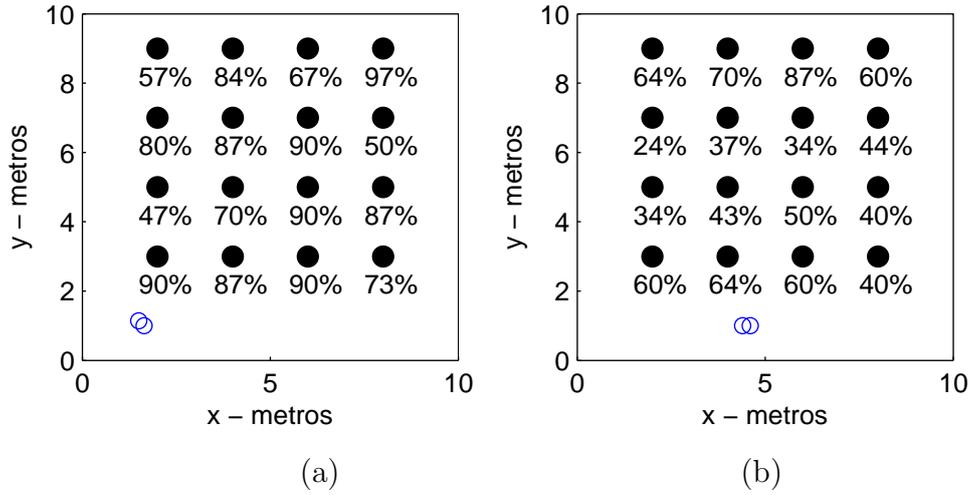


Figura 7.1: Taxa de sucesso nas estimativas das DOAs usando o método do Campo Distante em cada possível posição da fonte para $T_{60} = 200$ ms, considerando (a) Caso 1 com $\Delta\theta = 5^\circ$ e (b) Caso 2 para $\Delta\theta = 2^\circ$.

a erros de localização.

As posições com $y = 9$ m estão na região apropriada da Fig. 5.13(a), admitindo o erro máximo nas estimativas das DOAs igual a 3° . As taxas de sucesso para $\Delta\theta = 3^\circ$ nas posições (2,9), (4,9), (6,9) e (8,9) são respectivamente iguais a 44%, 67%, 60% e 80%, esperando-se valores de erros moderados na localização.

Na Fig. 7.1(b) observa-se que o método do Campo Distante apresentou taxas de sucesso baixas no Caso 2 para $\Delta\theta = 2^\circ$. Portanto, considerando que erros máximos dessa ordem nas estimativas das DOAs são necessários, de acordo com os resultados da Fig. 5.14, estimamos que o posicionamento dos conjuntos de microfones do Caso 2 seja inapropriado para a localização.

A Fig. 7.2 apresenta a variação do valor médio da taxa de sucesso para um erro máximo nas estimativas das DOAs entre 1° e 15° . Nessa figura, observa-se que não há diferenças significativas entre os Casos 1 e 2, e que para o erro máximo nas estimativas das DOAs igual a 5° a taxa de sucesso é 80%, e para 10° é praticamente 100%.

A Fig. 7.3 apresenta a taxa de sucesso da localização de uma das fontes com erro máximo admitido $D_{MAX} = 1.2$ m para cada possível posição. Para obter essa taxa, a segunda fonte foi colocada em todas as outras posições, avaliando-se em cada caso o sucesso da localização da primeira fonte. A Fig. 6.7(a) mostra os percentuais para o Caso A, enquanto que a Fig. 6.7(b) mostra os percentuais correspondentes para o Caso B.

Os resultados do Caso A para as posições com $y = 3$, $y = 5$, (4,7) e (6,7)

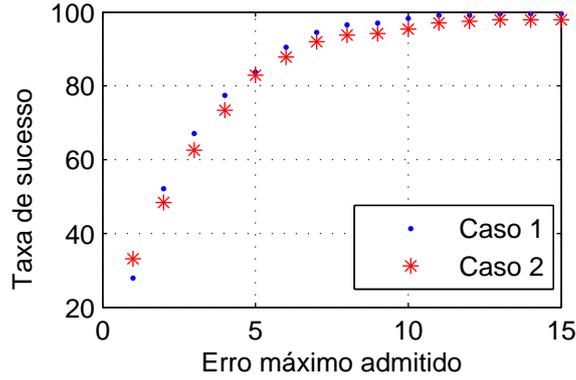


Figura 7.2: Taxa de sucesso média ao variar $\Delta\theta$, com $T_{60} = 200$ ms e utilizando o método do Campo Distante.

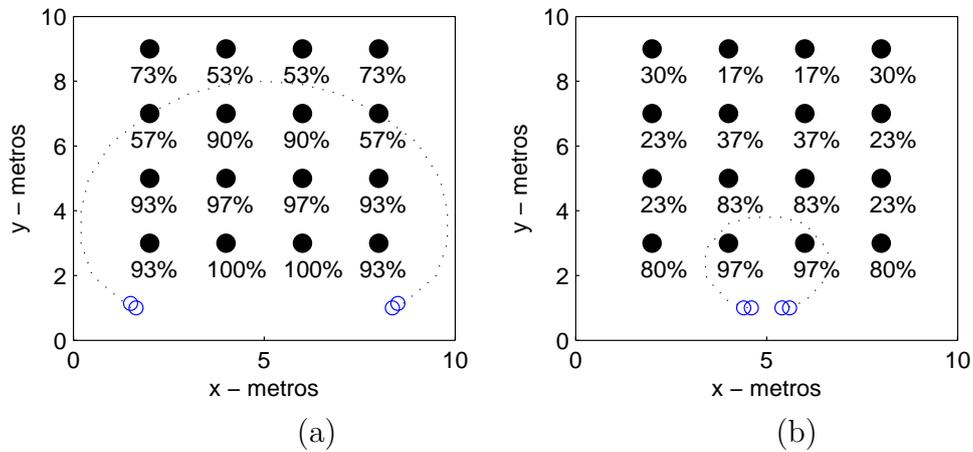


Figura 7.3: Taxa de sucesso da localização em cada possível posição da fonte, com $T_{60} = 200$ ms e usando o método do Campo Distante, considerando $D_{MAX} = 1.2$ m para (a) Caso A e (b) Caso (B).

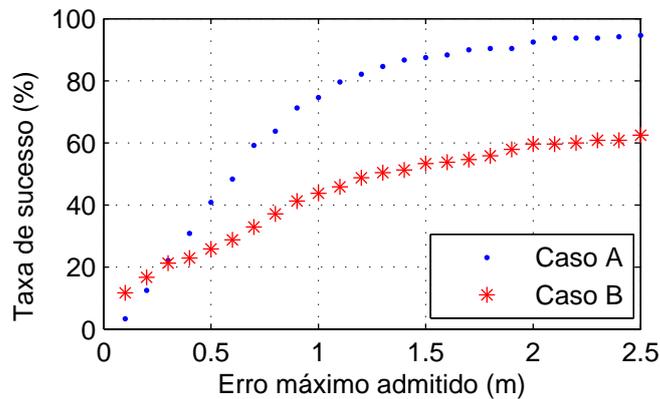


Figura 7.4: Taxa de sucesso média da localização ao variar D_{MAX} no ambiente simulado com $T_{60} = 200$ ms usando o método do Campo Distante.

estão dentro da região apropriada da Fig. 5.13(b) (mostrada em tracejado na Fig. 7.3) e apresentaram altas taxas de sucesso, sendo a taxa mínima igual a 90%. As posições (2,9) e (8,9) apresentaram taxas de sucesso iguais a 73%, resultante da boa performance obtida nas estimativas das DOAs na posição (8,9) do Caso 1. As posições (2,7), (8,7), (4,9) e (6,9) estão fora da região apropriada da Fig. 5.13(b) e dentro região apropriada da Fig. 5.13(a), e apresentaram taxas de sucesso em torno de 55%, que são compatíveis com as taxas de sucesso obtidas na estimação das suas DOAs.

No Caso B, as posições (3,4) e (3,6) possuem taxas de acerto iguais a 97%, sendo esse valor esperado por estarem na região apropriada da Fig. 5.14(b).

As posições (2,3), (8,3), (4,5) e (8,5) possuem taxas de sucesso em torno de 80%, apesar de estarem fora da região apropriada. Nesses casos a posição estimada a partir das DOAs com menor diferença (posição P_{E_n} mais distante na Fig. 5.2) possui grande peso no cálculo do D_E da Eq. (5.9), e o modelo se torna muito severo ao utilizar a média. Porém, a julgar pela probabilidade, admitindo o erro de 4º nas estimativas das DOAs, a quantidade de erros de até 1.2 m nas posições estimadas no cálculo do D_E representaram 50% dos casos. Esse fato, associado à pequena distância entre a fonte e o microfone, que favorece a acurácia nas estimativas das DOAs, beneficiaram a performance dessas posições.

A proposta da região apropriada indicou o d_{PAR} do Caso A como melhor escolha para esse parâmetro, conforme mostram as taxas de acerto médias para $D_{MAX} = 1.2$ m da Fig. 7.3, iguais a 82% e 49% para os Casos A e B, respectivamente. A julgar pela taxa de sucesso por posição, o Caso A foi superior em todas as situações.

A Fig. 7.4 mostra o comportamento do valor médio da taxa de sucesso ao variar o erro máximo de localização entre 0.1 m e 2.5 m, para os Casos A e B.

Os resultados da Fig. 7.4 mostram que o Caso A possui desempenho significativamente superior ao do Caso B, conforme antecipado pela análise apresentada no Cap. 5. Para $D_{MAX} \geq 0.3$ m, o Caso A apresenta taxas de sucesso maiores que o Caso B, sendo que para $D_{MAX} = 1.5$ m, a taxa de sucesso do Caso A é igual a 90%, enquanto que o Caso B não atinge esta taxa para nenhum valor de D_{MAX} no intervalo analisado. Então, é possível afirmar que a escolha da d_{PAR} otimizada permitiu melhorar o desempenho do sistema de localização, indicando que o procedimento proposto para encontrar a região apropriada de acordo com d_{PAR} e D_{MAX} foi bem sucedido.

7.1.2 Ambiente Simulado com Alta Reverberação

Nesta seção apresentaremos os resultados experimentais no ambiente simulado com $T_{60} = 800$ ms. Seguiremos a metodologia abordada na seção anterior, utilizando as

mesmas posições de fontes e microfones e os mesmos sinais de voz.

A Fig. 7.5 apresenta a taxa de sucesso nas estimativas das DOAs com erro máximo admitido $\Delta\theta = 7^\circ$ para os Casos 1 e 2, para cada posição da fonte. Esse valor de $\Delta\theta$ foi escolhido devido à baixa acurácia das DOAs em ambientes com alta reverberação, sendo adequado considerar as regiões apropriadas das Figs. 5.13(c) e 5.14(c) da Seção 5.2.3.

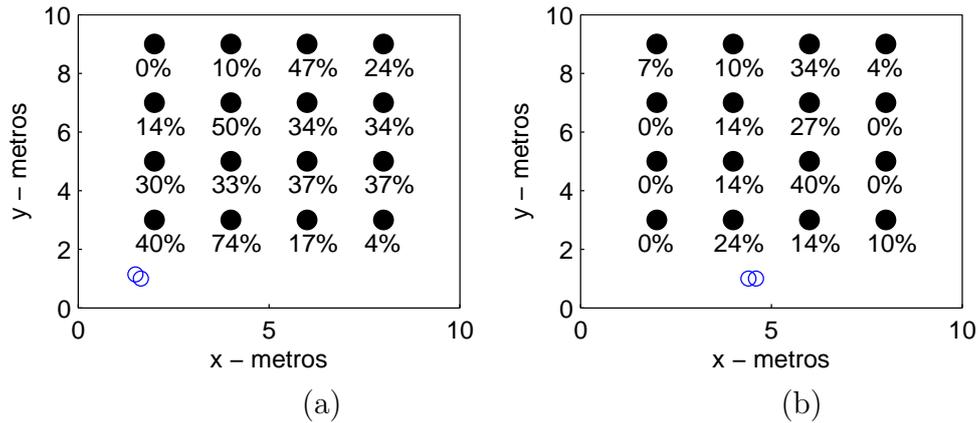


Figura 7.5: Taxa de sucesso nas estimativas das DOAs usando o método do Campo Distante em cada possível posição da fonte para $T_{60} = 800$ ms e $\Delta\theta = 7^\circ$, considerando (a) Caso 1 e (b) Caso 2.

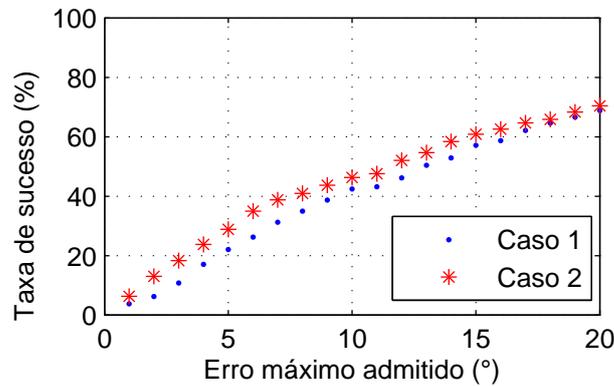


Figura 7.6: Taxa de sucesso média ao variar $\Delta\theta$, com $T_{60} = 800$ ms e utilizando o método do Campo Distante.

Na Fig. 7.5, observa-se que o método do Campo Distante apresentou baixas taxas de sucesso para todas as posições tanto para o Caso 1 quanto para o Caso 2, com valores médios iguais a 30% e 12%, respectivamente. Portanto, espera-se desempenho ruim do sistema de localização para DOAs estimadas pelo método de Campo Distante.

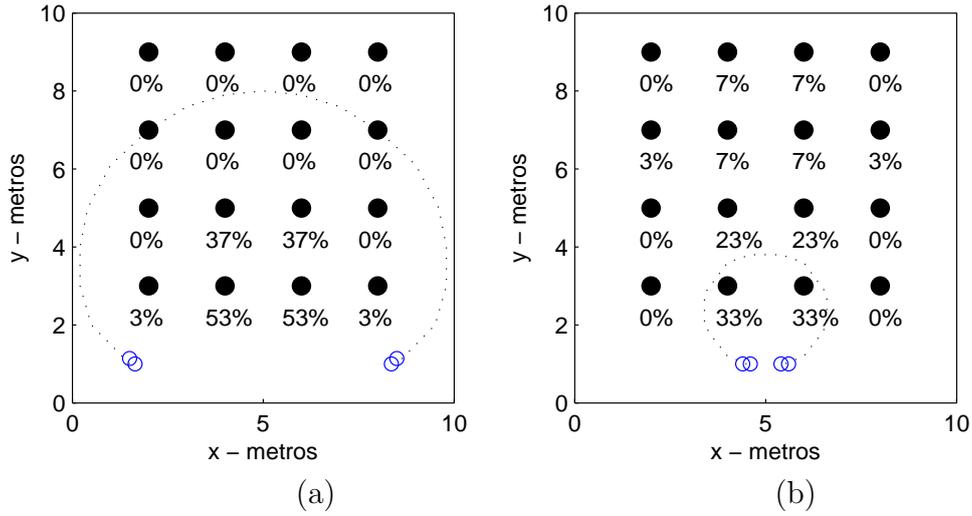


Figura 7.7: Taxa de sucesso da localização em cada possível posição da fonte, com $T_{60} = 800$ ms e usando o método do Campo Distante, considerando $D_{MAX} = 1.2$ m para (a) Caso A e (b) Caso (B).

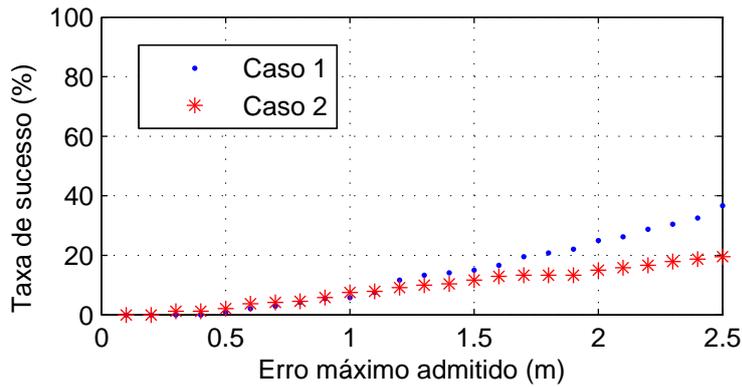


Figura 7.8: Taxa de sucesso média da localização ao variar D_{MAX} no ambiente simulado com $T_{60} = 800$ ms usando o método do Campo Distante.

A Fig. 7.6 apresenta a variação do valor médio da taxa de sucesso para erro máximo nas estimativas das DOAs entre 1° e 20° . Nessa figura, observa-se que não há diferenças significativas entre os Casos 1 e 2, e que a performance esperada do sistema é muito ruim, pois para alcançar a já reduzida região apropriada da Fig. 5.13(c) seriam necessárias estimativas das DOAs com $\Delta\theta = 7^\circ$, sendo a taxa de sucesso para este valor de $\Delta\theta$ menor que 40%.

A Fig. 7.7 apresenta a taxa de sucesso da localização de uma das fontes com erro máximo admitido $D_{MAX} = 1.2$ m para cada possível posição. A Fig. 7.7(a) mostra os percentuais para o Caso A, enquanto que a Fig. 7.7(b) mostra os percentuais correspondentes para o Caso B.

Os resultados dos Casos A e B apresentam baixas taxas de sucesso em todas as posições, cujas respectivas taxas médias de sucesso de localização dos Casos A e B

são 11.67% e 9.17%. Este resultado já era previsto, devido ao baixo desempenho do método do Campo Distante na estimação das DOAs.

A Fig. 7.8 apresenta o comportamento do valor médio da taxa de sucesso para uma variação do erro máximo de localização entre 0.1 m e 2.5 m nos Casos A e B.

Os resultados da Fig. 7.8 mostram que em ambos os casos, as taxas de sucesso são inferiores a 40% para erro máximo de 2.5 m, indicando um baixo desempenho do sistema com o método de Campo Distante para ambientes com alta reverberação.

7.1.3 Ambiente Real

Nesta seção avaliaremos os resultados de localização de duas fontes ativas simultaneamente no ambiente real descrito na Seção 5.3.1, com tempo de reverberação medido $T_{60} = 900$ ms. Seguindo a metodologia utilizada nas seções anteriores, iniciaremos investigando a acurácia das estimativas das DOAs para todas as possíveis posições das fontes e do conjunto de microfones mostradas na Fig. 5.16 e, em seguida, avaliaremos os erros das estimativas das posições das fontes considerando os pares de arranjos ([M1,M2] e [M7,M8]) e ([M3,M4] e [M5,M6]).

A Fig. 7.9 apresenta as taxas de sucesso nas estimativas das DOAs em cada par de microfones para todas as possíveis posições da fonte. As Figs. 7.9(a) e 7.9(b) apresentam as taxas de sucesso com os arranjos [M1,M2] (Caso 1) e [M7,M8] (Caso 2), respectivamente, para $\Delta\theta = 7^\circ$, enquanto que as Figs. 7.9(c) e 7.9(d) apresentam as taxas de sucesso com os arranjos [M3,M4] (Caso 3) e [M5,M6] (Caso 4), respectivamente, para $\Delta\theta = 5^\circ$. Esses valores de $\Delta\theta$ foram escolhidos por resultarem em erros de localização $D_{MAX} < 1.2$ m, de acordo com a análise da Seção 5.3.

Analisando os resultados dos Casos 1 e 2 nas Figs. 7.9(a) e (b), as respectivas taxas de sucesso médias são iguais a 50% e a 23%, respectivamente. Considerando apenas as posições na região apropriada da Fig. 5.19(c), derivada para $\Delta\theta = 9^\circ$, as taxas de sucesso aumentam para 64% e 34%, indicando desempenho aceitável do sistema de localização para essas posições ao utilizar os arranjos de microfones nas posições [M1,M2] e [M7,M8] com o método do Campo Distante.

Os resultados dos Casos 3 e 4, mostrados nas Figs. 7.9(c) e (d), apresentam taxas de sucesso iguais a 17% e 18%, respectivamente. Portanto, considerando que na maioria dos casos os erros nas estimativas das DOAs foram maiores que $\Delta\theta = 5^\circ$, cuja região apropriada correspondente se encontra na Fig. 5.20(b), espera-se desempenho ruim do sistema de localização ao utilizar os arranjos de microfones nas posições [M3,M4] e [M5,M6] com o método do Campo Distante.

As posições com $y \geq 4.47$ apresentaram taxas de sucesso nas estimativas das DOAs muito baixas para os erros máximos exigidos na análise da Seção 5.3. Assim, essas posições não serão consideradas nas demais análises.

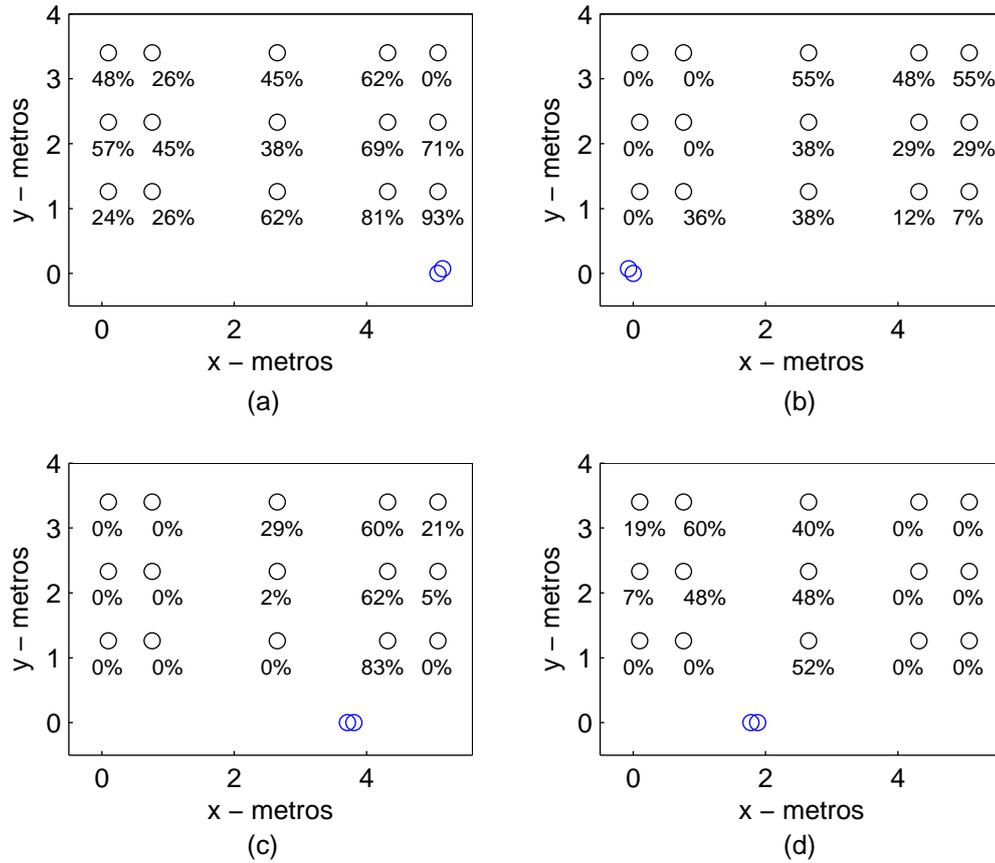


Figura 7.9: Taxa de sucesso nas estimativas das DOAs em cada possível posição da fonte considerando $\Delta\theta = 7^\circ$ para (a) Caso 1 e (b) Caso 2, e $\Delta\theta = 5^\circ$ para (c) Caso 3 e (d) Caso 4.

A Fig. 7.10 apresenta as taxas de sucesso médias variando-se o erro máximo admitido nas estimativas das DOAs entre 0° e 20° .

Analisando os resultados da Fig. 7.10, verifica-se que as melhores taxas de sucesso médias foram obtidas nos Casos 1 e 2, sendo que a diferença nos resultados dos dois casos deve-se à falta de simetria da sala. O Caso 2 apresentou pior desempenho, sendo a taxa de sucesso média igual a apenas 40% para o erro de 10° , indicando desempenho ruim do sistema de localização com o método de Campo Distante, tendo em vista a pequena região apropriada estabelecida na Fig. 5.19(d). Os Casos 3 e 4 apresentaram taxas de sucesso médias ainda mais baixas em todo intervalo de $\Delta\theta$ analisado, indicando uma redução do desempenho do sistema causada pela utilização dos arranjos de microfones nas posições correspondentes a esses casos com o método do Campo Distante.

A Fig. 7.11 apresenta a taxa de sucesso da localização de uma das fontes com erro máximo admitido $D_{MAX} = 1.2$ m para cada possível posição, sendo variada a posição da segunda fonte. A Fig. 7.11(a) mostra os percentuais de estimativas das

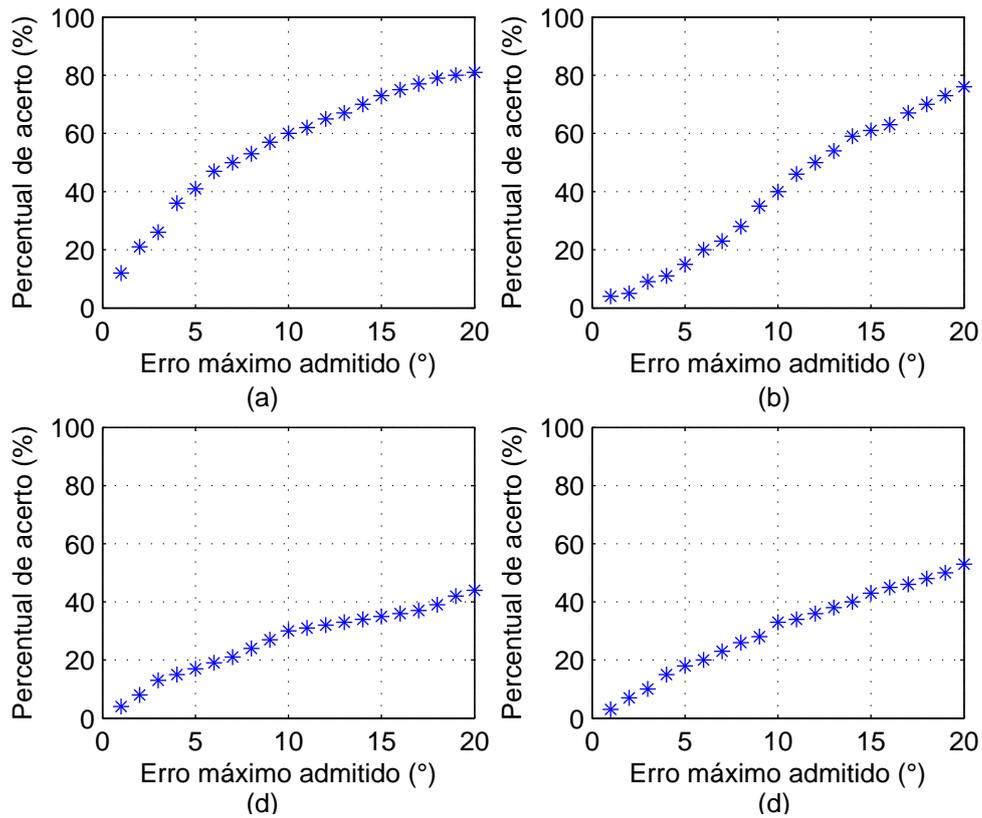


Figura 7.10: Taxa de sucesso média nas estimativas das DOAs ao variar $\Delta\theta$ para (a) Caso 1, (b) Caso 2, (c) Caso 3 e (d) Caso 4.

posições com erros menores que 1.2 m obtidas no Caso A, com os dois conjuntos de microfones [M1,M2] e [M7,M8], enquanto que a Fig. 7.11(b) mostra os percentuais correspondentes para o Caso B, com os conjuntos de microfones [M3,M4] e [M5,M6].

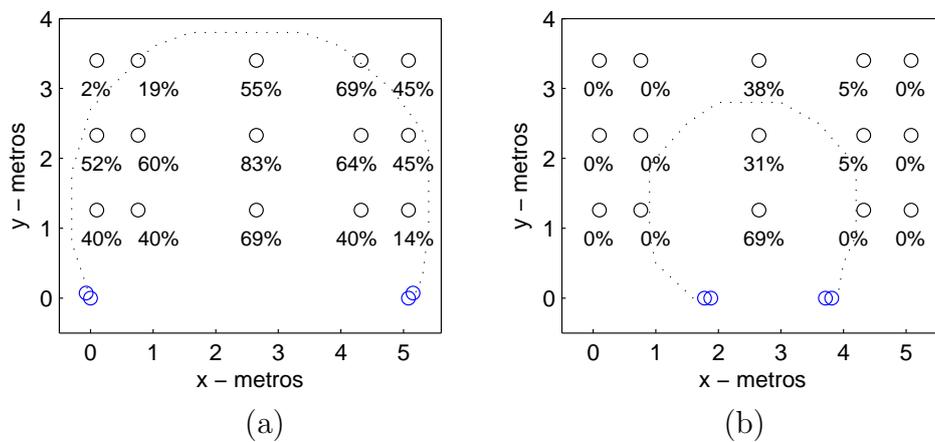


Figura 7.11: Taxa de sucesso da localização em cada possível posição da fonte, usando o método do Campo Distante no Ambiente Real, considerando $D_{MAX} = 1.2$ m para (a) Caso A e (b) Caso B.

Analisando os resultados do Caso A na Fig. 7.11(a), conclui-se que a taxa de sucesso média é igual a 47%. Considerando que as taxas de sucesso média para erros nas DOAs $\Delta\theta = 7^\circ$ foram 50% e 23% para os Casos 1 e 2, respectivamente, a baixa performance do sistema de localização com o método de Campo Distante já era esperado. Apenas as posições próximas aos microfones e na região central da sala teriam maior probabilidade de apresentarem resultados aceitáveis.

Os resultados do Caso B, mostrados na Fig. 7.11(b), indicam taxas de acerto ainda mais baixas na localização, sendo que 10 das 15 posições apresentaram taxas de sucesso iguais a zero. Este desempenho do sistema com o método de Campo Distante no Caso B já era esperado, dada a baixa acurácia nas estimativas das DOAs com os conjuntos de microfones dos Casos 3 e 4.

A Fig. 7.12 apresenta a variação do valor médio da taxa de sucesso para erro máximo admitido nas estimativas das posições D_{MAX} entre 0 e 2.5 m.

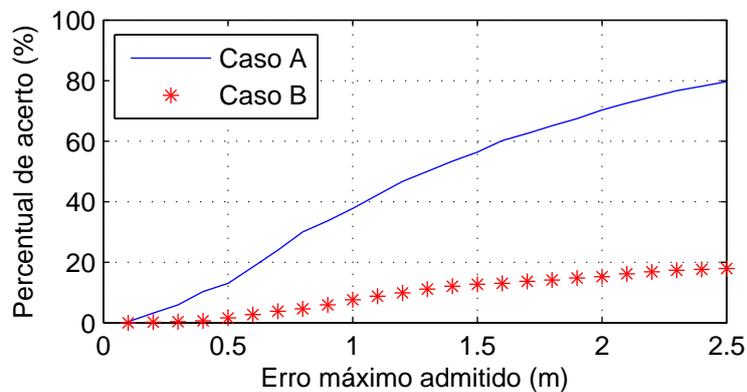


Figura 7.12: Taxa de sucesso média para a localização ao variar D_{MAX} para o ambiente real usando o método do Campo Distante.

Analisando os resultados, verifica-se que o Caso B não apresenta taxa de sucessos maior que 20 %, devido à baixa acurácia nas estimativas das DOAs em relação às exigidas na Fig. 5.20. O Caso A apresenta baixa performance, porém coerente com a baixa acurácia obtida nas estimativas das DOAs. O Caso A apresentou taxas de sucesso significativamente superiores às do Caso B, indicando sucesso na escolha da d_{PAR} ideal.

7.1.4 Considerações Finais sobre Localização com o Método do Campo Distante

A localização de duas fontes usando o método do campo distante no ambiente simulado com $T_{60} = 200$ ms apresentou taxa de sucesso média, para erro máximo de localização $D_{MAX} = 1.2$ m, igual a 82% no Caso A e 49% no Caso B, mostrando bom desempenho do procedimento proposto para a escolha da distância entre arranjos

d_{PAR} e para a determinação da região apropriada. Os resultados das Figs. 5.19 e 5.20 indicam o sucesso do Caso A e o fracasso do Caso B, conhecidas as acurácias das estimativas das DOAs.

No ambiente simulado com $T_{60} = 800$ ms e no ambiente real, as taxas de sucesso médias foram significativamente menores. Porém, o emprego da distância entre arranjos d_{PAR} otimizada, correspondente ao Caso A, resultou em melhores desempenhos experimentais em todos os ambientes.

O aumento da reverberação no ambiente simulado praticamente inviabilizou o uso do sistema de localização com o método do campo distante, uma vez que a taxa de sucesso média, ao admitir erro máximo de 1.2 m, não superou 20%, e foi de 40% ao admitir erro máximo de 2.5 m. Nesse caso, existe a necessidade de aprimorar o método de estimativa das DOAs, de forma a obter uma quantidade significativa de estimativas com erros de até 7° , conforme indicado na análise da região apropriada apresentada no Cap. 5.

O ambiente real utilizado nos experimentos apresenta alta reverberação³. A taxa de sucesso para $D_{MAX} = 1.2$ m no Caso A, com distância entre microfones otimizada, foi de 47%, indicando um baixo desempenho do sistema na maioria das localizações testadas. Nesse ambiente também é necessário melhorar a performance do método de estimativa das DOAs, de forma a obter uma quantidade significativamente maior de estimativas com erros de até 9° .

7.2 Método SCT

Nesta seção avaliaremos o desempenho do sistema de localização através de resultados experimentais em ambientes simulados e real com as DOAs estimadas pelo método SCT apresentado na Seção 4.2.2.

7.2.1 Ambiente Simulado com Baixa Reverberação

Apresentaremos em seguida os resultados experimentais no ambiente simulado com $T_{60} = 200$ ms, seguindo a mesma metodologia da Seção 7.1.1.

A Fig. 7.13 mostra a taxa de sucesso nas estimativas das DOAs com erro máximo admitido $\Delta\theta = 5^\circ$ para o Caso 1 e $\Delta\theta = 2^\circ$ para o Caso 2, para cada posição de uma das fontes. Esses valores de $\Delta\theta$ foram escolhidos por resultarem em erros de localização $D_{MAX} < 1.2$ m, de acordo com as análises das Figs. 5.13 e 5.14 na Seção 5.2.3.

Na Fig. 7.13(a) observa-se que o método SCT apresentou bons resultados no

³Na sala de aula utilizada, não há nenhum tratamento acústico, há uma grande vidraça e encontrava-se vazia durante as gravações.

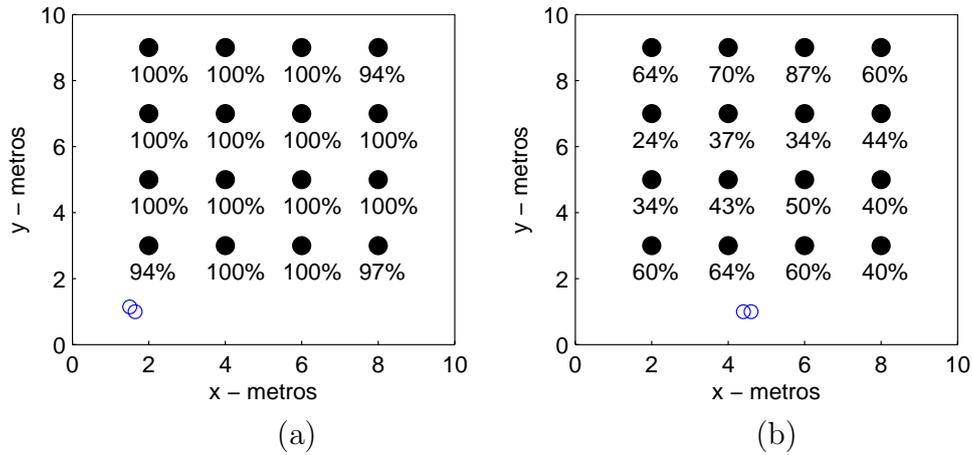


Figura 7.13: Taxa de sucesso nas estimativas das DOAs usando o método SCT em cada possível posição da fonte para $T_{60} = 200$ ms, considerando (a) Caso 1 com $\Delta\theta = 5^\circ$ e (b) Caso 2 com $\Delta\theta = 2^\circ$.

Caso 1 para todas as posições, sendo a taxa mínima de acerto igual a 94%, criando excelente expectativa para a utilização do sistema de localização na região apropriada da Fig. 5.13(b). Na Fig. 7.13(b) observa-se que o método SCT apresentou taxas de sucesso baixas no Caso 2 para $\Delta\theta = 2^\circ$, não sendo esperado, portanto, um bom desempenho do sistema de localização mesmo para as posições que se encontram dentro da região apropriada da Fig. 5.14(a).

A Fig. 7.14 apresenta a variação do valor médio da taxa de sucesso para erro máximo nas estimativas das DOAs entre 0.1° e 20° . Nessa figura observa-se que não há diferenças significativas entre os Casos 1 e 2, e que para $\Delta\theta = 5^\circ$ a taxa de sucesso é praticamente 100%.

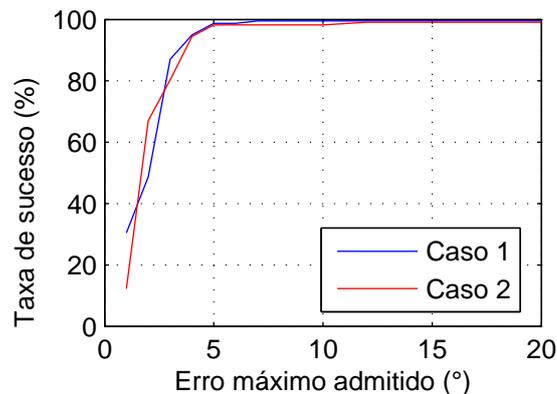


Figura 7.14: Taxa de sucesso média ao variar $\Delta\theta$, com $T_{60} = 200$ ms e utilizando o método SCT.

A Fig. 7.15 apresenta a taxa de sucesso da localização de uma das fontes com erro máximo admitido $D_{MAX} = 1.2$ m para cada possível posição. A Fig. 7.15(a) mostra

os percentuais para o Caso A, enquanto que a Fig. 7.15(b) mostra os percentuais correspondentes para o Caso B.

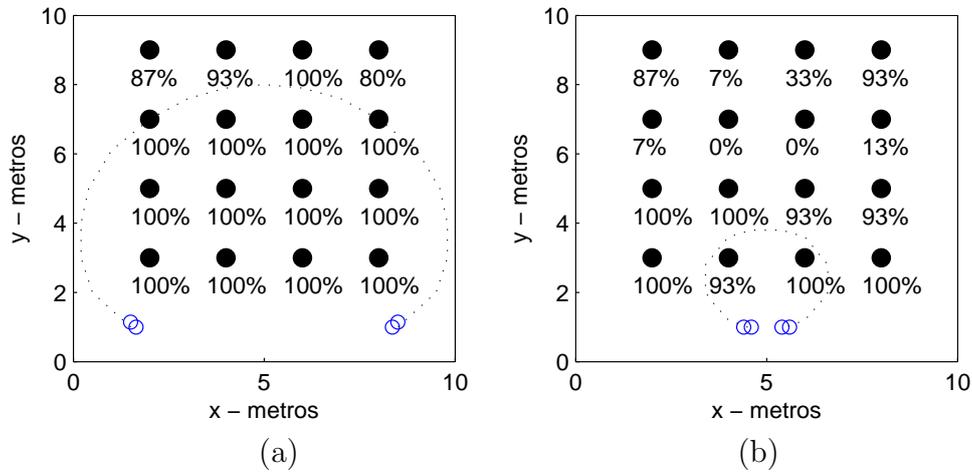


Figura 7.15: Taxa de sucesso da localização em cada possível posição da fonte, com $T_{60} = 200$ ms e usando o método SCT, considerando $D_{MAX} = 1.2$ m para (a) Caso A e (b) Caso (B).

Analisando-se os resultados do Caso A na Fig. 7.15(a), observa-se uma taxa de sucesso média igual a 97.5%, com taxa de sucesso igual a 100% em 12 das 16 posições avaliadas. As posições com $y = 9$ necessitam de $\Delta\theta = 3^\circ$ para serem bem localizadas, tendo apresentado taxas de sucesso nas estimativas das DOAs igual a 85%, que podem ser consideradas altas. Conseqüentemente, pode-se considerar que a localização na região apropriada da Fig. 5.13(a) será bem sucedida usando o método SCT.

Na Fig. 7.15(b) observamos que o Caso B apresenta taxa de sucesso média igual a 63.8%. As posições (4,3) e (6,3) estão dentro das regiões apropriadas da Fig. 5.14(a) e apresentaram altas taxas de sucesso. As posições (2,3), (8,3), (2,5), (4,5), (6,5), (8,5), (2,9) e (8,9) apresentaram taxas de acertos superiores a 87%, superando as expectativas. Nesses casos, a posição estimada a partir das DOAs com menor diferença possui grande peso no cálculo do D_E , e o procedimento proposto para obter a região apropriada se mostra conservador.

A Fig. 7.16 mostra o comportamento do valor médio da taxa de sucesso ao variar o erro máximo admitido de localização entre 0.1 m e 2.5 m para os Casos A e B.

Os resultados da Fig. 7.4 mostram que o Caso A possui desempenho significativamente superior ao Caso B, conforme esperado da análise da Seção 5.2.3. Para $D_{MAX} \geq 0.3$ m, o Caso A apresenta taxas de superiores às do Caso B, sendo que para $D_{MAX} = 0.5$ m, a taxa de sucesso do Caso A é igual a 80%, enquanto que a do Caso B é aproximadamente 38%. Para $D_{MAX} = 1.1$ m, a taxa de sucesso do Caso A é praticamente 100%, enquanto que o Caso B não supera 80% para nenhum valor

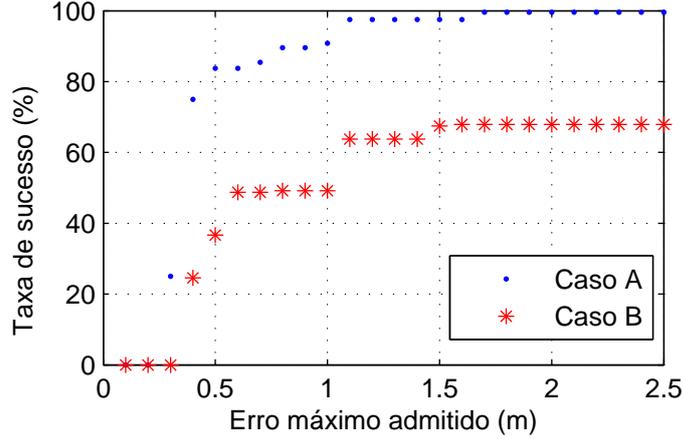


Figura 7.16: Taxa de sucesso média da localização ao variar D_{MAX} no ambiente simulado com $T_{60} = 200$ ms usando o método SCT.

de D_{MAX} no intervalo analisado. Então, é possível afirmar que a escolha da d_{PAR} otimizada permitiu melhorar o desempenho do sistema de localização.

7.2.2 Ambiente Simulado com Alta Reverberação

Nesta seção apresentaremos os resultados experimentais no ambiente simulado com $T_{60} = 800$ ms.

A Fig. 7.17 indica a taxa de sucesso nas estimativas das DOAs com erro máximo admitido $\Delta\theta = 7^\circ$ para os Casos 1 e 2, para cada posição da fonte. Estes valores foram escolhidos devido à baixa acurácia das estimativas das DOAs em ambientes com alta reverberação, sendo adequado considerar as regiões apropriadas das Figs. 5.13(b) e 5.14(c) da Seção 5.2.3.

Na Fig. 7.17(a) observa-se que o método SCT apresentou taxas de sucesso moderadas em todas as posições para o Caso 1, com média igual a 65%. Então, de acordo com a acurácia exigida na Fig. 5.13(b), não é esperada uma boa performance do sistema de localização.

A Fig. 7.17(b) mostra taxas de sucesso predominantemente baixas para o Caso B, com apenas 3 posições superando 50%, com média igual a 22%. Então, de acordo com a acurácia exigida na Fig. 5.14(b), também não é esperado um bom desempenho do sistema de localização.

A Fig. 7.18 apresenta a variação do valor médio da taxa de sucesso para erro máximo nas estimativas das DOAs entre 1° e 20° . Os resultados do Caso 1 apresentam performance levemente superior, e fica nítida a baixa taxa de sucesso para os erros menores que 5° , deixando as posições com $y = 7$ e $y = 9$ com perspectiva de baixo rendimento na localização, de acordo com as acurácias exigidas na Fig. 5.13. Os resultados do Caso 2 apresentam baixas taxas de sucesso com o erro máximo de

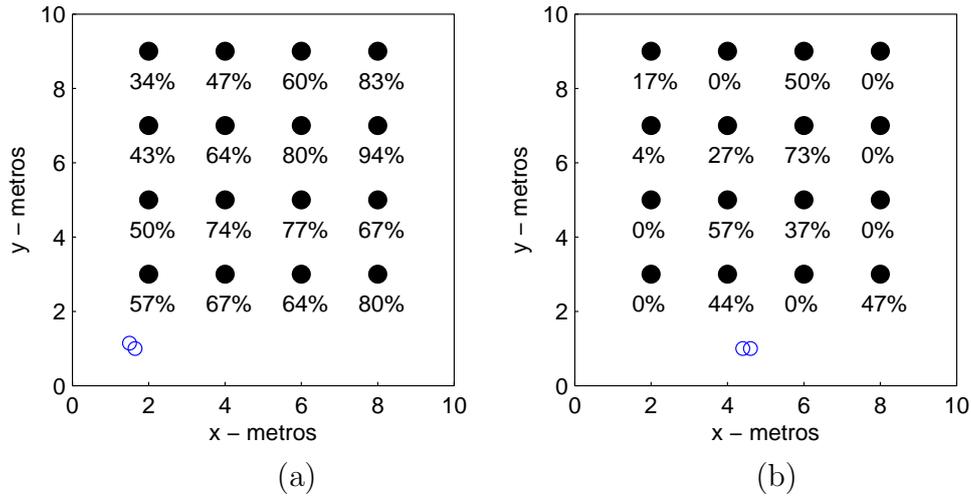


Figura 7.17: Taxa de sucesso nas estimativas das DOAs em cada posição, usando o método SCT, para $T_{60} = 800$ ms e $\Delta\theta = 7^\circ$, considerando (a) Caso 1 e (b) Caso 2.

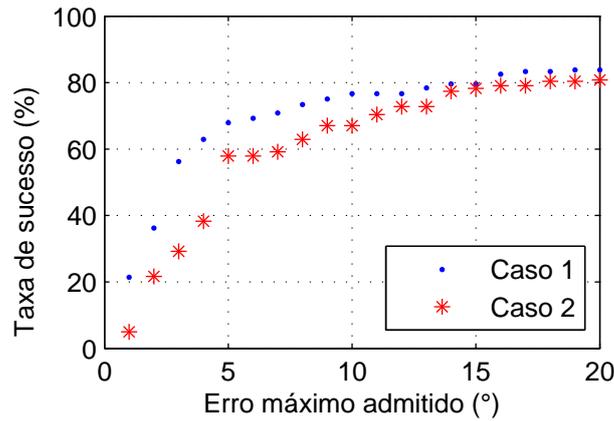


Figura 7.18: Taxa de sucesso média ao variar $\Delta\theta$ entre 1° e 20° , com $T_{60} = 800$ ms usando o método SCT.

2° , indicando a possibilidade de ocorrer baixa performance no sistema de localização.

A Fig. 7.19 apresenta a taxa de sucesso da localização de uma fonte com erro máximo admitido $D_{MAX} = 1.2$ m para cada possível posição da fonte. A Fig. 7.19(a) mostra os percentuais para o Caso A, enquanto que a Fig. 7.19(b) mostra os percentuais correspondentes para o Caso B.

Analisando os resultados do Caso A, as posições (2,3) e (8,3), próximas aos microfones, apresentaram taxas de acertos iguais a 53%, enquanto as posições (4,3) e (6,3) apresentaram taxas iguais a 67%, todas compatíveis com as acurácias das estimativas das DOAs exigidas da Fig. 5.13. As posições (2,5) e (8,5) apresentaram taxas de acertos iguais a 50%, compatíveis com as performances nas estimativas das DOAs e com a acurácia exigida na Fig. 5.13, enquanto que as posições (4,5) e (6,5) apresentaram taxas de 100%, superando as expectativas pelo fato de estarem numa

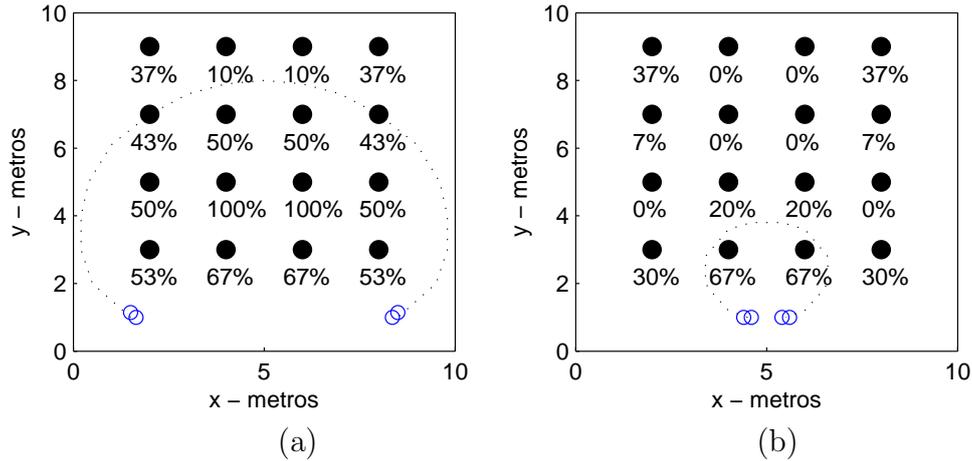


Figura 7.19: Taxa de sucesso para a localização em cada posição no ambiente simulado com T_{60} igual a 800 ms usando o método método SCT.

região onde o encontro das DOAs admite maiores erros nas estimativas. As posições (2,7) e (8,7), que estão fora da região apropriada para $\Delta\theta = 5^\circ$, apresentaram taxas de acerto iguais a 43%, enquanto que as posições (4,7) e (6,7), que estão no limite da região apropriada (mostrada em tracejado na Fig. 7.19), apresentaram taxas de 50%. As posições mais afastadas dos microfones, com $y = 9$, apresentaram baixas taxas de acerto, iguais a 37% ou 10%, e considerando que essas posições estão na região apropriada da Fig. 5.13(a), mas que a acurácia das estimativas das DOAs é baixa, correspondem a resultados compatíveis com a análise do Cap. 5.

Todas as posições do Caso B apresentaram baixa performance, como esperado, dadas as taxas de sucesso nas estimativas das DOAs para a acurácia exigida na Fig. 5.14.

A Fig. 7.20 apresenta o comportamento do valor médio da taxa de sucesso para uma variação do erro máximo admitido nas estimativas das posições D_{MAX} entre 0.1 m e 2.5 m nos Casos A e B.

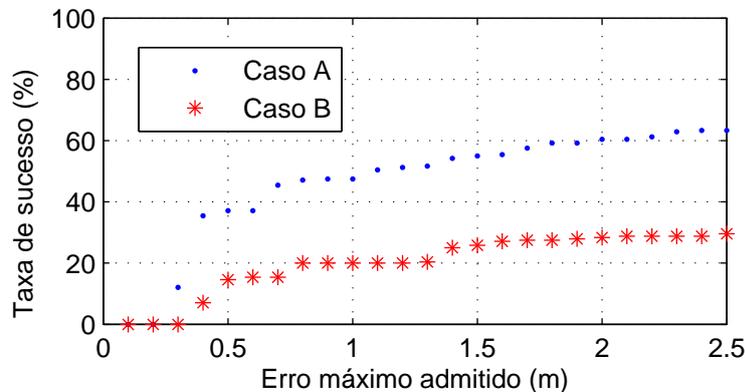


Figura 7.20: Taxa de sucesso média da localização ao variar D_{MAX} no ambiente simulado com $T_{60} = 800$ ms usando o método SCT.

O Caso A apresentou melhor desempenho em praticamente todo o intervalo de erro de localização analisado, como previsto pela análise da região apropriada ao considerar diferentes distâncias entre os arranjos de microfones. A taxa de acertos do Caso A é maior que 50% a partir de $D_{MAX} = 1.1$ m, chegando a 65% para $D_{MAX} = 2.5$ m. A taxa de acerto do Caso B para $D_{MAX} = 2.5$ m é igual a 30%, confirmando a análise do Cap. 5 que considerou este caso como inapropriado.

7.2.3 Ambiente Real

Nesta seção avaliaremos os resultados de localização de uma fonte no ambiente real da Fig. 5.16, seguindo a metodologia utilizada para os ambientes simulados.

A Fig. 7.21 apresenta as taxas de sucesso nas estimativas das DOAs em cada par de microfones (Casos 1 a 4) para todas as possíveis posições da fonte.

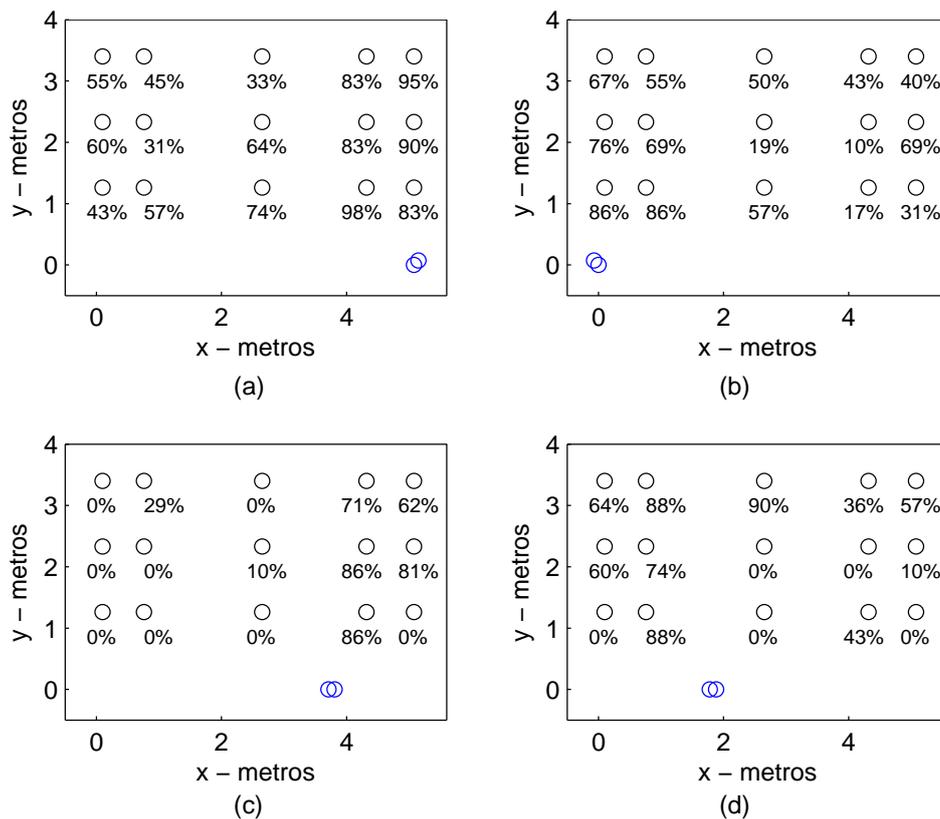


Figura 7.21: Taxa de sucesso nas estimativas das DOAs em cada possível posição da fonte no ambiente real, usando o método SCT e considerando $\Delta\theta = 7^\circ$ para (a) Caso 1 e (b) Caso 2, e $\Delta\theta = 2^\circ$ para (c) Caso 3 e (d) Caso 4.

Analisando os resultados do Caso 1, verificam-se taxas de sucesso inferiores a 50% nas posições (0.80,2.35), (2.60,3.40) e (0.8,3.4), sendo os respectivos valores iguais a 31%, 33% e 45%. Porém, considerando $\Delta\theta = 9^\circ$ com região apropriada

mostrada na Fig. 5.19(c), as respectivas taxas são 71%, 86% e 64%, estando apenas a posição (0.8,3.4) fora da região apropriada.

O Caso 2 apresenta taxas de sucesso inferiores a 50% nas posições (4.30,1.25), (5.10,1.25), (2.6,2.35), (4.3,2.35), (4.3,3.4) e (5.1,3.4), com respectivos percentuais de 17%, 31%, 19%, 10%, 43% e 40%. Avaliando as taxas de acertos para o erro máximo de 9°, os respectivos percentuais são 17%, 86%, 57%, 48%, 60% e 40%. Assim, os rendimentos das posições (4.30,1.25) e (4.3,2.35) permanecem baixos e as posições (4.3,3.4) e (5.1,3.4) estão fora da região apropriada para o erro de 9°, deixando estas posições sem perspectiva de apresentarem bons resultados na localização.

Os Casos 3 e 4 apresentaram baixas taxas de sucesso, sendo que o Caso 3 apresenta taxa de sucesso superior a 50% em apenas 5 das 15 posições e o Caso 4 em 6 das 15 posições. Assim, o sistema de localização com estes casos fica com perspectiva de performance muito ruim, dada a acurácia exigida na Fig. 5.20.

A Fig. 7.22 apresenta as taxas de sucesso médias variando-se o erro máximo admitido nas estimativas das DOAs entre 0° e 20°.

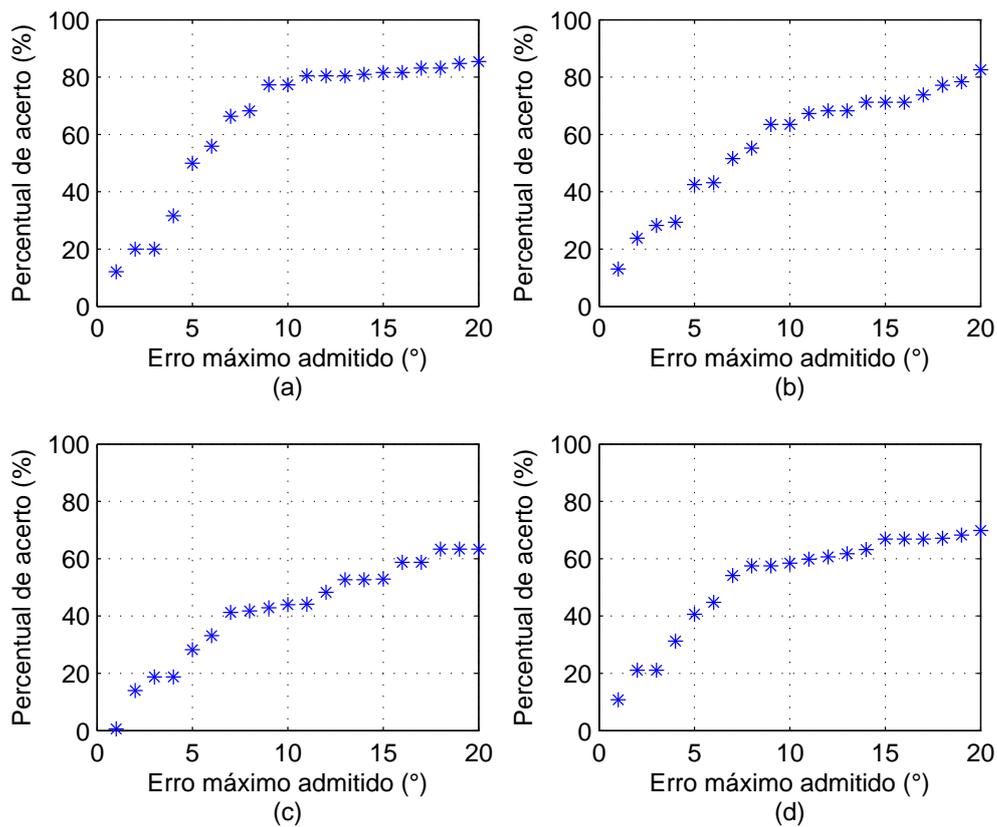


Figura 7.22: Taxa de sucesso média ao variar $\Delta\theta$, no ambiente real e usando o método SCT, para (a) Caso 1, (b) Caso 2, (c) Caso 3 e (d) Caso 4.

Analisando os resultados da Fig. 7.22(a), verifica-se que o Caso 1 é o que apresenta melhor taxa de sucesso, igual a 80% para o $\Delta\theta \geq 10^\circ$, enquanto que os Casos 2,

3 e 4 apresentam as respectivas taxas de 63%, 43% e 59%. As taxas de sucesso obtidas para os Casos 3 e 4 são incompatíveis com as condições de sucesso estabelecidas na análise da região apropriada da Fig. 5.20.

A Fig. 7.23 apresenta a taxa de sucesso da localização de uma das fontes com erro máximo admitido $D_{MAX} = 1.2$ m, para todas as possíveis posições. A Fig. 7.23(a) mostra os percentuais de estimativas das posições com erros menores que 1.2 m obtidas no Caso A, enquanto que a Fig. 7.23(b) mostra os percentuais correspondentes no Caso B.

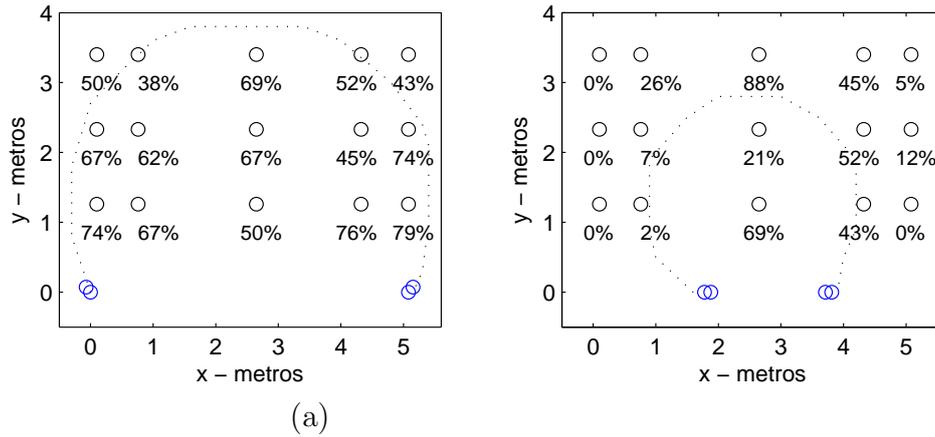


Figura 7.23: Taxa de sucesso da localização em cada possível posição da fonte no ambiente real, usando o algoritmo do SCT e considerando $D_{MAX} = 1.2$ m, para (a) Caso A e (b) Caso B.

Analisando-se os resultados do Caso A na Fig. 7.23(a), observam-se taxas de sucesso moderadas, com taxa de sucesso média igual a 61%, compatíveis com as acurácias das DOAs obtidas pelo método SCA e as exigidas para a região apropriada da Fig. 5.19, que apresentaram taxas de sucesso média para nos Casos 1 e 2 iguais a 66% e 52%, respectivamente. A d_{PAR} utilizada no Caso A viabilizou a melhor performance do sistema de localização em 13 das 15 posições.

Analisando os resultado do Caso B, verificam-se baixas taxas de sucesso, conforme previsto, dada a baixa acurácia nas estimativas das DOAs perante as exigidas pela Fig. 5.20.

A Fig. 7.24 apresenta a variação do valor médio da taxa de sucesso para erro máximo admitido nas estimativas das posições D_{MAX} entre 0 e 2.5 m.

Analisando os resultados da Fig. 7.24, verifica-se que o Caso A apresenta melhor performance. Para $D_{MAX} = 1$ m, sua taxa de sucesso é 60%, enquanto que a do Caso B é 20%. As taxas máximas de sucesso para os Casos A e B no intervalo analisado são 80% e 40%, respectivamente, indicando que a escolha do valor de d_{PAR} ótimo encontrado na Seção 5.3 faz com que a performance do sistema de localização melhore significativamente.

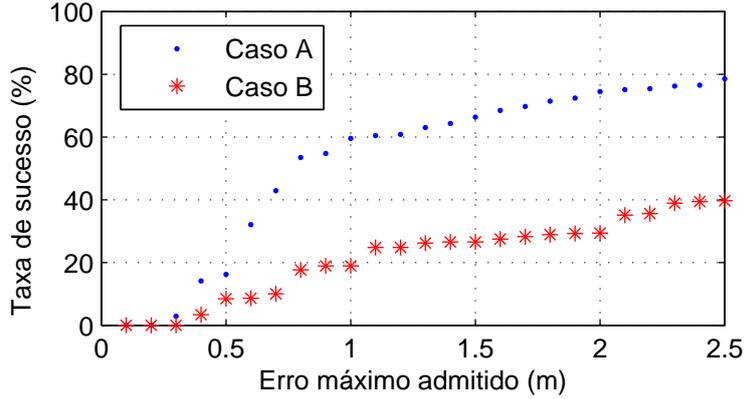


Figura 7.24: Taxa de sucesso média da localização ao variar D_{MAX} no ambiente real usando o método SCT.

7.2.4 Considerações Finais sobre a Localização com o Método SCT

A localização de duas fontes usando o método SCT no ambiente simulado com $T_{60} = 200$ ms apresentou taxa de acerto média, ao admitir o erro máximo de 1.2 m, igual a 99% no Caso A e 50% no Caso B, indicando que o método é adequado a ambientes com baixa reverberação. No ambiente simulado com $T_{60} = 800$ ms e no ambiente real, as taxas de acertos médias foram significativamente menores. Porém, a escolha da distância otimizada pelo procedimento proposto no Cap. 5, correspondente ao Caso A, resultou em melhor desempenho do sistema de localização em todos os ambientes. No ambiente real, com $T_{60} = 900$ ms, a taxa de acerto média ao admitir erro máximo de localização de 1.2 m foi igual a 61%.

As regiões apropriadas indicadas para cada ambiente no Cap. 5 foram validadas pelos testes experimentais. Com o método SCT e utilizando a distância otimizada entre os conjuntos de microfones, as taxas de acertos médias nas estimativas das DOAs e na localização no ambiente simulado com $T_{60} = 200$ ms foram, respectivamente, 99% e 97.5%; no ambiente simulado com $T_{60} = 800$ ms foram, respectivamente, 65% e 52%; e no ambiente real foram, respectivamente, 59% e 61%.

7.3 Método TRINICON

Nesta seção, avaliaremos o desempenho do sistema de localização através de resultados experimentais em ambientes simulados e real com as DOAs estimadas pelo algoritmo TRINICON apresentado na Seção 4.2.3.

7.3.1 Ambiente Simulado com Baixa Reverberação

Inicialmente apresentaremos os resultados experimentais no ambiente simulado com $T_{60} = 200$ ms. Seguiremos a metodologia utilizada nas seções anteriores, sendo avaliados inicialmente os resultados de estimação, pelo algoritmo TRINICOM, das DOAs nos pares de microfones nas duas posições consideradas (Casos 1 e 2) e em seguida a acurácia do sistema de localização empregando esse algoritmo para as duas combinações dos arranjos correspondentes aos Casos A e B.

A Fig. 7.25 apresenta a taxa de sucesso média nas estimativas das DOAs com erro máximo admitido $\Delta\theta = 5^\circ$ para o Caso 1 e $\Delta\theta = 2^\circ$ para o Caso 2, para uma das fontes na posição indicada pelo valor percentual. Assim como nas Seções 7.1.1 e 7.2.1, esses valores de $\Delta\theta$ foram escolhidos por resultarem em erros de localização $D_{MAX} < 1.2$ m, de acordo com as análises das Figs. 5.13 e 5.14 na Seção 5.2.3.

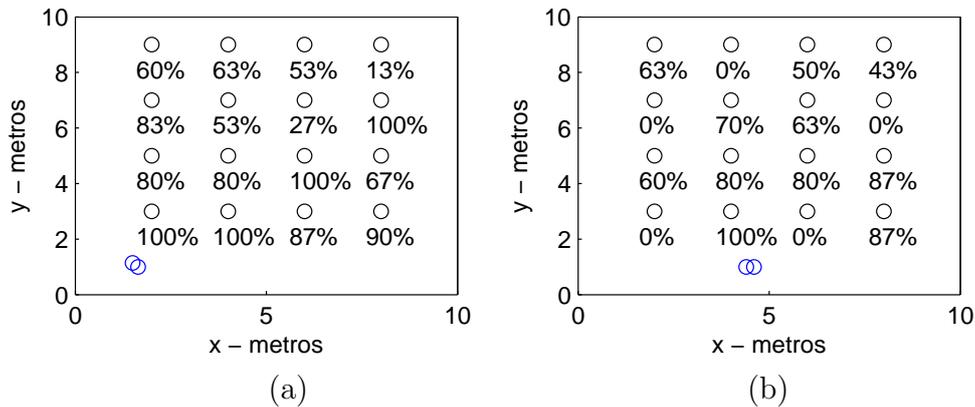


Figura 7.25: Taxa de sucesso nas estimativas das DOAs usando o método TRINICOM em cada possível posição da fonte para $T_{60} = 200$ ms, considerando (a) Caso 1 com $\Delta\theta = 5^\circ$ e (b) Caso 2 para $\Delta\theta = 2^\circ$.

Analisando os resultados do Caso 1, mostrados na Fig. 7.25(a), verifica-se que o método TRINICOM apresentou bons resultados, com valor médio da taxa de sucesso igual a 67%, gerando uma boa perspectiva de sucesso na etapa de localização, conforme mostram as propriedades na Seção 5.2.3. As posições com $y = 3$ e 5 estão dentro da região apropriada da Fig. 5.13(b) e apresentaram taxa de acerto mínima igual a 67%. As posições (2,7) e (8,7) estão na região apropriada da Fig. 5.13(a) e apresentaram elevadas taxas de sucesso, indicando boa localização. Já as posições (4,7), (6,7), e com $y = 9$ apresentaram baixas taxas de sucesso e, portanto, apesar de estarem na região apropriada da Fig. 5.13(a), não devem apresentar boas taxas de sucesso na localização.

O algoritmo TRINICOM no Caso 2, cujos os resultados estão mostrados na Fig. 7.25(b), apresentou baixas taxas de sucesso em todas as posições para a acurácia

exigida na obtenção da região apropriada da Fig. 5.13(a), gerando a expectativa de baixo rendimento para o sistema de localização.

A Fig. 7.26 apresenta a variação do valor médio da taxa de sucesso para erro máximo nas estimativas das DOAs entre 1° e 20° . Nessa figura, observa-se que não há diferenças significativas entre os resultados dos Casos 1 e 2, e que para o erro máximo nas estimativas das DOAs igual a 5° a taxa de sucesso é 80%.

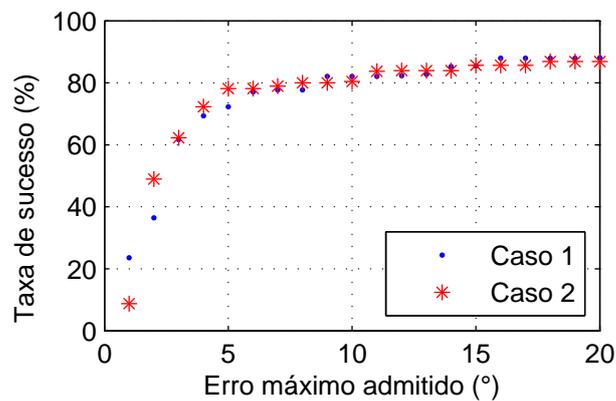


Figura 7.26: Taxa de sucesso média ao variar $\Delta\theta$, com $T_{60} = 200$ ms e utilizando o método TRINICON.

A Fig. 7.27 apresenta a taxa de sucesso da localização de uma das fontes com erro máximo admitido $D_{MAX} = 1.2$ m para cada possível posição. A Fig. 7.27(a) mostra os percentuais para o Caso A, enquanto que a Fig. 7.27(b) mostra os percentuais correspondentes para o Caso B.

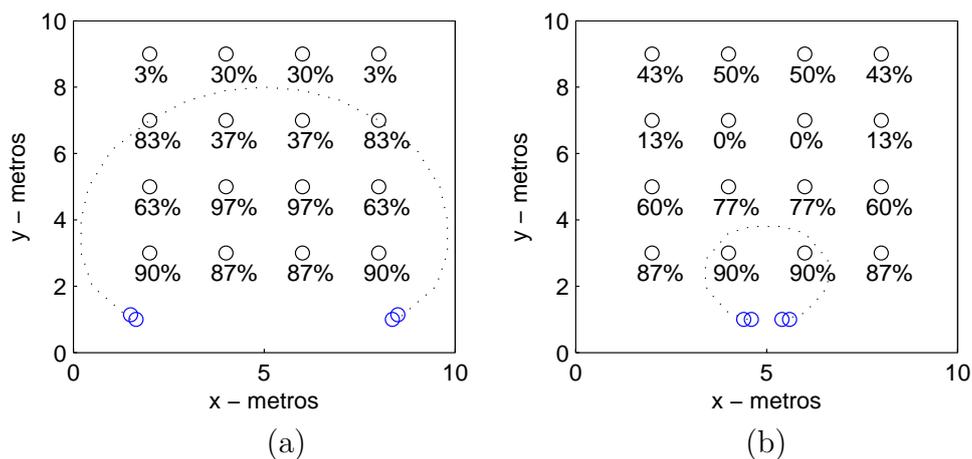


Figura 7.27: Taxa de sucesso da localização, com $T_{60} = 200$ ms e usando o algoritmo do TRINICON, considerando $D_{MAX} = 1.2$ m para (a) Caso A e (b) Caso (B).

Analisando os resultados do Caso A, mostrados na Fig. 7.27(a), as posições (2,7), (8,7) e com $y = 3$ e $y = 5$ apresentaram elevadas taxas de sucesso, conforme esperado devido ao bom desempenho do método TRINICON na estimação das suas DOAs.

As posições (4,7), (6,7) e com $y = 9$ apresentaram baixas taxas de sucesso, conforme esperado devido à baixa acurácia nas estimativas das DOAs correspondentes em relação às exigidas pela análise da Seção 5.2.3.

Analisando os resultados do Caso B, mostrados na Fig. 7.27(b), as posições com $y = 3$, (4,5) e (6,5) apresentaram altas taxas de sucesso, o que era esperado pelo fato dessas posições admitirem maiores erros e continuarem na região apropriada, conforme pode ser confirmado nas Figs. 5.14(b) e (c). As posições (2,5) e (8,5) estão no limite da região apropriada da Fig. 5.14(c) e apresentaram moderadas taxas de sucesso. As posições com $y = 7$ e $y = 9$ apresentaram baixas taxas de sucesso. Podemos concluir que os resultados do Caso B apresentaram resultados satisfatórios apenas para posições próximas aos arranjos de microfones.

A Fig. 7.28 mostra o comportamento do valor médio da taxa de sucesso supondo erro máximo admitido de localização entre 0.1 e 2.5 m, para os Casos A e B.

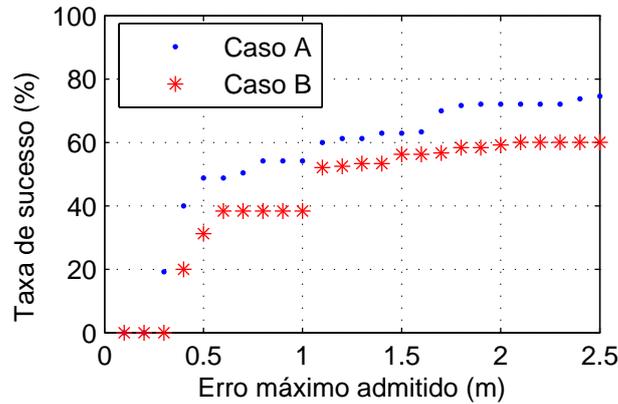


Figura 7.28: Taxa de acerto média da localização ao variar D_{MAX} no ambiente simulado com $T_{60} = 200$ ms usando o método TRINICON.

Os resultados da Fig. 7.28 mostram baixas taxas de acerto do sistema de localização com o algoritmo TRINICON, tanto para o Caso A quanto para o Caso B, cujas respectivas taxas de sucesso médias estão em torno de 60% para $D_{MAX} = 1.2$ m. Esse desempenho do sistema pode ser considerado ruim quando comparado ao do sistema com outros algoritmos de estimação de DOA em ambientes de baixa reverberação, conforme será discutido na Seção 7.4.

7.3.2 Ambiente Simulado com Alta Reverberação

Nesta seção apresentaremos os resultados experimentais no ambiente simulado com $T_{60} = 800$ ms empregando o algoritmo TRINICON para estimação das DOAs.

A Fig. 7.29 mostra as taxas de acertos nas estimativas das DOAs em cada posição ao admitir erro máximo de $\Delta\theta = 5^\circ$ para o Caso 1 e $\Delta\theta = 2^\circ$ para o Caso

2. Os valores de $\Delta\theta$ foram obtidos da análise apresentada na Seção 5.5 de modo a ter $D_{MAX} < 1.2$ m.

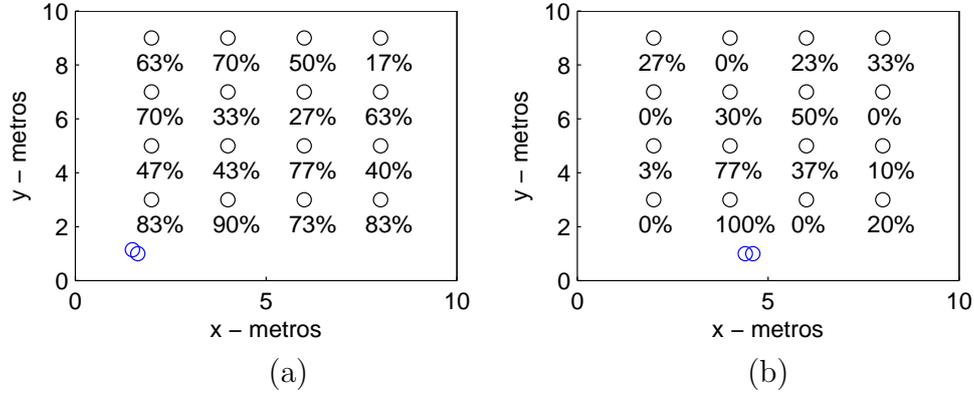


Figura 7.29: Taxa de sucesso nas estimativas das DOAs usando o método TRINICON em cada possível posição da fonte para $T_{60} = 800$ ms, considerando (a) Caso 1 com $\Delta\theta = 5^\circ$ e (b) Caso 2 para $\Delta\theta = 2^\circ$.

Analisando os resultados do Caso 1, mostrados na Fig. 7.29(a), verifica-se que nas posições mais próximas aos microfones, com $y = 3$, o método TRINICON apresentou bons resultados, com valor mínimo da taxa de sucesso igual a 73%, gerando uma boa perspectiva de sucesso na etapa de localização. As posições com $y = 5$ apresentaram taxas de acertos moderadas, com valor mínimo de 40%, criando perspectivas moderadas para a localização. As posições (2,7) e (8,7) apresentaram taxas de acertos iguais a 70% e 63%, consideravelmente maiores que as das posições (4,7) e (6,7), que foram de 33% e 27%. Esse resultado pode ser justificado pelo fato de haver, no centro da sala, um número maior de posições da segunda fonte próximas à da primeira, do que nas extremidades, dificultando a separação pelo algoritmo BSS na região central. As posições com $y = 9$, segundo a análise do Cap. 5, admitem erro máximo de 3° , sendo as taxas de acerto para esse erro ainda menores, o que gera a expectativa de baixo rendimento do sistema de localização.

Os resultados do Caso 2, mostrados na Fig. 7.29(b), indicam a baixa acurácia das estimativas das DOAs pelo método TRINICON, com a taxa média de acerto igual a 26%, criando expectativa de baixo rendimento na localização para este posicionamento do arranjo de microfones.

A Fig. 7.30 apresenta a variação do valor médio da taxa de sucesso para erro máximo nas estimativas das DOAs entre 1° e 20° para o algoritmo TRINICON.

Os resultados da Fig. 7.30 mostram que não há diferenças significativas entre os Casos 1 e 2. A maior diferença ocorre para valores de erro máximo admitido entre 3° e 4° , quando o Caso 1 apresenta rendimento em torno de 15% superior. A Fig. 7.31 apresenta a taxa de sucesso da localização de uma fonte com erro máximo

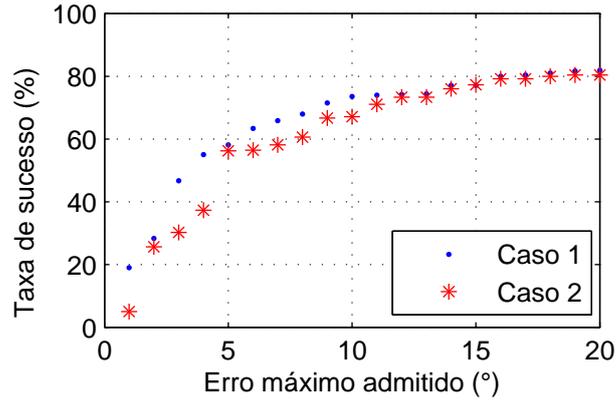


Figura 7.30: Taxa de acerto média ao variar $\Delta\theta$ entre 1° e 20° , com $T_{60} = 800$ ms usando o método TRINICON.

admitido $D_{MAX} = 1.2$ m para cada possível posição da fonte. A Fig. 7.31(a) mostra os percentuais para o Caso A, enquanto que a Fig. 7.31(b) mostra os percentuais correspondentes ao Caso B.

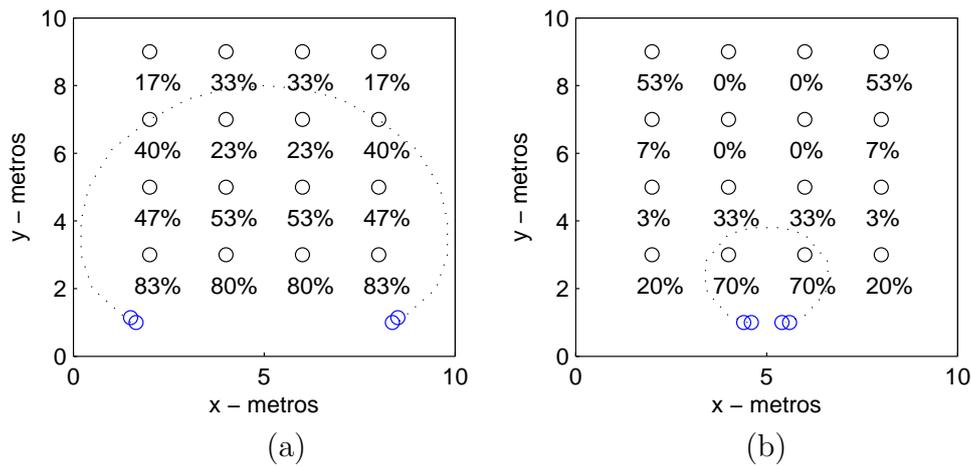


Figura 7.31: Taxa de sucesso da localização no ambiente simulado com $T_{60} = 800$ ms usando o método TRINICON, considerando $D_{MAX} = 1.2$ m para (a) Caso A e (b) Caso (B).

Analisando os resultados do Caso A, mostrados na Fig. 7.31(a), as posições com $y = 3$ apresentaram boas taxas de sucesso, com valor mínimo igual a 80%. As posições com $y = 5$ apresentaram taxas em torno de 50%, enquanto as posições com $y = 7$ e 9 apresentaram taxas iguais ou menores que 40%, indicando bom desempenho do sistema de localização apenas para posições próximas aos microfones quando o algoritmo TRINICON é utilizado.

Os resultados do Caso B, mostrados na Fig. 5.15(b), apresentaram taxa média de sucesso igual a 26%, indicando um desempenho fraco do sistema em todo o ambiente, em conformidade com a análise das estimativas das DOAs e da região apropriada.

A Fig. 7.32 mostra o comportamento do valor médio da taxa de sucesso ao variar o erro máximo admitido de localização entre 0.1 m e 2.5 m. Os resultados da Fig.

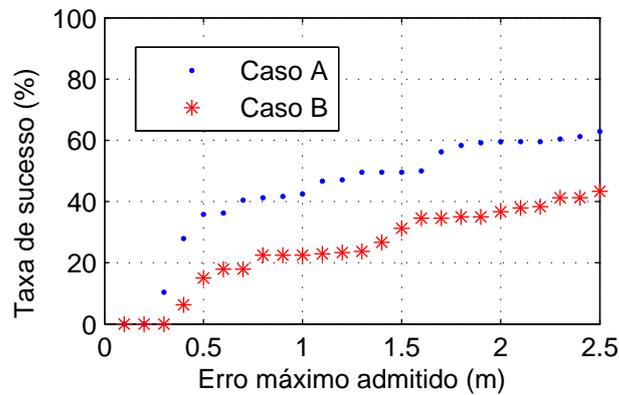


Figura 7.32: Taxa de acerto média da localização ao variar D_{MAX} no ambiente simulado com $T_{60} = 800$ ms usando o método TRINICON.

7.28 mostram que o Caso A apresenta performance predominantemente superior à do Caso B, apresentando taxa de sucesso da localização aproximadamente 20% acima da do Caso B para erro máximo admitido maior que 0.4 m. As taxas de sucesso não atingem valores altos mesmo no Caso A, sendo de 63% no Caso A e de 43% no Caso B para erro máximo de localização igual 2.5 m.

7.3.3 Ambiente Real

Nesta seção avaliaremos os resultados de localização de duas fontes ativas simultaneamente no ambiente real quando o algoritmo TRINICON é empregado para estimar as DOAs. Devido à falta de simetria da sala, são apresentados os resultados das quatro posições consideradas para o arranjo de microfones da Fig. 5.16.

A Fig. 7.33 apresenta as taxas de sucesso nas estimativas das DOAs em cada par de microfones para todas as possíveis posições da fonte. O erro máximo admitido foi de 7° para os Casos 1 e 2, conforme exigido da análise da região apropriada na Seção 5.7, e de 5° para os Casos 3 e 4, acima do exigido tendo em vista a baixa performance do algoritmo TRINICON nesse cenário.

Analisando os resultados do Caso 1, mostrados na Fig. 7.33(a), verifica-se uma grande variação nas taxas de sucesso, com valor médio igual a 50%. As taxas foram maiores nas posições mais próximas aos microfones. Os resultados do Caso 2, mostrados na Fig. 7.33(b), apresentaram taxas de sucesso iguais a 0% em 8 das 15 posições avaliadas, gerando expectativa de desempenho ruim na etapa de localização quando utilizado esse par de microfones.

Analisando os resultados do Caso 3, mostrados na Fig. 7.33(c), verificam-se baixas taxas de sucesso, sendo o valor médio igual a 4%. O Caso 4, com resultados

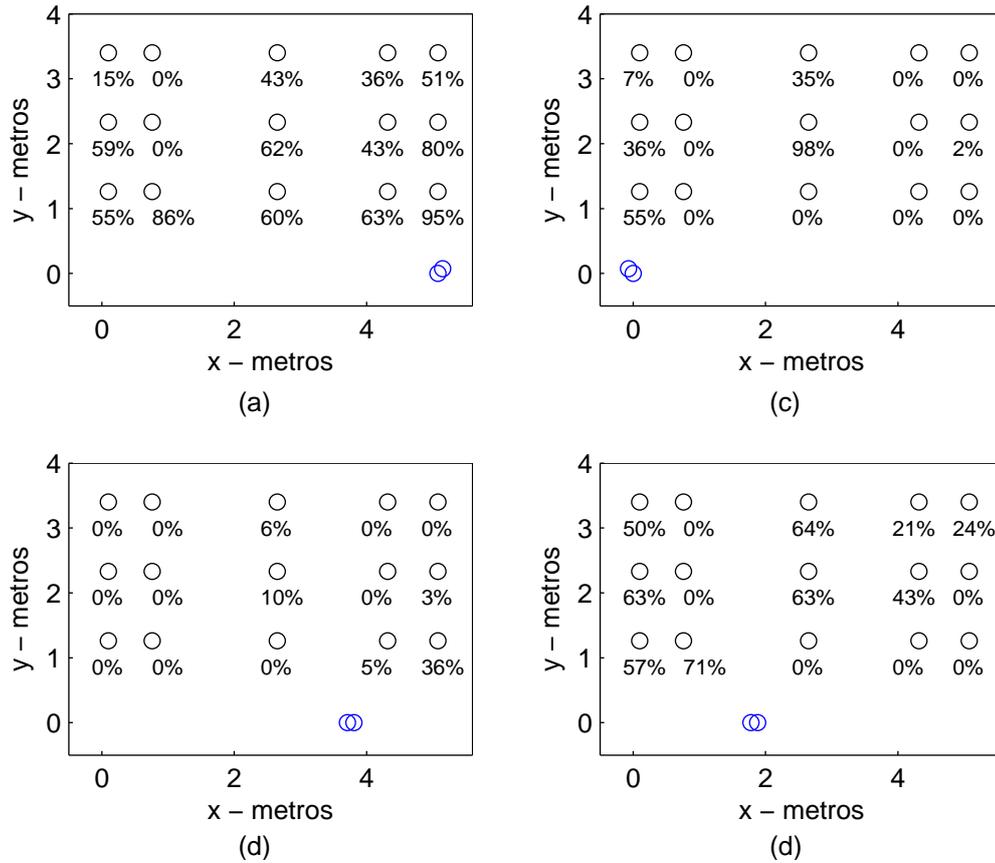


Figura 7.33: Taxa de sucesso nas estimativas das DOAs em cada possível posição da fonte, usando o método TRINICON no ambiente real, considerando $\Delta\theta = 7^\circ$ para (a) Caso 1 e (b) Caso 2, e $\Delta\theta = 5^\circ$ para (c) Caso 3 e (d) Caso 4.

mostrados na Fig. 7.33(d), também apresenta baixas taxas de sucesso, com média igual a 30%. Portanto, espera-se desempenho ruim do sistema de localização com os arranjos de microfones nas posições dos Casos 3 e 4. A Fig. 7.34 apresenta as taxas de sucesso médias variando-se o erro máximo admitido nas estimativas das DOAs entre 0° e 20° .

Analisando os resultados da Fig. 7.34, verifica-se que o Caso 1 apresenta as melhores taxas de sucesso médias, sendo que a diferença nos resultados em relação ao Caso 2 deve-se à falta de simetria da sala. Os Casos 2, 3 e 4 apresentaram baixas taxas de sucesso, sendo inferiores a 40% para $\Delta\theta = 10^\circ$, correspondendo a desempenhos muito ruins em relação ao exigido nas Figs. 5.19 e 5.20 e indicando baixa performance na etapa de localização.

A Fig. 7.35 apresenta a taxa de sucesso da localização de uma das fontes com erro máximo admitido $D_{MAX} = 1.2$ m para cada possível posição, sendo a posição da segunda fonte variada. A Fig. 7.35(a) mostra os percentuais de estimativas das posições com erros menores que 1.2 m obtidas no Caso A, enquanto que a Fig.

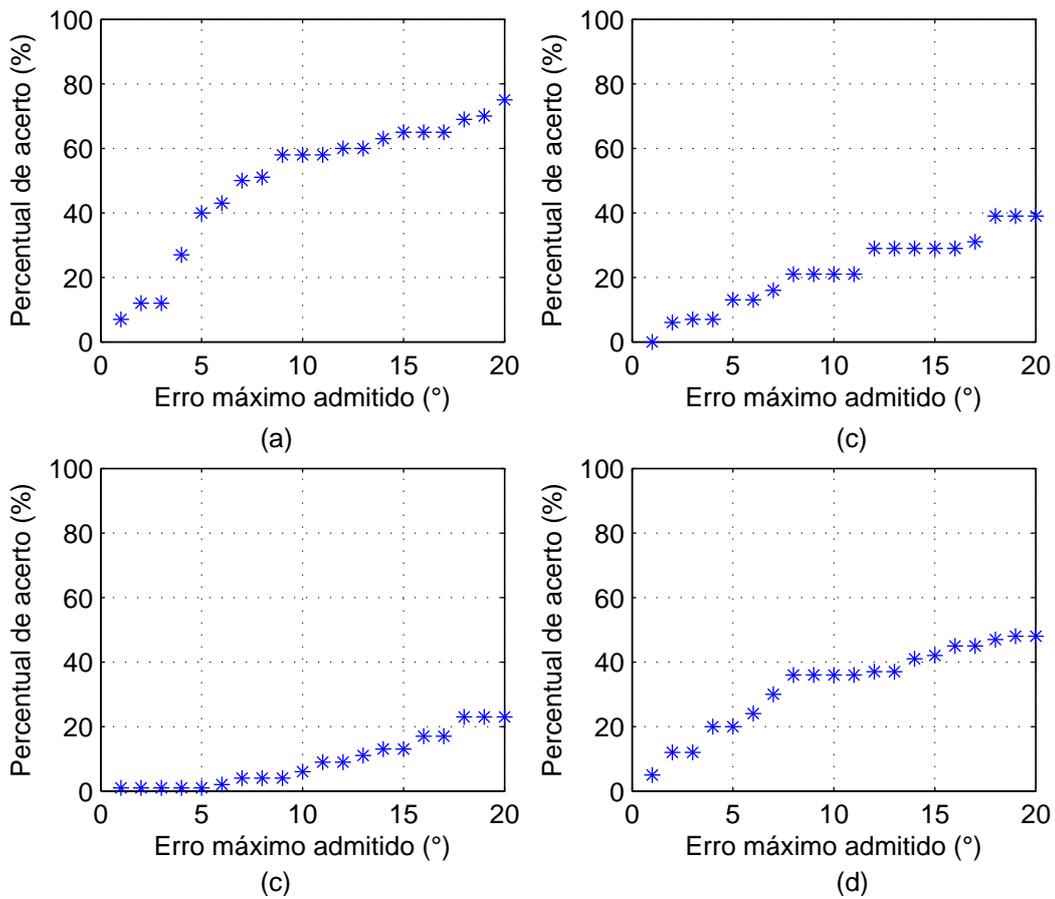


Figura 7.34: Taxa de sucesso ao variar $\Delta\theta$ para (a) Caso 1, (b) Caso 2, (c) Caso 3 e (d) Caso 4.

7.35(b) mostra os percentuais correspondentes para o Caso B.

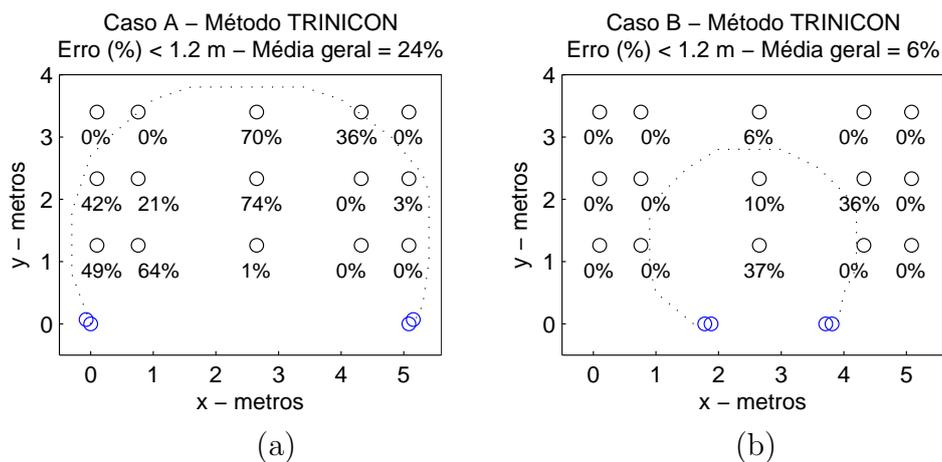


Figura 7.35: Taxa de sucesso da localização em cada possível posição da fonte, usando o método TRINICON no ambiente real, considerando $D_{MAX} = 1.2$ m para (a) Caso A e (b) Caso B.

Analisando-se os resultados da Fig. 7.35(a), o Caso A apresentou baixas taxas

de sucesso, sendo seu valor médio igual a 24%. Esse resultado está compatível com o baixo rendimento nas estimativas das DOAs do Caso 2 em relação ao exigido na Fig. 5.19. O Caso B, com resultados mostrados na Fig. 7.35(b), apresentou baixíssimas taxas de sucesso, sendo seu valor médio igual a 6%, o que já era esperado pela baixa performance do algoritmo TRINICON para estimar as DOAs dos Casos 3 e 4 perante ao exigido na Fig. 5.20.

A Fig. 7.36 apresenta a variação do valor médio da taxa de sucesso para erro máximo admitido nas estimativas das posições D_{MAX} entre 0.1 m e 2.5 m. Os

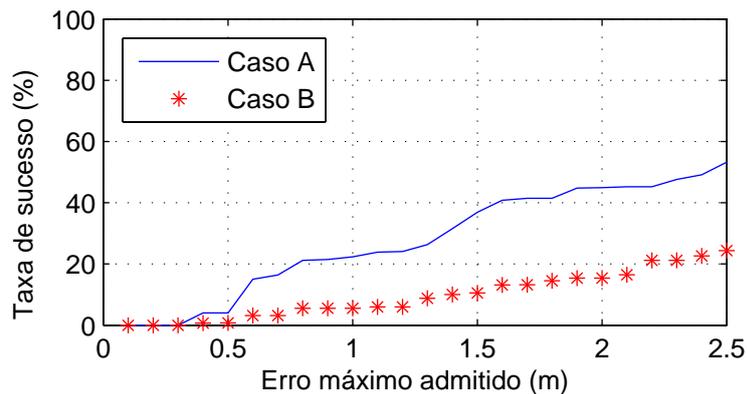


Figura 7.36: Taxa de acerto média da localização ao variar D_{MAX} no ambiente real usando o método TRINICON.

resultados da Fig. 7.36 confirmam o baixo rendimento do algoritmo TRINICON nas condições testadas, uma vez que para $D_{MAX} = 2.5$ m as taxas de acerto do Caso A não superam 55% e as do Caso B não superam 30%.

7.3.4 Considerações Finais sobre a Localização com o Método TRINICON

A localização de duas fontes usando o método TRINICON no ambiente simulado com $T_{60} = 200$ ms apresentou taxa de sucesso média, para o erro máximo de localização $D_{MAX} = 1.2$ m, igual a 62% no Caso A e 52% no Caso B, indicando um desempenho pior quando comparado aos dos outros métodos em ambiente de baixa reverberação.

No ambiente simulado com $T_{60} = 800$ ms, a diminuição da performance do sistema ocorreu nos dois casos, sendo menos expressiva no Caso A. Porém, mesmo nesse caso, o sistema não apresentou performance satisfatória com o algoritmo TRINICON. No ambiente real, as taxas de sucesso foram em geral baixas, com valores médios para $D_{MAX} = 1.2$ m iguais a 24% no Caso A e 6% no Caso B. Assim, podemos considerar o método inapropriado para as condições testadas.

O procedimento proposto para a escolha da distância entre arranjos d_{PAR} e as regiões apropriadas indicadas no Cap. 5 mostrou bom desempenho para os diferentes

ambientes testados.

7.4 Comparações entre Métodos de Localização de Duas Fontes

Nesta seção, compararemos os resultados obtidos com os diferentes métodos empregados para estimar as DOAs na presença de duas fontes, para os ambientes simulados e real apresentados nas seções anteriores. Apresentaremos também uma análise da influência da reverberação na localização, variando o valor do T_{60} do ambiente simulado entre 200 ms e 900 ms, com passo de 100 ms.

7.4.1 Ambientes Simulados

Inicialmente, compararemos as acurácias das estimativas das DOAs e em seguida as das estimativas das posições obtidas com os métodos de Campo Distante (CD), SCT e TRINICON, através da avaliação das suas taxas de sucesso médias nos ambientes simulados. Os resultados experimentais utilizados nas comparações serão os do Caso 1 para os erros nas estimativas das DOAs e os dos Casos A nas estimativas das posições da fonte, uma vez que esses foram os casos que apresentaram os melhores rendimentos para todos os métodos.

A Fig. 7.37 apresenta a variação do valor médio da taxa de sucesso para erro máximo nas estimativas das DOAs entre 1° e 20° , em ambos os ambientes simulados e com todos os métodos investigados. Foram também incluídos os resultados obtidos com o método GCC-PHAT quando apenas uma fonte estava presente no ambiente. Portanto, esse resultado serve apenas para avaliar a redução de desempenho causada pela presença da segunda fonte, não sendo o algoritmo GCC-PHAT apropriado para a estimação das DOAs no caso de múltiplas fontes.

Analisando os resultados da Fig. 7.37, verificam-se melhores desempenhos de todos os algoritmos no ambiente com menor reverberação, com $T_{60} = 200$ ms. O método GCC-PHAT, não sujeito às dificuldades impostas pela separação, apresentou taxas de sucesso de 100% para $\Delta\theta > 2^\circ$.

Dentre os métodos válidos para múltiplas fontes, o método SCT apresentou os melhores resultados nos dois ambientes (com baixa e com alta reverberações). Suas taxas de sucesso médias para $\Delta\theta > 5^\circ$ foram de 98% no ambiente com $T_{60} = 200$ ms e de 68% no ambiente com $T_{60} = 800$ ms. O método do Campo Distante apresentou bons resultados para $T_{60} = 200$ ms, obtendo taxa de sucesso média igual a 84% para $\Delta\theta = 5^\circ$. No entanto, a sua performance caiu significativamente no ambiente com maior reverberação, sendo a taxa de sucesso média igual a 22% para $\Delta\theta = 5^\circ$ e $T_{60} = 800$ ms.

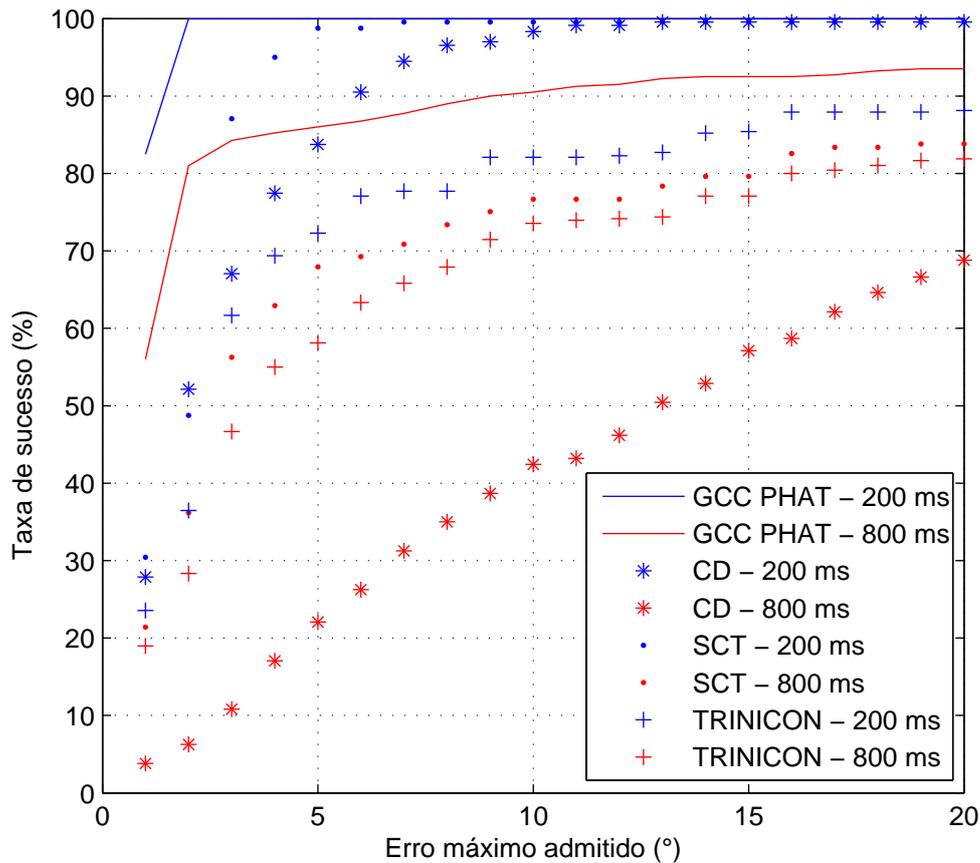


Figura 7.37: Taxa de sucesso média ao variar $\Delta\theta$ obtidas com os diversos métodos de estimação das DOAs para o Caso 1 nos ambientes simulados.

O método TRINICON apresentou o menor rendimento no ambiente com menor reverberação. Porém, sua performance não foi muito afetada com o aumento do tempo de reverberação. Considerando erro máximo de 5° , sua taxa de sucesso foi igual a 73% para $T_{60} = 200$ ms, e 58% para $T_{60} = 800$ ms. A Fig. 7.38 apresenta a variação do valor médio da taxa de sucesso para erro máximo de localização entre 0.1 m e 2.5 m, em ambos os ambientes simulados e com todos os métodos investigados, para o Caso A.

Analisando os resultados da Fig. 7.38, verificam-se resultados compatíveis com os da Fig. 7.37, e as melhores performances de todos os métodos foram obtidas no ambiente com $T_{60} = 200$ ms. O método SCT apresentou a melhor performance dentre os métodos válidos para múltiplas fontes. O método SCT foi o mais afetado pela reverberação, enquanto que o método TRINICON obteve os piores resultados para o ambiente com menor tempo de reverberação. No ambiente com maior reverberação, os algoritmos SCT e TRINICON apresentaram desempenhos semelhantes.

Portanto, a partir dos resultados experimentais nos ambientes simulados, o método SCT pode ser considerado o melhor método de estimação das DOAs para

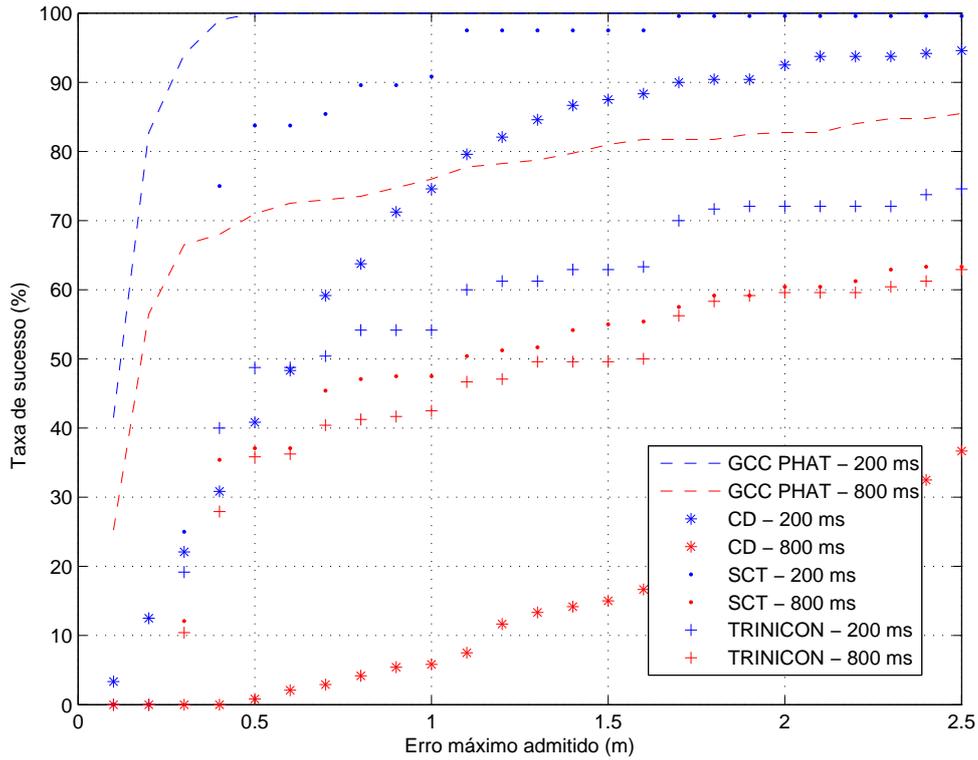


Figura 7.38: Taxa de sucesso média ao variar D_{MAX} obtidas com os diversos métodos de estimação das DOAs para o Caso A nos ambientes simulados.

duas fontes ativas dentre os avaliados, apresentando excelente desempenho no ambiente de baixa reverberação.

Influência da Reverberação na Localização

Nesta seção, analisaremos os resultados experimentais obtidos com o método SCT para valores intermediários de T_{60} , de forma a avaliar o efeito desse parâmetro no desempenho do sistema de localização e considerar uma variedade maior de ambientes. Apenas o método SCT será considerado nessa seção devido à sua melhor performance.

A Fig. 7.39 apresenta os valores médios da taxa de sucesso para erro máximo de localização entre 0.1 m e 2.5 m, para o Caso A, com T_{60} entre 200 ms e 900 ms. Analisando os resultados, verifica-se que para $D_{MAX} = 0.5$ m o sistema de localização apresenta taxas de sucesso superiores a 70% para $T_{60} \leq 300$ ms, indicando bom desempenho do sistema nesses casos. Para $D_{MAX} = 1.2$ m, obtêm-se taxas de sucesso superiores a 65% com $T_{60} \leq 500$ ms.

A Fig. 7.40 apresenta a taxa de sucesso da localização de uma das fontes com erro máximo admitido $D_{MAX} = 1.2$ m para cada possível posição para $T_{60} = 300$

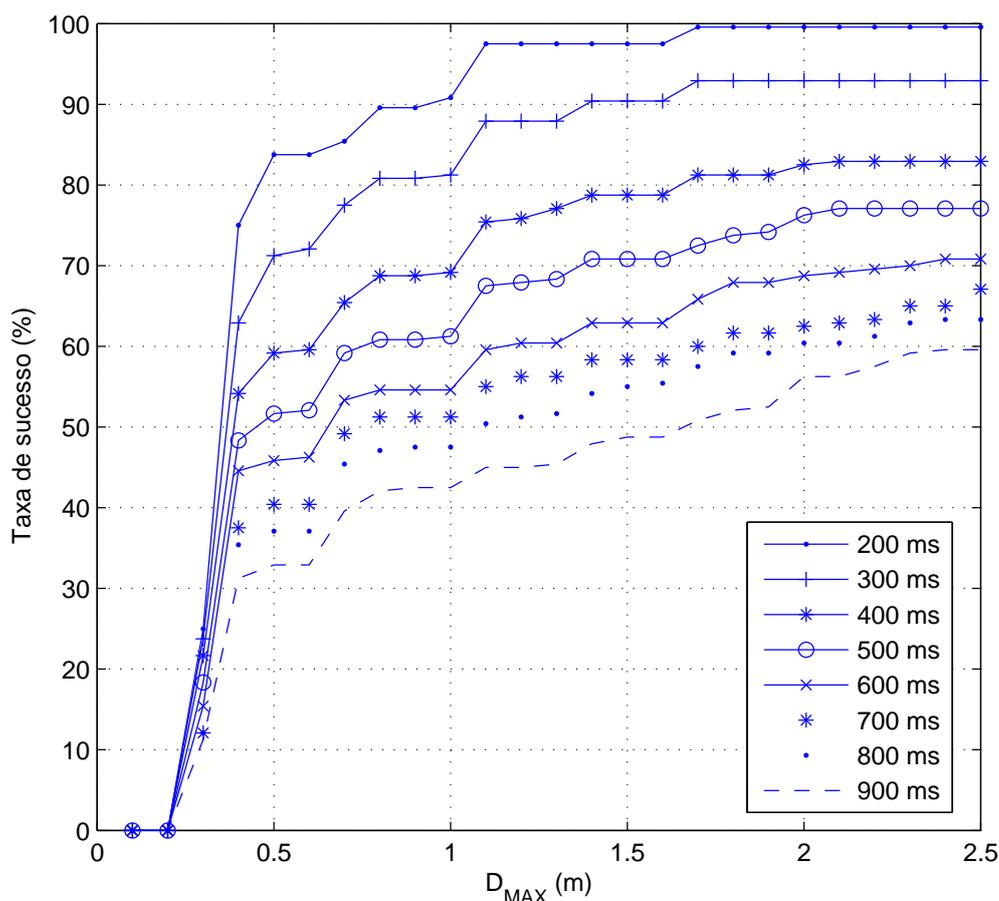


Figura 7.39: Taxa de sucesso média da localização ao variar D_{MAX} no ambiente simulado para T_{60} entre 200 ms e 900 ms, usando o método SCT.

ms, 400 ms, 500 ms e 600 ms.

Analisando os resultados da Fig. 7.40, verifica-se que o aumento do T_{60} provoca perda de performance gradual. Para $T_{60} = 300$ ms até mesmo as posições mais afastadas dos microfones, com $y = 9$, apresentam boas taxas de sucesso, com valor médio igual a 79%. Para $T_{60} = 400$ ms, as taxas de sucesso são moderadas para as posições com $y = 9$ e boas para as posições com $y = 7$, com respectivos valores médios iguais a 58% e 70%; para $T_{60} = 500$ ms e 600 ms as posições com $y = 9$ apresentam baixas taxas de sucesso e as posições com $y = 7$ taxas de sucesso moderadas.

Assim, podemos concluir que o sistema de localização apresenta bom desempenho, considerando erro máximo nas estimativas das posições igual a 1,2 m, no ambiente simulado nas posições com $y = 9$ para $T_{60} \leq 300$ ms e nas posições com $y = 7$ para $T_{60} \leq 500$ ms.

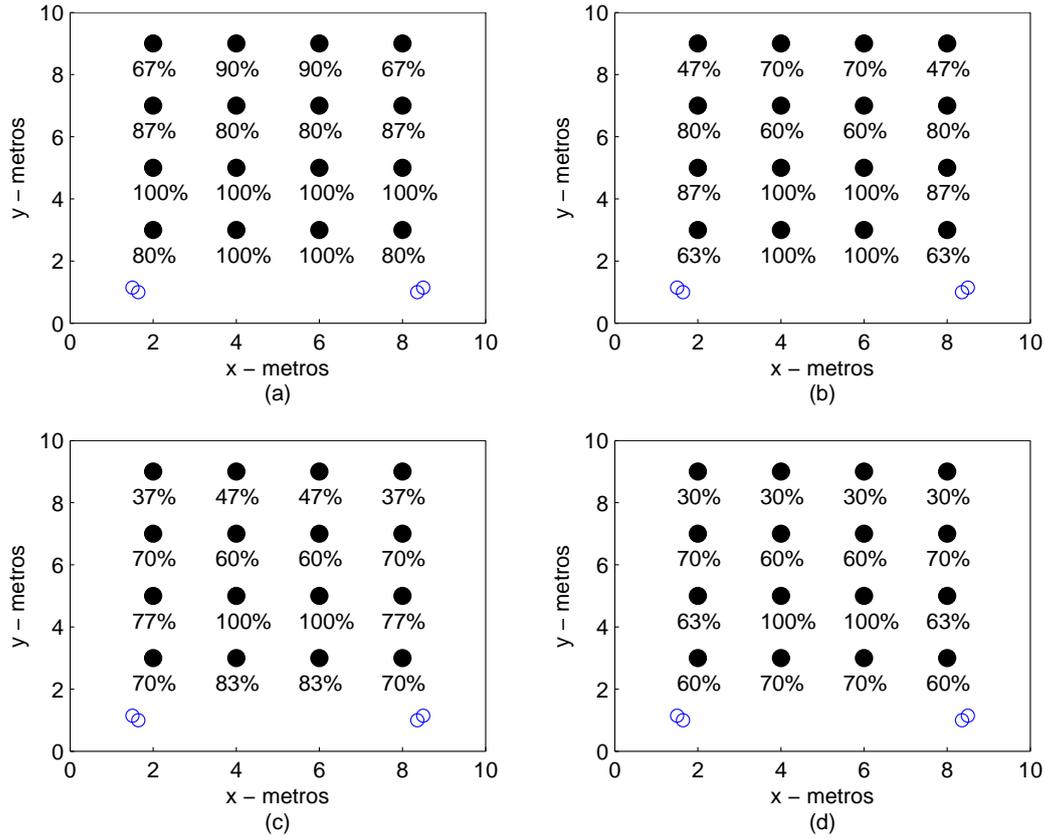


Figura 7.40: Taxa de sucesso da localização em cada possível posição da fonte usando o método SCT e considerando $D_{MAX} = 1.2$ m para (a) $T_{60} = 300$ ms, (b) $T_{60} = 400$ ms, (c) $T_{60} = 500$ ms e (d) $T_{60} = 600$ ms.

7.4.2 Ambiente Real

Nesta seção compararemos as taxas de acertos médias das estimativas das DOAs dos Casos 1 e 2, e as taxas de acertos médias das estimativas das posições do Caso A, no ambiente real, para todos os métodos investigados. As estimativas das DOAs dos Casos 3 e 4 e as estimativas das posições do Caso B não serão avaliadas devido aos seus baixos rendimentos em todos os testes.

A Fig. 7.41 apresenta a variação do valor médio da taxa de sucesso para erro máximo nas estimativas das DOAs entre 1° e 20° , no ambiente real e com todos os métodos investigados, para os Casos 1 e 2. Analisando os resultados da Fig. 7.41, observa-se que os rendimentos de todos os métodos são melhores no Caso 1 do que no Caso 2. Esse rendimento diferenciado se justifica pela utilização de diferentes microfones e pela assimetria da sala.

O método GCC-PHAT, por ter sido aplicado ao sinal de apenas uma fonte e não estar sujeito aos erros da etapa de separação, apresentou os melhores resultados, com melhor performance no Caso 1 para $\Delta\theta \leq 10^\circ$. Os métodos do Campo Distante e TRINICON apresentaram resultados semelhantes para $\Delta\theta \leq 10^\circ$ no Caso 1 e

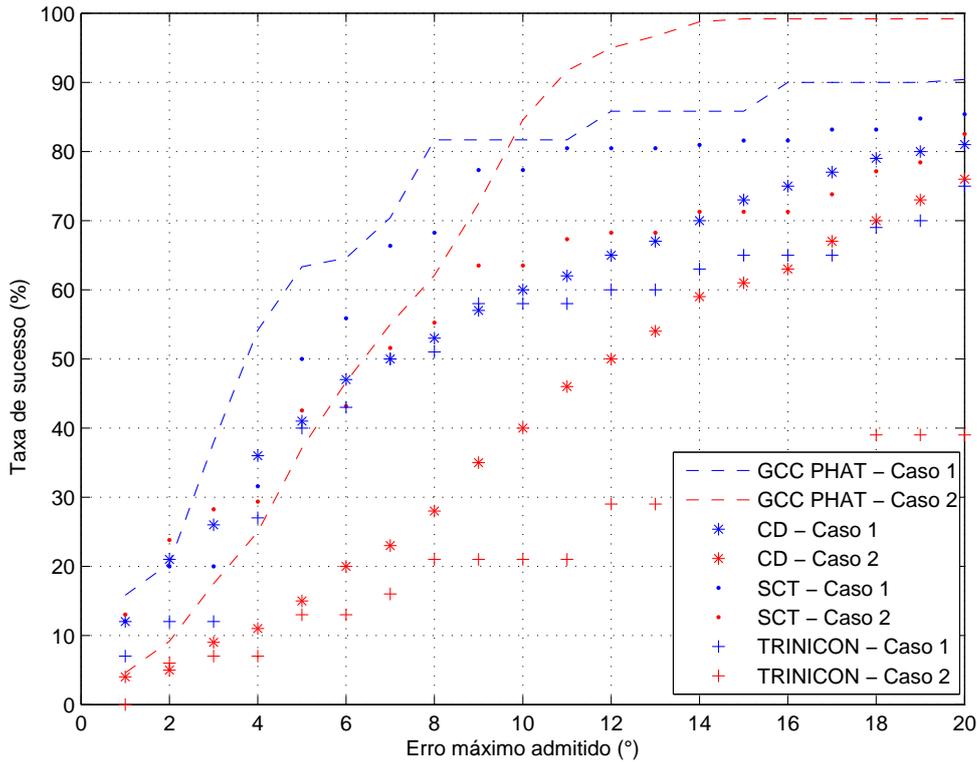


Figura 7.41: Taxa de sucesso média ao variar $\Delta\theta$ obtidas com os diversos métodos de estimação das DOAs para os Casos 1 (em azul) e 2 (em vermelho) no ambiente real.

$\Delta\theta \leq 5^\circ$ no Caso 2. Para valores maiores de $\Delta\theta$ o método do Campo Distante apresentou melhor desempenho.

O método SCT apresentou o melhor desempenho dentre todos os algoritmos nos Casos 1 e 2, próximo do desempenho do algoritmo GCC-PHAT com apenas uma fonte ativa para o Caso 1. Considerando a análise da região apropriada apresentada na Seção 5.7, é possível indicar erro máximo admitido de até 7° como um valor satisfatório para boa localização nestas condições. Assim, o método SCT se mostra a melhor opção para ser utilizado no sistema de localização, uma vez que para o Caso 1 ele apresenta taxa de sucesso igual a 68% para $\Delta\theta = 7^\circ$.

A Fig. 7.42 mostra a variação da taxa de sucesso média de localização com erro máximo admitido entre 0.1 m e 2.5 m para todos os métodos no Caso A. O método GCC PHAT apresentou taxa de acerto igual a 86% ao admitir erro máximo de 1 m. Esse bom desempenho se deve à sua boa performance na estimação das DOAs, à escolha otimizada da distância entre os microfones e a não estar submetido a erros introduzidos pela mistura de fontes.

O método SCT apresentou o melhor desempenho dentre os métodos apropriados para múltiplas fontes (baseados em BSS), sendo para $D_{MAX} = 1$ m a sua taxa

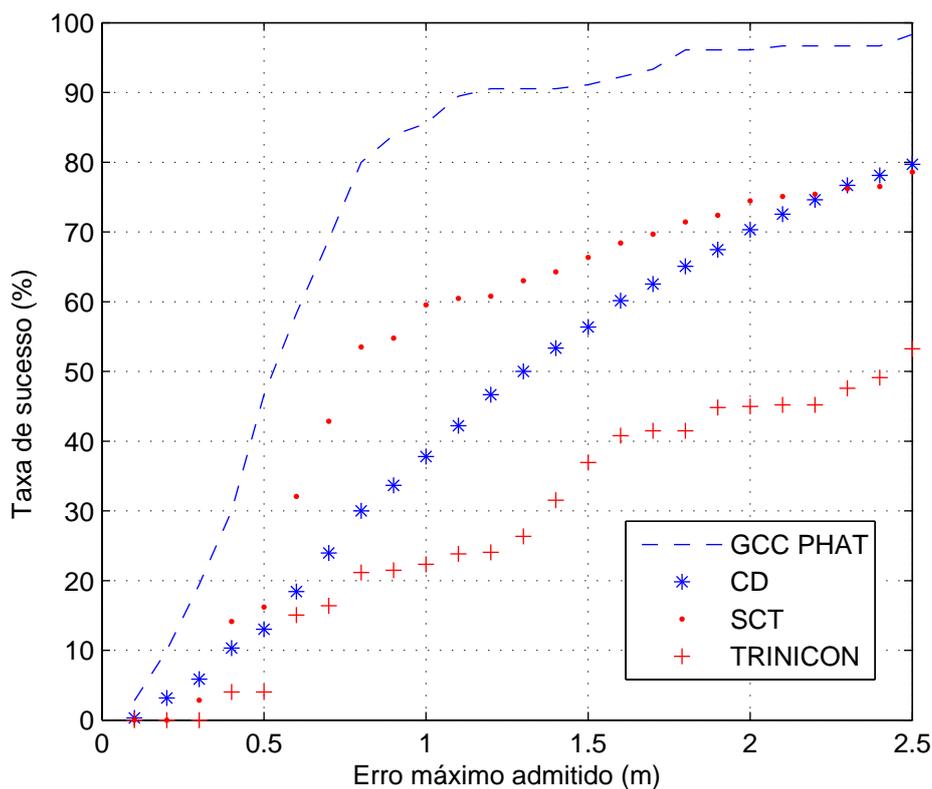


Figura 7.42: Taxa de sucesso média das estimativas das posições ao variar D_{MAX} para os diversos algoritmos no Caso A em ambiente real.

de sucesso igual a 60%, enquanto que os métodos TRINICON e Campo Distante apresentaram taxas de 23% e 38%, respectivamente.

Portanto, pode-se concluir que, dentre os métodos investigados, o SCT é a melhor escolha para o sistema de localização baseado no encontro das DOAs.

7.5 Emprego das Modificações Propostas nos Algoritmos de BSS

Nesta seção analisaremos os desempenhos dos métodos SCT e TRINICON para diferentes comprimentos dos sinais de mistura. Avaliaremos a acurácia das DOAs estimadas pelas implementações convencionais e com as modificações propostas na Seção 4.3 nos ambientes simulados com baixa e alta reverberações.

7.5.1 Método SCT

Realizamos testes com trechos dos sinais de mistura de durações iguais a 1 s, 2 s e 4 s. Para avaliar a performance do algoritmo SCT com a modificação descrita na Seção

4.3.1, variamos a quantidade de raias de maior potência N_p utilizadas para estimar as DOAs. Verificamos que para $N_p < 200$ o algoritmo apresentou desempenho ruim, enquanto para $N_p > 500$ não houve melhora no desempenho. Apresentamos a seguir os resultados obtidos com N_p igual a 200, 500 e 2048, sendo esse último valor igual ao tamanho da DFT empregada, correspondendo, portanto, ao algoritmo convencional.

A Fig. 7.43 apresenta as taxas de sucesso médias variando-se o erro máximo admitido nas estimativas das DOAs entre 1° e 7° no ambiente simulado com $T_{60} = 200$ ms. Na Fig. 7.43 verifica-se que a utilização das raias de maior potência resultou

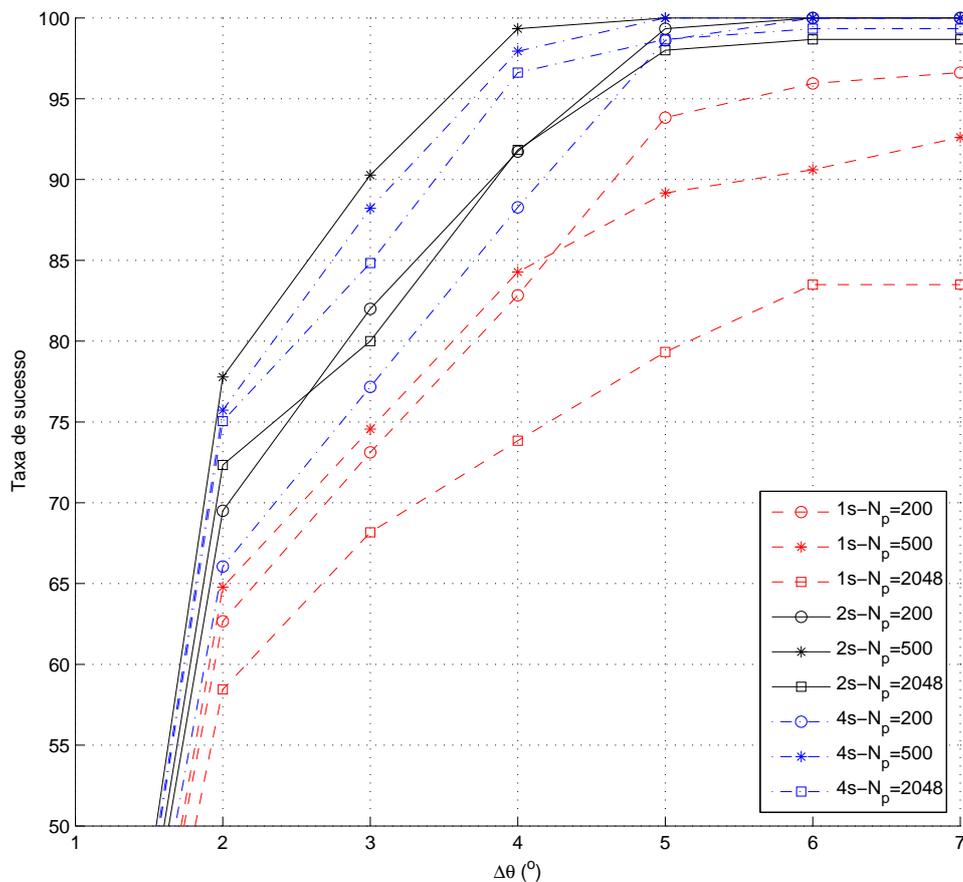


Figura 7.43: Taxa de sucesso média ao variar $\Delta\theta$ do método SCT com N_p raias de frequência e diferentes comprimentos dos sinais no ambiente simulado com $T_{60} = 200$ ms.

em melhor desempenho para todos os intervalos de observação dos sinais. Para o intervalo de 1 s, o algoritmo SCT com $N_p = 200$ raias apresentou taxa de sucesso igual a 94% para erro máximo admitido de 5° , enquanto com $N_p = 2048$ (utilizando todas as raias) a taxa de sucesso foi igual a 78%. Para o intervalo de 2 s, o algoritmo SCT com $N_p = 500$ apresentou o melhor desempenho dentre os casos avaliados, sendo sua taxa de sucesso igual a 99% para erro máximo admitido igual a 4° . Para

o intervalo de 4 s, o melhor desempenho foi obtido com $N_p = 500$, apresentando diferença pequena em relação ao desempenho com $N_p = 2048$. Portanto, para o ambiente simulado com baixa reverberação, o emprego de um número reduzido de raias produz resultados superiores, principalmente quando é necessário obter as estimativas das DOAs em um intervalo pequeno de tempo. Outra vantagem é a diminuição do custo computacional da implementação.

A Fig. 7.44 apresenta as taxas de sucesso médias para erro máximo admitido nas estimativas das DOAs entre 1° e 7° no ambiente simulado com $T_{60} = 800$ ms. Na

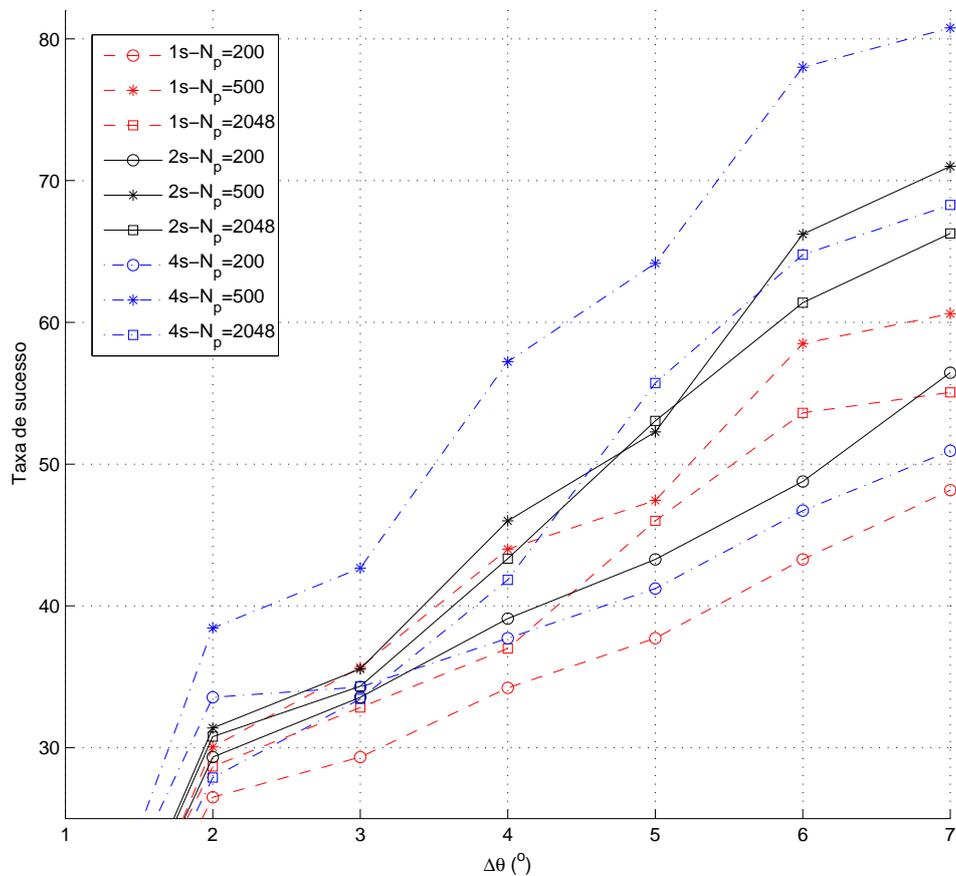


Figura 7.44: Taxa de sucesso média ao variar $\Delta\theta$ do método SCT com N_p raias de frequência e diferentes comprimentos dos sinais no ambiente simulado com $T_{60} = 800$ ms.

Fig. 7.43 verifica-se que a seleção das raias de maior potência resultou em melhor desempenho para praticamente todos os intervalos de observação dos sinais. O maior ganho foi alcançado para o intervalo de 4 s, sendo a taxa de sucesso com $N_p = 500$ cerca de 10% superior à obtida com $N_p = 2048$. Para $\Delta\theta = 7^\circ$, alcançou-se taxa de sucesso média igual a 81% com $N_p = 500$, enquanto que utilizando-se todas as raias a taxa de sucesso média foi apenas 68%. O emprego de $N_p = 200$ raias não

se mostrou adequado no ambiente simulado de alta reverberação, sendo necessário o uso de $N_p = 500$ raias.

7.5.2 Método TRINICON

Foram realizados testes com o método TRINICON utilizando trechos dos sinais de misturas de comprimentos 1 s, 2 s e 4 s. Os desempenhos da implementação convencional e da proposta na Seção 4.3.2 foram avaliados. Observou-se que não houve melhora de desempenho do algoritmo com a modificação proposta quando avaliada a taxa de sucesso média das estimativas da DOA considerando as fontes em todas as posições mostradas na Fig. 5.5. A modificação proposta mostrou-se vantajosa apenas quando as fontes estavam afastadas com diferenças de DOAs entre 45° e 135° . As Figs. 7.45 e 7.46 mostram as taxas de sucesso médias resultantes para essas posições nos ambientes simulados com $T_{60} = 200$ ms e $T_{60} = 800$ ms, respectivamente. Na Fig. 7.45 observa-se uma maior vantagem da modificação

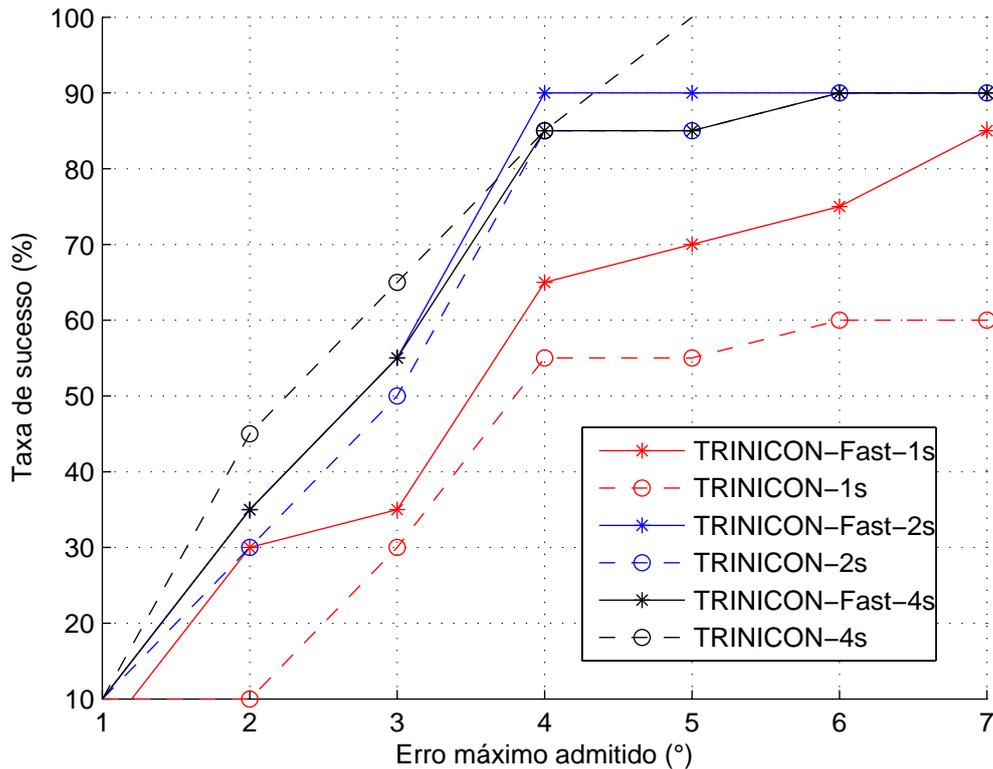


Figura 7.45: Taxa de sucesso média ao variar $\Delta\theta$ do método TRINICON no ambiente simulado com $T_{60} = 200$ ms, para diferenças de DOAs das fontes entre 45° e 135° .

proposta para o intervalo de observação igual a 1 s. Para o intervalo de observação de 4 s, o algoritmo convencional apresentou melhor resultado. Na Fig. 7.46 verifica-se uma melhora considerável do desempenho do algoritmo quando a modificação

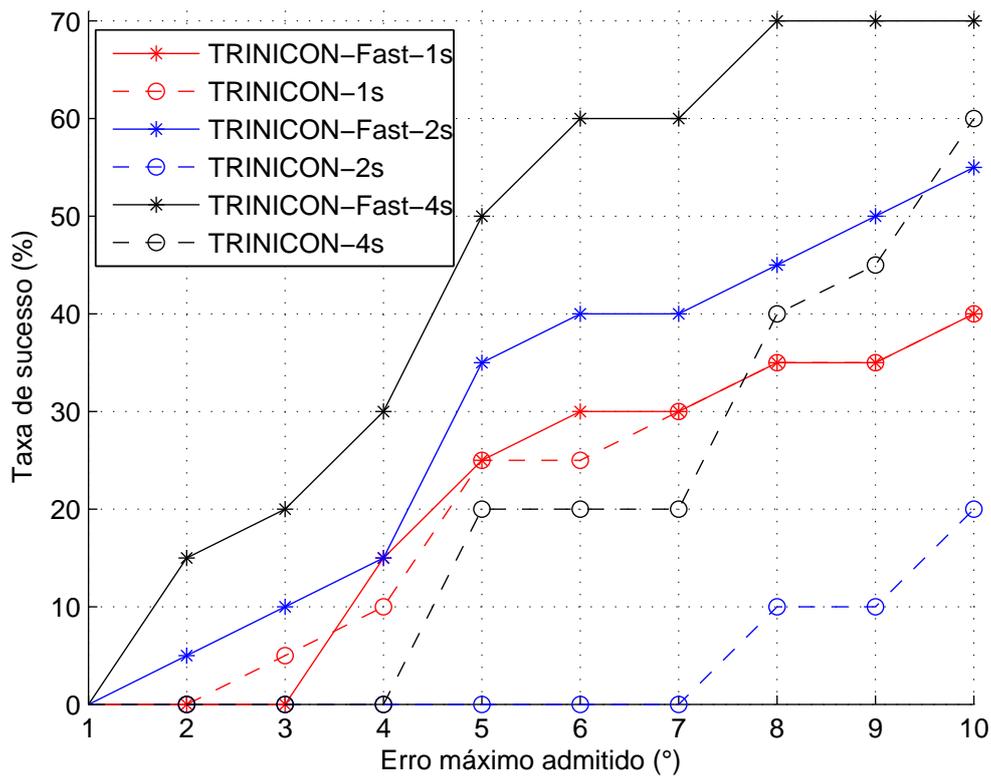


Figura 7.46: Taxa de sucesso média ao variar $\Delta\theta$ do método TRINICON no ambiente simulado com $T_{60} = 800$ ms, para diferenças de DOAs das fontes entre 45° e 135° .

proposta é empregada, principalmente para intervalo de observação igual a 4 s. Portanto, podemos concluir que o método proposto em geral se mostrou superior quando as fontes não se encontram posicionadas em direções muito próximas em relação aos arranjos de microfones.

Capítulo 8

Conclusões e Trabalhos Futuros

Nesta tese investigamos o desempenho de um sistema de localização de múltiplas fontes acústicas baseado em algoritmos de separação cega de fontes, indicando as condições para torná-lo funcionalmente robusto. Ao contrário de outras propostas, o sistema investigado é composto por apenas dois arranjos lineares de sensores, os quais são posicionados em uma única parede, facilitando a sua utilização em diferentes aplicações. Sendo a localização obtida pelo encontro das retas correspondentes às direções de chegada do sinal de cada fonte nos dois arranjos de microfones, a acurácia da localização é afetada pelo desempenho do método de estimação das DOAs e pela distância entre os arranjos de microfones. Analisamos os efeitos do tempo de reverberação do ambiente e da disposição dos dois arranjos de microfones na acurácia de três métodos de estimação das DOAs de múltiplas fontes, para diferentes posições das fontes. A partir dos erros de estimação das DOAs esperados, determinamos a região da sala em que a localização apresentará um erro máximo dentro de um limite aceitável.

Os conceitos de misturas convolutivas e a modelagem de canais acústicos foram apresentados no Cap. 2, enquanto que os principais métodos de separação cega de fontes para os casos de misturas instantâneas e convolutivas foram discutidos no Cap. 3. No Cap. 4 foi descrito o método GCC-PHAT, usado para estimar as direções de chegadas dos sinais nos arranjos de microfones para o caso de haver uma única fonte ativa, e foram apresentados os métodos baseados em separação cega de fontes empregados para estimar as DOAs no caso de haver múltiplas fontes simultaneamente ativas.

No Cap. 5 foi descrito o sistema de localização estudado nesta tese e foi apresentada a proposta de análise da região apropriada de operação, a qual utiliza como parâmetros a acurácia das estimativas das DOAs e o erro máximo admitido nas estimativas das posições. A partir dessa análise, é possível obter diretrizes para tornar o sistema de localização robusto, indicando o posicionamento otimizado dos arranjos e estabelecendo a região apropriada para a localização.

No Cap. 6 foram apresentados os resultados de experimentos para a localização de uma fonte em ambientes simulados de baixa e alta reverberação, com T_{60} iguais a 200 e 800 ms, respectivamente, e em um ambiente real, com T_{60} medido em torno de 900 ms. Dos resultados obtidos no ambiente simulado com baixa reverberação, observaram-se erros de estimativas das DOAs com valores máximos de 1° , permitindo erros de localização da ordem de 25 cm em praticamente toda a sala. Foram verificados experimentalmente os resultados teóricos do Cap. 5, obtendo-se taxas de acerto significativamente superiores quando utilizada a distância ótima entre arranjos. As posições da fonte testadas que se encontravam dentro da região apropriada obtiveram taxas de sucesso de localização bem maiores que as das outras posições.

No caso mais desafiador de localização de uma fonte em ambiente com alta reverberação, os erros nas estimativas das DOAs apresentaram valores máximos em torno de 3° , permitindo taxas de sucesso aceitáveis apenas para erros de localização permitidos iguais a 1.2 m. Nesse caso, também foi observada a validade da análise teórica do Cap. 5, obtendo-se taxas de acertos significativamente superiores nas posições dentro da região apropriada, sendo a sua área aumentada utilizando-se o valor otimizado da distância entre arranjos de microfones.

O ambiente real utilizado foi uma sala de aula com 14 m de comprimento e 10 m de largura, com alta reverberação, caracterizando um ambiente desafiador. Observou-se a limitação do emprego apropriado do sistema para distâncias entre a fonte e os microfones de até 4 m, coincidindo com as posições internas à região apropriada obtida teoricamente. Em geral, as taxas de acertos aumentaram significativamente com a escolha otimizada da posição dos arranjos de microfones.

Portanto, da análise dos resultados dos experimentos de localização de uma fonte, conclui-se que a análise de erro apresentada no Cap. 5 permitiu o aprimoramento do desempenho do sistema em todos os casos, indicando o posicionamento ideal dos pares e a região apropriada para a localização e as condições para o funcionamento robusto do sistema.

No Cap. 7 foram realizados experimentos para a localização de duas fontes simultaneamente ativas nos mesmos ambientes simulados e real usados na avaliação da localização de uma fonte. Dos três algoritmos testados, todos baseados em BSS, o método SCT apresentou os melhores resultados, com maiores taxas de sucesso na localização, em praticamente todos os experimentos.

Quando testado no ambiente simulado com baixa reverberação, o sistema de localização usando o método SCT apresentou taxa média de acerto igual a 97.5% ao admitir erro máximo de 1.2 m. Nos ambientes com alta reverberação, seu desempenho foi significativamente degradado, apresentando taxas médias de acerto iguais a 52% e 62%, respectivamente, para os ambientes simulado e real, considerando o mesmo erro máximo de localização.

Na localização de duas fontes ativas, observou-se também que o posicionamento otimizado dos conjuntos de microfones permitiu melhor rendimento do sistema em todos os casos testados, e que as posições dentro da região apropriada obtida no Cap. 5 mostraram taxas de acertos predominantemente superiores às das outras posições. As taxas de acerto da localização obtidas foram compatíveis com as das DOAs exigidas pela análise teórica, indicando que os resultados da análise podem ser usados para tornar a performance do sistema mais robusta e permitir melhores desempenhos.

O emprego da modificação proposta no algoritmo SCT resultou em estimativas de DOAs mais acuradas e em redução no seu custo computacional. O uso de passos de atualização proporcionais aos valores dos coeficientes no algoritmo TRINICON mostrou-se vantajoso apenas quando as fontes encontravam-se em direções não muito próximas.

Podemos propor, para continuidade deste trabalho, os seguintes tópicos de pesquisa:

- Estabelecer uma relação entre o tempo de reverberação e o erro de estimativa da DOA. Seria possível, assim, através da estimação do T_{60} a partir das misturas [74], usar essa relação para obter o erro máximo das DOAs e automaticamente determinar a região considerada apropriada para o processo de localização;
- Testar a compatibilidade do modelo proposto no Cap. 5 em novos ambientes e utilizá-lo para otimizar a quantidade de conjuntos de sensores;
- Utilizar o método de localização de múltiplas fontes proposto neste trabalho para um número maior de fontes ativas, utilizando mais microfones em cada arranjo;
- Testar o modelo de localização no caso em que as fontes estão fixas, em posições conhecidas, e deseja-se localizar os sensores;
- Investigar outras modificações nos algoritmos de BSS, visando melhorar os seus desempenhos quando empregados na estimação da DOA;
- Estender e testar o modelo proposto em 3 dimensões.

Referências Bibliográficas

- [1] OMOLOGO, M., SVAIZER, P. “Acoustic Event Localization Using a Crosspower-Spectrum Phase Based Technique”, *IEEE International Conference on Acoust., Speech, and Signal Processing*, pp. 273–276, 1994.
- [2] SHOJI MAKINO, H. S., ARAKI, S. “Frequency-Domain Blind Source Separation”. In: *Blind Speech Separation*, 2 ed., cap. 10, Springer Netherlands, 2007.
- [3] MORELANDE, M., KREUCHER, C., KASTELLA, K. “A Bayesian Approach to Multiple Target Detection and Tracking”, *IEEE Trans. on Signal Processing*, v. 55, n. 5, pp. 1589–1604, 2007.
- [4] J. ROMANO, R. ATTUX, C. C., SUYAMA, R. *Unsupervised Signal Processing: Channel Equalization and Source Separation*. CRC Press, 2010.
- [5] HYVARINEN, A., KARHUNEN, J., OJA, E. *Independent Component Analysis*. Wiley, 2004.
- [6] DIBIASE, J. H. *A High-Accuracy, Low-Latency Technique for Talker Localization in Reverberant Environments Using Microphone Arrays*. Tese de Doutorado, Brown University, 2000.
- [7] BRANDSTEIN, M. S., ADCOCK, J. E., SILVERMAN, H. F. “Microphone-Array Localization Error Estimation with Application to Sensor Placement”, *The Journal of the Acoustical Society of America*, v. 99, pp. 3807–3816, 1996.
- [8] NAKANO, A. Y., NAKAGAWA, S., YAMAMOTO, K. “Automatic Estimation of Position and Orientation of an Acoustic Source by a Microphone Array Network”, *The Journal of the Acoustical Society of America*, v. 126, pp. 3084–3094, 2009.
- [9] E G. C. CARTER, C. H. K. “The Generalized Correlation Method for Estimation of Time Delay”, *IEEE Trans. on Acoust., Speech and Signal Processing*, v. 24, pp. 320–327, 1976.

- [10] BUCHNER, H., AICHNER, R., KELLERMANN, W. “TRINICON-Based Blind System Identification with Application to Multiple-Source Localization and Separation”. In: *Blind speech separation*, Springer, pp. 101–147, 2007.
- [11] NESTA, F., OMOLOGO, M. “Generalized State Coherence Transform for Multidimensional TDOA Estimation of Multiple Sources”, *IEEE Trans. on Audio, Speech and Language Processing*, v. 20, n. 1, pp. 246–260, 2012.
- [12] NESTA, F., SVAIZER, P., OMOLOGO, M. “Robust Two-Channel TDOA Estimation for Multiple Speaker Localization by Using Recursive ICA and a State Coherence Transform”. In: *IEEE International Conference on Acoust., Speech and Signal Processing*, pp. 4597–4600, 2009.
- [13] NESTA, F., OMOLOGO, M. “Generalized State Coherence Transform for Multidimensional Localization of Multiple Sources”. In: *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 237–240, 2009.
- [14] NESTA, F., OMOLOGO, M., SVAIZER, P. “Multiple TDOA Estimation by Using a State Coherence Transform for Solving the Permutation Problem in Frequency-Domain BSS”. In: *IEEE Workshop on Machine Learning for Signal Processing.*, pp. 43–48, 2008.
- [15] SAWADA, H., ARAKI, S., MAKINO, S. “Measuring Dependence of Bin-wise Separated Signals for Permutation Alignment in Frequency-domain BSS”. In: *IEEE International Symposium on Circuits and Systems.*, pp. 3247–3250, 2007.
- [16] KURITA, S., SARUWATARI, H., KAJITA, S., et al. “Evaluation of Blind Signal Separation Method Using Directivity Pattern Under Reverberant Conditions”. In: *IEEE International Conference Acoust., Speech, and Signal Processing*, pp. 3140–3143, 2000.
- [17] A. LOMBARD, H. B., KELLERMANN, W. “Multidimensional Localization of Multiple Sound Sources Using Blind Adaptive MIMO System Identification”, *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2006.
- [18] OMOLOGO, M., SVAIZER, P. “Acoustic Event Localization Using a Crosspower-Spectrum Phase Based Technique”. In: *IEEE International Conference on Acoust., Speech, and Signal Processing*, pp. 273–276, 1994.

- [19] OMOLOGO, M., SVAIZER, P. “Acoustic Source Location in Noisy and Reverberant Environment Using CSP Analysis”. In: *IEEE International Conference on Acoust., Speech, and Signal Processing*, v. 2, pp. 921–924, 1996.
- [20] DIBIASE, J., SILVERMAN, H., BRANDSTEIN, M. “Robust Localization in Reverberant Rooms”. In: Brandstein, M., Ward, D. (Eds.), *Microphone Arrays*, Springer Berlin Heidelberg, pp. 157–180, 2001.
- [21] NAKANO, A. Y., NAKAGAWA, S., YAMAMOTO, K. “Distant Speech Recognition Using a Microphone Array Network”, *IEICE Transactions on Information and Systems*, , n. 9, pp. 2451–2462, 2010.
- [22] NAKANO, A. Y., BURT, P. M. S. “Directional Acoustic Source Orientation Estimation Using Only Two Microphones.” *Digital Signal Processing*, v. 23, pp. 1918–1922, 2013.
- [23] WANG, L., KITAOKA, N., NAKAGAWA, S. “Robust Distant Speech Recognition by Combining Multiple Microphone-Array Processing with Position-Dependent CMN”, *EURASIP J. Appl. Signal Process*, 2006.
- [24] BUCHNER, H., AICHNER, R., KELLERMANN, W. “Blind Source Separation for Convolutional Mixtures: A Unified Treatment”. In: *Audio signal processing for next-generation multimedia communication systems*, Springer, pp. 255–293, 2004.
- [25] ABAD, A., SEGURA, C., MACHO, D., et al. “Audio Person Tracking in a Smart-Room Environment”. In: *International Conference on Spoken Language Processing - Interspeech*, 2006.
- [26] FALLON, M., GODSILL, S. “Acoustic Source Localization and Tracking of a Time-Varying Number of Speakers”, *IEEE Trans. on Audio, Speech, and Language Processing*, v. 20, pp. 1409–1415, 2012.
- [27] CHEN, J., YIP, L., ELSON, J., et al. “Coherent Acoustic Array Processing and Localization on Wireless Sensor Networks”, *Proceedings of the IEEE*, v. 91, pp. 1154–1162, 2003.
- [28] WANG, H., CHEN, C. E., ALI, A., et al. “Acoustic Sensor Networks for Woodpecker Localization”. In: *SPIE Conference on Advanced Signal Processing Algorithms, Architectures, and Implementations*, 2005.
- [29] OMOLOGO, M., SVAIZER, P. “Use of the Crosspower-Spectrum Phase in Acoustic Event Location”, *IEEE Trans. on Speech and Audio Processing*, v. 5, pp. 288–292, 1997.

- [30] DOBLINGER, G. “Localization and Tracking of Acoustical Sources”. In: *Topics in Acoustic Echo and Noise Control*, Springer, pp. 91–120, Berlin - Heidelberg, 2006.
- [31] A. Y. NAKANO, S. N., YAMAMOTO, K. “Distant Speech Recognition Using a Microphone Array Network”, *IEICE Transactions*, v. 93-D, n. 9, pp. 2451–2462, 2010.
- [32] WANG, L., KITAOKA, N., NAKAGAWA, S. “Robust Distant Speaker Recognition Based on Position-Dependent CMN by Combining Speaker-Specific GMM with Speaker-Adapted HMM”, *Speech Communication*, 2007.
- [33] HANSLER, E., SCHMIDT, G. *Topics in Acoustic Echo and Noise Control*. Springer-Verlag New York, Inc., 2006.
- [34] BENESTY, J. “Adaptive Eigenvalue Decomposition Algorithm for Passive Acoustic Source Localization”, *The Journal of the Acoustical Society of America*, v. 107, pp. 384–391, 2000.
- [35] FERRARA, E. R. “Fast Implementations of LMS Adaptive Filters”, *IEEE Trans. on Acoust., Speech and Signal Processing*, v. 28, pp. 474–475, 1980.
- [36] HUANG, Y., BENESTY, J., HUANG, Y. *Audio Signal Processing for Next-Generation Multimedia Communication Systems*. Boston, [Mass.] ; London : Kluwer Academic, 2004.
- [37] MEESOOKHO, C., MITRA, U., NARAYANAN, S. “On Energy-Based Acoustic Source Localization for Sensor Networks”, *IEEE Trans. on Signal Processing*, v. 56, pp. 365–377, 2008.
- [38] BRANDSTEIN, M., ADCOCK, J., SILVERMAN, H. “A Closed-Form Location Estimator for Use With Room Environment Microphone Arrays”, *IEEE Trans. on Speech and Audio Processing*, v. 5, n. 1, pp. 45–50, 1997.
- [39] BENESTY, J., CHEN, J., HUANG, Y. *Microphone Array Signal Processing*. Springer Topics in Signal Processing. Springer, 2008.
- [40] MA, W.-K., VO, B.-N., SINGH, S., et al. “Tracking an Unknown Time-Varying Number of Speakers Using TDOA Measurements: A Random Finite Set Approach”, *IEEE Trans. on Signal Processing*, v. 54, n. 9, pp. 3291–3304, 2006.
- [41] COMON, P. “Independent Component Analysis, A New Concept?” *Elsevier Science*, v. 36, n. 3, pp. 287–314, 1994.

- [42] J. HERAULT, C. J., ANS, B. “Detection de Grandeurs Primitives dan un Message Composite Par Une Architecture de Calcul Neuromimetique en Apprentissage Non Supervise”, *Proceeding of the 10th GRETSI Symposium on Signal and Image Processing*, 1988.
- [43] LACOUME, J. L., RUIZ, P. “Source Identification: A Solution Based on Cumulants”, *Fourth Annual ASSP Workshop on Spectrum Estimation and Modeling*, 1988.
- [44] CARDOSO, J.-F. “Blind Signal Separation: Statistical Principles”, *Proceedings of the IEEE*, v. 86, n. 10, pp. 2009–2025, 1998.
- [45] BELL, A. J., SEJNOWSKI, T. J. “An Information-Maximization Aproach to Blind Separation and Blind Deconvolution”, *Neural Computing - Massachusetts Institute of Technology*, 1995.
- [46] CICHOCKI, A., UNBEHAUEN, R. “Robust Neural Networks with On-Line Learning for Blind Identification and Blind Separation of Sources”, *IEEE Trans. on Circuits and Systems I: Fundamental Theory and Applications*, v. 43, n. 11, pp. 894–906, 1996.
- [47] HYVARINEN, A. “Fast and Robust Fixed-Point Algorithms for Independent Component Analysis”, *IEEE Trans. on Neural Networks*, v. 10, pp. 626–634, 1999.
- [48] MAKINO, S., SAWADA, H., MUKAI, R., et al. “Blind Source Separation of Convulsive Mixtures of Speech in Frequency Domain”, *IEICE Trans. Fundamentals*, v. E88-A, n. 7, pp. 1640–1655, 2005.
- [49] SMARAGDIS, P. “Blind Separation of Convolved Mixtures in The Frequency Domain”, *Neurocomputing - Elsevier*, v. 22, n. 1, pp. 21–34, 1998.
- [50] MITIANOUDIS, N., DAVIES, M. “Audio Source Separation of Convulsive Mixtures”, *IEEE Trans. on Speech and Audio Processing*, v. 11, n. 5, pp. 489–497, 2003.
- [51] ASANO, F., IKEDA, S., OGAWA, M., et al. “Combined Approach of Array Processing and Independent Component Analysis for Blind Separation of Acoustic Signals”, *IEEE Trans. on Speech and Audio Processing*, v. 11, pp. 204–215, 2003.
- [52] SCHOBEN, D. W. E., SOMMEN, P. “A Frequency Domain Blind Signal Separation Method Based on Decorrelation”, *IEEE Trans. on Signal Processing*, v. 50, pp. 1855–1865, 2002.

- [53] MATSUOKA, K., NAKASHIMA, S. “Minimal Distortion Principle for Blind Source Separation”, *Proceedings of the International Conference on Independent Component Analysis and Blind Signal Separation*, v. 4, pp. 2138–2143, 2001.
- [54] MATSUOKA, K. “Minimal Distorsion Principle for Blind Source Separation”, *Proceeding of the 41st SICE annual conference*, v. 4, pp. 2138–2143, 2002.
- [55] BUCHNER, H., AICHNER, R., KELLERMANN, W. “A Generalization of Blind Source Separation Algorithms for Convolutional Mixtures Based on Second-Order Statistics”, *IEEE Trans. on Speech and Audio Processing*, v. 13, pp. 120–134, 2005.
- [56] HADDAD, D. B. *Estruturas em Subbandas para Filtragem Adaptativa e Separacao Cega e Semi-Cega de Sinais de Voz*. Tese de Doutorado, COPPE/URFJ, 2013.
- [57] KUTTRUFF, H. *Room Acoustics*. Spon press, 2000.
- [58] BLAUERT, J. *Communication Acoustics*. Springer, 2005.
- [59] Y. HUANG, J. B., CHEN, J. *Acoustic MIMO Signal Processing*. Springer, 2006.
- [60] LEHMANN, E., JOHANSSON, A., NORDHOLM, S. “Reverberation-Time Prediction Method for Room Impulse Responses Simulated with the Image-Source Model”. In: *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 159–162, 2007.
- [61] LEHMANN, E. A., JOHANSSON, A. M. “Prediction of Energy Decay in Room Impulse Responses Simulated with an Image-Source Model”, *The Journal of the Acoustical Society of America*, v. 124, pp. 269–277, 2008.
- [62] ALLEN, J. B., BERKLEY, D. A. “Image Method For Efficiently Simulating Small Room Acoustics”, *The Journal of the Acoustical Society of America*, v. 65, pp. 943–950, 1979.
- [63] MATHWORKS. *The Exponential Integral $Ei(.)$ Is a Standard Function Building Into Several Mathematical Packages Such as MATLAB, MAPLE and MATHEMATICAL*, 2014.
- [64] LI, K. M., LAM, P. M. “Prediction of Reverberation Time And Speech Transmission Index in Long Enclosures”, *The Journal of the Acoustical Society of America*, v. 117, pp. 3716–3726, 2005.

- [65] SCHROEDER, M. R. “New Method of Measuring Reverberation Time”, *The Journal of the Acoustical Society of America*, v. 37, n. 6, pp. 1187–1188, 1965.
- [66] AMARI, S.-I., CICHOCKI, A., YANG, H. H., et al. “A New Learning Algorithm for blind Signal Separation”, *Advances in Neural Information Processing Systems*, pp. 757–763, 1996.
- [67] ANEMÜLLER, J., KOLLMEIER, B. “Amplitude Modulation Decorrelation for Convolutional Blind Source Separation”, *Proc. Second International Workshop*, pp. 215–220, 2000.
- [68] BUCHNER, H., AICHNER, R., KELLERMANN, W. “TRINICON: A Versatile Framework for Multichannel Blind Signal Processing”. In: *IEEE International Conference on Acoust., Speech, and Signal Processing, 2004*, pp. 889–892, 2004.
- [69] HAYKIN, S. *Unsupervised Adaptive Filtering: Blind Deconvolution*. Wiley, 2000.
- [70] AMARI, S. “Natural Gradient Works Efficiently in Learning”, *Neural Computation*, v. 10, n. 2, pp. 251–276, Feb 1998.
- [71] VINCENT, E., GRIBONVAL, R., FEVOTTE, C. “Performance Measurement in Blind Audio Source Separation”, *IEEE Trans. on Audio, Speech, and Language Processing*, v. 14, pp. 1462–1469, 2006.
- [72] KWON, B., PARK, Y., SIK PARK, Y. “Analysis of the GCC-PHAT Technique for Multiple Sources”. In: *International Conference on Control Automation and Systems*, pp. 2070–2073, 2010.
- [73] H. BUNCHER, R. A., KELLERMANN, W. “TRINICON: A Versatile for Multichannel Blind Signal Processing”, *IEEE International Conference on Acoust., Speech, and Signal Processing*, v. 3, pp. 889–892, 2004.
- [74] BARRETO HADDAD, D., REMBOLD PETRAGLIA, M., BULKOO BATALHEIRO, P. “Using Reverberation Time Estimates in Blind Separation of Acoustic Sources”. In: *Digital Signal Processing and Signal Processing Education Meeting*, pp. 153–157, 2013.