

Universidade Federal do Rio de Janeiro

Escola Politécnica

Departamento de Engenharia Eletrônica e de Computação

Métodos para Localização de Sensores Acústicos

Autor:

Maurício do Vale Madeira da Costa

Banca Examinadora

Orientador:

Wallace Alves Martins

Orientador:

Luiz Wagner Pereira Biscainho

Examinador:

Marcello Luiz Rodrigues de Campos

Examinador:

Tadeu Nagashima Ferreira

DEL

Agosto de 2013

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Escola Politécnica - Departamento de Eletrônica e de Computação

Centro de Tecnologia, bloco H, sala H-217, Cidade Universitária

Rio de Janeiro - RJ CEP 21949-900

Este exemplar é de propriedade da Universidade Federal do Rio de Janeiro, que poderá incluí-lo em base de dados, armazená-lo em computador, microfilmá-lo ou arquivá-lo em qualquer outra forma que decidir adotar.

É permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do(s) autor(es) e do(s) orientador(es).

Dedico este trabalho a todos os meus amigos e familiares.

“O livre-pensamento eleva a dignidade do homem, dele fazendo um ser ativo, inteligente, em vez de uma *máquina de crer*.”

Allan Kardec, Revista Espírita de fev/1867

Agradecimentos

Agradeço, primeiramente, a Deus, pelo dom da vida e por todas as oportunidades de crescimento que tive. Agradeço a meus pais pelo apoio e formação que me deram, me possibilitando alcançar mais uma conquista. Agradeço a minha irmã Isabela pelo convívio e amizade durante toda minha vida e pela companhia, muitas vezes, nas horas de trânsito indo ou voltando da faculdade. Agradeço à Julia, minha namorada, que durante quatro longos anos me acompanhou e participou de muitos momentos importantes para minha formação. Agradeço também a todos os meus outros familiares — que são muitos — mas, em especial, ao meu avô Edson Madeira, que durante toda a minha vida foi peça fundamental como incentivo à busca pelo conhecimento.

Gostaria de agradecer a meus orientadores Luiz Wagner, Wallace Martins e Leonardo Nunes pela paciência e tempo que me dedicaram durante minha formação acadêmica e pela amizade que cultivamos. Agradeço também aos professores Tadeu Ferreira e Marcello Campos por aceitarem participar da banca deste trabalho.

Agradeço a todos os meus amigos e colegas que compartilharam comigo dessa longa e trabalhosa jornada, tornando-a, muitas vezes, mais agradável e divertida. Dentre muitos, gostaria de citar nomes como Gabriel Gouvêa, Leonardo Muricy, Isabela Apolinário, Lucas Thomaz, Lucas Maia, Lucas Frucht, Allan Freitas, Felipe Mayer, João Bernardo, Camila Vasconcellos, Paulo Catunda, Luiz Tavares, Jonathan Gois, Rick Ferreira, Hugo Andrade, Igor Gameleiro e Bernardo Cid.

Agradeço, igualmente, a todos os companheiros dos laboratórios LPS e SMT, por muito terem me ensinado e compartilhado vivências, em especial, os companheiros do Grupo de Processamento de Áudio. Agradeço também a todos os companheiros que trabalharam comigo no projeto de pesquisa MuQuE e MuQuE-AS, sem os quais este trabalho não seria realizado.

Finalmente, gostaria de agradecer aos professores e funcionários do DEL, que, em sua maioria, realizaram um trabalho de excelência, contribuindo para que esse curso de graduação mereça carregar consigo o nome dessa universidade.

Resumo

Em sistemas modernos de comunicações, já se recorre ao uso de arranjos de microfones [1] para localização de fontes sonoras e captura de seus sinais acústicos, possibilitadas pela multiplicidade de versões dos mesmos sinais medidas em diferentes posições (diversidade espacial). Em tais aplicações, é comum que haja a necessidade de se conhecer precisamente as posições dos sensores acústicos.

Visando a atacar esse problema, este trabalho faz parte de uma linha de pesquisa cujo interesse é a localização dos sensores, e não das fontes sonoras. O ponto de partida foi a implementação das técnicas propostas no artigo [2], onde os autores descrevem um método desenvolvido para calibração automática das posições espaciais relativas de múltiplos microfones e alto-falantes, tendo como única restrição que exista um microfone acoplado a um alto-falante. Tendo obtido uma primeira estimativa das posições dos microfones utilizando o método de mínimos quadrados, é realizada uma otimização não-linear envolvendo o tempo de percurso entre cada sinal de calibração emitido pelos alto-falantes e cada microfone.

O algoritmo proposto em [2] foi implementado e testado através de simulações em diversos ambientes virtuais, como forma de aferir seu desempenho sob condições reais envolvendo, por exemplo, reverberação, fontes sonoras indesejadas e diversas dimensões de sala. Na etapa posterior, foi investigado o desempenho do método em situações reais, verificando-se a possibilidade de estender sua aplicação a contextos menos controlados. Com esses resultados em mãos, foi possível propor formas alternativas de processamento que permitem a substituição da etapa não-linear presente em [2]. Tal substituição permitiu que o algoritmo resultante apresentasse robustez a situações com reverberação, sendo menos custoso do ponto de vista computacional.

Sumário

Agradecimentos	iv
Resumo	vi
Lista de Figuras	ix
Lista de Tabelas	xii
1 Apresentação	1
1.1 Introdução	1
1.2 Organização	3
2 Fundamentos de Acústica	4
2.1 Fenômeno do Som	4
2.2 Ambientes Acústicos	6
2.2.1 Reflexão de ondas sonoras	7
2.2.2 Acústica de Salas	8
2.3 Alto-falantes: Atuadores Acústicos	11
2.4 Microfones: Sensores Acústicos	14
2.4.1 Tipos e Princípios Básicos	14
2.4.2 Padrões Polares Básicos	16
3 Localização de Sensores Acústicos	20
3.1 Definições Preliminares	20
3.2 Localização por Sistemas Ideais	21
3.3 Localização em Sistemas Reais	22
3.3.1 Estimação de TOF	23

3.3.2	Método de Mínimos Quadrados	25
3.3.3	Método de Otimização Não-Linear	26
3.3.4	Método de Seleção de TOFs [8]	28
4	Experimentos: Procedimentos, Resultados e Discussões	33
4.1	Estudo da Ortogonalidade dos Sinais Emitidos	33
4.2	Simulações	39
4.2.1	Metodologia	39
4.2.2	Resultados	41
4.3	Experimento Prático	50
4.3.1	<i>Setup</i> Experimental	50
4.3.2	Resultados	57
5	Conclusões e Trabalhos Futuros	61
5.1	Conclusões	61
5.2	Trabalhos Futuros	62
	Bibliografia	63

Lista de Figuras

1.1	Ilustração do uso de um arranjo de microfones.	2
2.1	Ondas de pressão no ar.	5
2.2	Modelo de fonte sonora esférica.	6
2.3	Ilustração da interação da energia do som com materiais.	7
2.4	Sinal acústico refletindo-se em uma superfície.	8
2.5	Ilustração detalhada de uma frente de onda refletindo-se em superfície.	8
2.6	Sinal acústico refletindo-se em superfícies de uma sala.	10
2.7	Resposta ao impulso de uma sala real.	10
2.8	Resposta em frequência medida de uma sala real.	11
2.9	Ilustração do princípio de funcionamento de um alto-falante.	12
2.10	Ilustração da direcionalidade de um alto-falante.	14
2.11	Ilustração do princípio de funcionamento de um microfone condensador.	16
2.12	Padrões polares mais comuns de microfones.	18
3.1	Exemplo da influência do ruído aditivo na correlação cruzada.	29
3.2	Exemplo de CC de um sinal gravado evidenciando uma situação onde o maior pico não está relacionado ao sinal de visada direta. Estas figuras foram adaptadas de [8].	30
4.1	Comparação entre autocorrelações de <i>chirps</i> variando largura de banda e faixa de frequências.	35
4.2	Comparação entre autocorrelações de <i>chirps</i> com diferentes números de amostras.	36
4.3	Comparação entre autocorrelações de sinais pseudoaleatórios com diferentes números de amostras.	37

4.4	Comparação entre auto-correlações de sinais pseudo-aleatórios corrompidos por ruído e com diferentes números de amostras (SNR = 0 dB).	38
4.5	Resultado para LS e seleção de TOFs das simulações variando a SNR em uma sala anecoica de dimensões 5,2 m × 7,5 m × 2,6 m.	42
4.6	Resultado para LS e seleção de TOFs das simulações variando o RT-60 de uma sala de dimensões 5,2 m × 7,5 m × 2,6 m com SNR idealmente infinita.	42
4.7	Histogramas das simulações de uma sala pequena com RT-60 = 300 ms.	44
4.8	Histogramas das simulações de uma sala pequena com RT-60 = 600 ms.	45
4.9	Histogramas das simulações de uma sala média com RT-60 = 400 ms.	46
4.10	Histogramas das simulações de uma sala média com RT-60 = 800 ms.	47
4.11	Histogramas das simulações de uma sala grande com RT-60 = 700 ms.	48
4.12	Histogramas das simulações de uma sala grande com RT-60 = 1400 ms.	49
4.13	Microfone condensador omnidirecional [16].	50
4.14	Especificações técnicas do microfone.	51
4.15	Fotografia do alto-falante utilizado.	52
4.16	Referência para medição da posição dos alto-falantes.	52
4.17	Microfones utilizados para o experimento prático montados no pedestal. Os círculos brancos indicam a localização dos pequenos microfones.	53
4.18	<i>Setup</i> utilizado no experimento prático. Há 12 alto-falantes (esferas pretas na imagem), e 16 microfones.	54
4.19	Esquemático do sistema de gravação e reprodução do experimento.	55
4.20	<i>Setup</i> utilizado para medir o atraso causado pelos equipamentos.	56
4.21	Histogramas das realizações do experimento com <i>chirps</i> reproduzidos individualmente.	58
4.22	Histogramas das realizações do experimento com <i>chips</i> pseudoaleatórios reproduzidos individualmente.	59

4.23 Histogramas das realizações do experimento com <i>chips</i> pseudoaleatórios reproduzidos com intervalo de 3 ms.	60
-------------------------------------------------------------------------------------------------------------------------------	----

Lista de Tabelas

2.1	Resumo dos padrões polares [12].	18
4.1	Resumo dos resultados da simulação da sala pequena.	45
4.2	Resumo dos resultados da simulação da sala média.	46
4.3	Resumo dos resultados da simulação da sala grande.	48
4.4	Resumo dos resultados do experimento com <i>chirps</i> reproduzidos individualmente.	58
4.5	Resumo dos resultados do experimento com <i>chips</i> pseudoaleatórios reproduzidos individualmente.	58
4.6	Resumo dos resultados do experimento com <i>chips</i> pseudoaleatórios reproduzidos com intervalo de 3 ms.	59

Capítulo 1

Apresentação

A Seção 1.1 introduz o tema de localização de sensores acústicos e motiva seu estudo e desenvolvimento. Em seguida, a Seção 1.2 apresenta brevemente o conteúdo abordado em cada capítulo.

1.1 Introdução

Atualmente, existem diversas atividades em que há interesse em se conhecer a posição de uma determinada fonte sonora. Em situações de investigação, por exemplo, pode ser de grande interesse localizar fontes como tiros disparados ou pessoas falando [3]. Em outros casos, o conhecimento da posição de uma determinada fonte pode ser útil como forma de se melhorar a qualidade do sinal capturado, como em discursos ou teleconferências.

Uma maneira eficaz de realizar este tipo de processamento é através da captura dos sinais de interesse por sistemas chamados de arranjos de microfones [1]. Estes sistemas consistem no uso de diversos microfones espacialmente espalhados, de forma a obter-se os sinais capturados com diferentes tempos de chegada.¹ Utilizando as informações de posição dos microfones e dos atrasos relativos dos sinais de chegada, é possível processar estes sinais com as finalidades citadas anteriormente, como estimar a posição do emissor, quando se fala de localização de fontes sonoras (*sound source localization* — SSL) [4][5], e realizar realce de fala (*speech*

¹Considera-se “tempo de chegada” como o tempo de viagem do estímulo sonoro do transmissor ao receptor (do alto-falante ao microfone).

enhancement) [3][6].

A Figura 1.1 ilustra o uso destes sistemas, onde há um arranjo de dois microfones composto por Mic 1 e Mic 2 capturando duas fontes sonoras, Fonte 1 e Fonte 2. É possível observar que pela diferença nas distâncias relativas, o som emitido pela Fonte 1 chegará mais rapidamente ao Mic 1 que o som da Fonte 2, enquanto ocorre o oposto no Mic 2. A partir do processamento dos sinais capturados é possível, por exemplo, realçar o sinal da fonte de maior interesse, melhorando sua inteligibilidade.

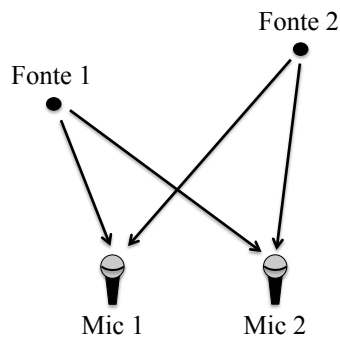


Figura 1.1: Ilustração do uso de um arranjo de microfones.

As aplicações de SSL consideram que as posições dos microfones são exatamente conhecidas previamente [2]. Porém, pequenos erros de medição das posições dos microfones podem causar erros substanciais na localização das fontes. Em alguns casos, medir as posições de microfones distantes ou em configurações espacialmente complexas pode se tornar inviável ou consideravelmente imprecisa. Em outros, a meta pode ser localizar dispositivos em campo [7]. Em tais aplicações, a localização de sensores acústicos (*acoustic sensor localization*—ASL) é a atividade fim, enquanto as posições das fontes sonoras são conhecidas a priori. A ASL pode ser considerada dual da SSL, sendo, entretanto, menos discutida na literatura. Como principal referência deste trabalho, utilizaremos um método de calibração do posicionamento espacial de microfones e alto-falantes desenvolvido por *Vikas C. Raykar* e *Ramani Duraiswami* em [2].

Com o intuito de deduzir uma expressão em forma fechada que permita estimar as posições dos microfones, o método implementado faz uso de medidas dos atrasos entre os sinais reproduzidos nos alto-falantes e os sinais recebidos nos microfones, chamados TOFs (*times of flight*). As posições dos microfones e dos próprios alto-falantes são desconhecidas a priori. O procedimento de calibração é realizado em

duas etapas: primeiro, são localizados os alto-falantes; posteriormente, são localizados os microfones, valendo-se das posições dos alto-falantes previamente estimadas. Após feita uma primeira estimativa das posições dos microfones, é realizado um refinamento dos resultados por otimização de uma função-custo não-linear envolvendo os TOFs.

Para avaliar o desempenho do método em situações reais, é necessário conhecer o efeito causado pelas condições ambientais mais comuns. Para que estas avaliações sejam feitas, serão realizados experimentos em ambientes de simulação e, posteriormente, em ambientes reais. As simulações serão de grande valia, pois facilitam a obtenção de dados estatísticos do método, o que é inviável em experimentos práticos. Em nosso caso, desejamos fazer um apanhado geral do desempenho do método, valendo-nos de diferentes dimensões de salas, diferentes tempos de reverberação e presença de ruído de fundo.

1.2 Organização

No Capítulo 2, é feito um breve estudo do som enquanto fenômeno físico e um estudo sobre sensores e atuadores acústicos, resumindo os fundamentos teóricos necessários para o melhor entendimento dos métodos utilizados e melhor interpretação dos experimentos realizados.

Em seguida, no Capítulo 3, são discutidos os métodos de localização de sensores acústicos presentes na literatura [2] utilizada como referência para o trabalho, assim como um novo método proposto [8].

O Capítulo 4 apresenta a descrição dos procedimentos experimentais realizados como testes de desempenho dos métodos apresentados. São descritas simulações realizadas com grande variação dos parâmetros acústicos e espaciais, de forma a abranger diversos possíveis cenários reais, assim como os experimentos práticos realizados. Ao longo do capítulo, são mostrados e discutidos os resultados obtidos em cada etapa.

Por fim, no Capítulo 5, são expostas as conclusões do trabalho, incluindo discussões acerca dos métodos implementados, dos resultados obtidos e de trabalhos futuros, que continuem o que foi realizado e documentado neste projeto.

Capítulo 2

Fundamentos de Acústica

O som está sempre presente em nossas vidas, quando falamos, ouvimos música ou percebemos o mundo à nossa volta. A audição é responsável por nos proporcionar uma sensação tridimensional do ambiente. Mas, o que é o som? Como se dá o fenômeno físico que envolve um dos mais importantes sentidos do ser humano? Neste capítulo, será feito um breve estudo do som enquanto fenômeno físico e um estudo sobre sensores e atuadores acústicos.

2.1 Fenômeno do Som

A Acústica, área da Física cujo objeto de estudo é o som, define-o como um fenômeno relacionado à variação de pressão em um meio material [9]. Sua geração, propagação e percepção estão diretamente relacionadas com vibrações ou oscilações mecânicas. A faixa audível, para os seres humanos, tem limites inferior de 20 Hz e superior de 20.000 Hz. Qualquer vibração mecânica com informação frequencial acima desta faixa será considerada ultrassom, e abaixo, infrassom, embora sejam fenômenos idênticos ao som quanto a sua natureza.

Para facilitar a explicação do fenômeno, consideremos que o som está sendo gerado pela vibração de uma membrana no ar, como um corpo de um instrumento de corda, ou a membrana de um alto-falante. A Figura 2.1 mostra três diferentes situações possíveis. As linhas verticais representam a concentração do ar à frente da membrana.

Analisando-se a Figura 2.1, é possível observar um primeiro instante, onde

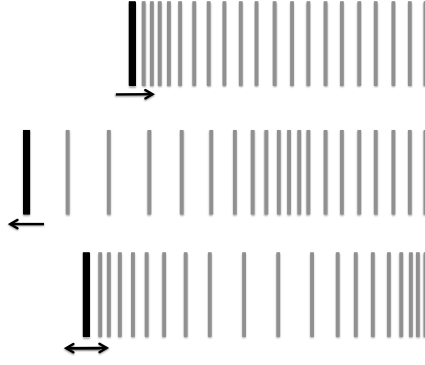


Figura 2.1: Ondas de pressão no ar.

a membrana está se deslocando para a frente, empurrando o ar da região próxima a ela e forçando um aumento na pressão local; no segundo, a membrana se desloca para trás, reduzindo a pressão local; o terceiro instante mostra a pressão resultante quando a membrana realiza um movimento oscilatório. Com o aumento da pressão, a taxa de colisão entre as moléculas aumenta, forçando o deslocamento gradativo das moléculas para regiões vizinhas, até que a pressão seja uniforme. O oposto ocorre quando há uma diminuição na pressão local, pois haverá uma pressão maior nas regiões vizinhas. Este fenômeno gera uma onda de pressão, levando o estímulo a viajar até que a energia se dissipe em forma de calor.

Podemos modelar este fenômeno como um sistema linear, onde a entrada é a posição horizontal $x_m(t)$ da membrana, enquanto a saída será a pressão $p_o(t)$ em um ponto de observação $o \in \mathbb{R}_+$. Consideraremos que a membrana oscila em torno de um ponto $f \in \mathbb{R}$. Como há um atraso na propagação da variação de pressão, teremos que a pressão no ponto o será $p_o(t) = \alpha x_m(t - \tau)$, onde $\alpha \in \mathbb{R}_+$ será uma constante de atenuação do sinal e τ será o TOF, citado anteriormente na Seção 1.1, um atraso definido em função da velocidade de propagação $c \in \mathbb{R}_+$ e das posições o e f . Assim, para este simples modelo unidimensional,

$$p_o(t) = \alpha x_m \left(t - \frac{\|o - f\|}{c} \right), \quad (2.1)$$

onde $\|\cdot\|$ denota a norma euclidiana. Para a velocidade de propagação no ar, será considerada a aproximação $c = 331 + 0,6T$ [2], com T medida em graus centígrados.

Como modelo mais geral, no espaço tridimensional, podemos considerar que,

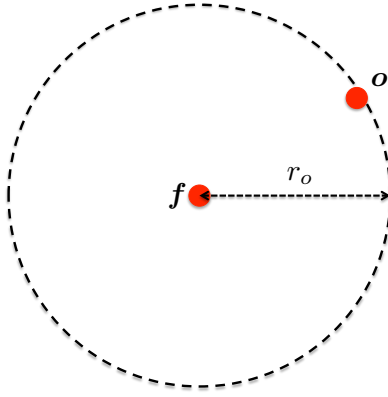


Figura 2.2: Modelo de fonte sonora esférica.

se a fonte for pontual e omnidirecional¹ e o meio for isotrópico, o estímulo se propagará de forma esférica. A Figura 2.2 ilustra o modelo, onde há dois pontos assinalados: o ponto do observador $\mathbf{o} \in \mathbb{R}^3$ e um ponto $\mathbf{f} \in \mathbb{R}^3$, onde se encontra o emissor. Como a emissão partirá de um ponto e se dará de forma esférica, teremos que a potência será igualmente distribuída pela superfície da esfera. Se considerarmos $P_{\mathbf{f}} \in \mathbb{R}_+$ a potência do sinal gerado no ponto \mathbf{f} , $I_{\mathbf{o}} \in \mathbb{R}_+$ a intensidade² em \mathbf{o} e $r_{\mathbf{o}}$ como a distância $r_{\mathbf{o}} = \|\mathbf{o} - \mathbf{f}\|$, teremos na equação (2.2) a chamada Lei do Quadrado da Distância:

$$I_{\mathbf{o}} = \frac{P_{\mathbf{f}}}{4\pi r_{\mathbf{o}}^2}. \quad (2.2)$$

2.2 Ambientes Acústicos

A onda sonora interage com objetos, gerando fenômenos conhecidos como reflexão, absorção, refração e difração. No caso de um ambiente acústico simples, como uma sala vazia, o som que se propaga no ar encontra frequentemente as superfícies das paredes, do teto e do chão. A Figura 2.3 ilustra a distribuição de potência quando uma onda sonora incide sobre uma superfície genérica. Nessa imagem, a seta com P_i representa a potência incidente; a seta com P_r representa a potência

¹Fontes omnidirecionais, ou onidirecionais, são fontes que geram mudanças de pressão iguais em todas as direções.

²Intensidade de radiação é a quantidade de energia que flui por unidade de área por unidade de tempo. É, portanto, a potência por unidade de área dada em [W/m²].

refletida, ou devolvida ao mesmo meio de incidência; a seta com P_a se refere à potência absorvida pelo material, convertendo a energia acústica em energia térmica; e a seta com P_t representa a potência transmitida do material ao meio posterior. Pelo princípio da conservação de energia, temos que $P_i = P_r + P_a + P_t$.

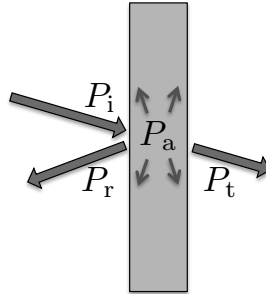


Figura 2.3: Ilustração da interação da energia do som com materiais.

A seguir, na Subseção 2.2.1, será estudado de forma mais detalhada o fenômeno da reflexão do ponto de vista geométrico. Para este breve estudo, não será considerado o efeito de difração, por não ser relevante para esta aplicação.

2.2.1 Reflexão de ondas sonoras

O fenômeno da reflexão se dá de forma semelhante ao estudado pela Óptica. Se o som é refletido por uma superfície plana, o efeito causado é equivalente a haver uma segunda fonte em posição simétrica à da verdadeira em relação à superfície. Porém, esta fonte secundária estaria emitindo com menor potência, por conta da energia perdida por absorção e transmissão. A Figura 2.4 ilustra este fenômeno, representando uma fonte pelo círculo preto, a propagação esférica por linhas cheias e a reflexão pelas linhas tracejadas. A linha do meio representa a superfície refletora, ou anteparo. Observando atentamente, é possível perceber as frentes de ondas incidentes sendo refletidas e, conforme invertem sua direção de propagação naquele eixo, se configurando como se tivessem sido emitidas pela fonte virtual (círculo vazio).

Para a observação mais detalhada do ponto de vista geométrico, a Figura 2.5 ilustra o fenômeno da reflexão da frente de onda na superfície. O vetor \mathbf{A}_f é o vetor área da superfície da frente de onda no ponto em contato com o anteparo, enquanto o vetor \mathbf{A}_r representa o vetor área da onda que seria gerada pela fonte virtual. Para que a fonte virtual r seja diametralmente oposta à fonte f , tendo o anteparo como

referência, teremos que o vetor \mathbf{A}_r será o próprio \mathbf{A}_f com a componente \hat{x} invertida. Com isso, os vetores ilustrados manterão um mesmo ângulo θ em relação ao eixo \hat{x} . Como o vetor área tem sua magnitude igual ao valor da área da superfície, este vetor crescerá a medida que se afasta da fonte. As perdas de energia na reflexão não irão influenciar nestes vetores, embora reduzam a amplitude do sinal acústico.

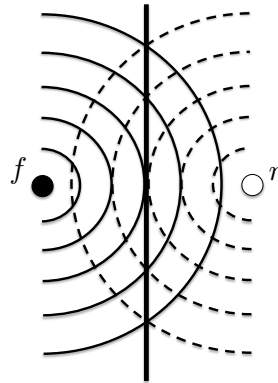


Figura 2.4: Sinal acústico refletindo-se em uma superfície.

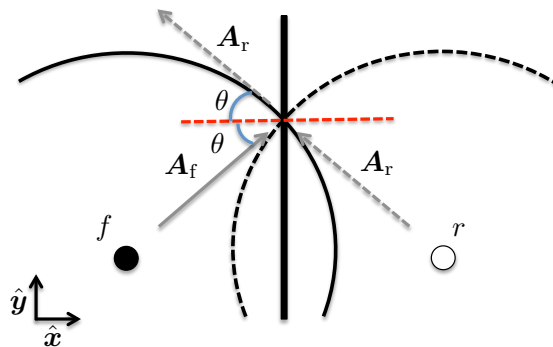


Figura 2.5: Ilustração detalhada de uma frente de onda refletindo-se em superfície.

Quando ondas sonoras se propagam em locais fechados, como salas, o fenômeno da reflexão se torna o principal fator a caracterizar o ambiente do ponto de vista acústico. A próxima subseção aborda uma discussão a esse respeito.

2.2.2 Acústica de Salas

Uma das características acústicas mais relevantes de uma sala é o tempo que um sinal acústico leva para ser atenuado em 60 dB a partir da sua intensidade

original, após cessar sua emissão [9]. A esta medida se dá o nome de RT-60 (tempo de reverberação com decaimento de 60 dB).

Há diversas formas de se medir o RT-60. De maneira geral, pode-se utilizar um sinal capaz de excitar o espectro de frequências por completo, como um sinal pseudobranco, sendo emitido até que se estabeleça um patamar de energia constante para, em seguida, cessá-lo abruptamente. Tendo a energia neste instante como referência, mede-se o tempo necessário para que decaia 60 dB daquele patamar. Em locais projetados para cinema ou concertos e *shows*, o RT-60 é medido por faixa de frequências, pela necessidade de se ter a resposta em frequência o mais próxima de plana.

Um dos modelos utilizados para cálculo teórico de RT-60 foi desenvolvido ao final do século XIX por Wallace C. Sabine, um acústico conhecido até os dias atuais pelas suas contribuições [10]. Seu modelo considera que a sala é formada por um somatório de superfícies com área S [m²], com fator α de absorção, onde $0 \leq \alpha \leq 1$. Para $\alpha = 1$, a superfície é totalmente absorviva; portanto, toda energia incidente é acumulada. Para $\alpha = 0$, pode-se afirmar que nenhuma energia será acumulada no material e, posteriormente, será devolvida para o ambiente, por ter-se refletido na superfície que o sucede. O fator α leva em consideração toda a energia perdida entre o sinal incidente e o sinal refletido. Com isso, calcula-se o RT-60 teórico utilizando-se a seguinte expressão:

$$T_{60} = \frac{kV}{A}, \quad (2.3)$$

onde $k = 0,161$ s/m para o ar com temperatura de 20°C, V denota o volume da sala em m³ e $A = \sum_i \alpha_i S_i$, sendo S_i a área da i -ésima superfície com coeficiente de absorção α .

A Figura 2.6 ilustra uma fonte sonora no ponto f e um microfone no ponto m . Utilizaremos a figura de um círculo com uma barra horizontal para representar um microfone. As setas com linha tracejada representam as ondas sonoras se deslocando nas direções indicadas e sendo refletidas nas superfícies, enquanto a seta com linha cheia representa o som se propagando na linha de visada direta.³ Como o som que

³Nesse contexto, chama-se visada direta ao percurso direto do som entre o alto-falante e o microfone. Se houver anteparos entre eles, não haverá visada direta.

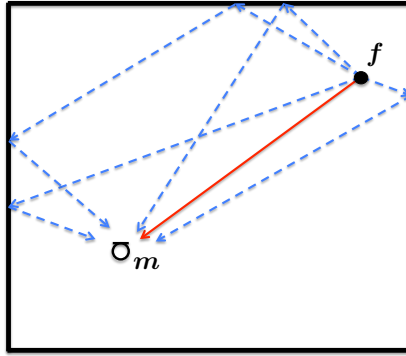


Figura 2.6: Sinal acústico refletindo-se em superfícies de uma sala.

viajou pela linha de visada direta percorreu o menor caminho, este será o primeiro a alcançar o receptor. A este sinal será atribuído o TOF. Todos os outros, provenientes de reflexões, chegarão atrasados e com intensidade reduzida, como exemplificado na Figura 2.7, que ilustra a resposta ao impulso medida na sala onde foram realizados os experimentos práticos.

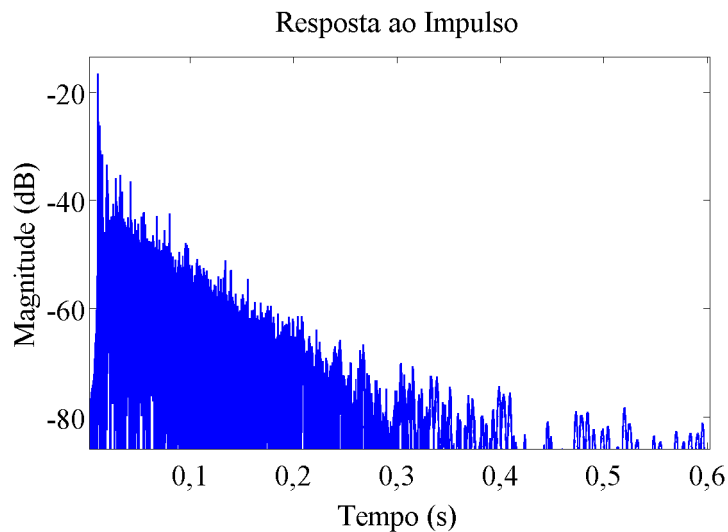


Figura 2.7: Resposta ao impulso de uma sala real.

Podemos modelar salas como sistemas lineares, tendo o sinal reproduzido pela fonte como entrada do sistema e o sinal que chega ao receptor como sinal de saída. Portanto, o sistema realiza apenas uma filtragem no sinal, entregando à saída o sinal de entrada modificado. O fato de as superfícies não refletirem o som de maneira uniforme ao longo do espectro juntamente com a geometria da sala causam grande influência na resposta em frequência do sistema. Por exemplo, paredes paralelas

tendem a originar ondas estacionárias, pelas sucessivas reflexões, originando pontos de máximos e mínimos para determinadas frequências reproduzidas, fenômeno cujo efeito depende das posições onde se encontram a fonte e o receptor. Para contornar estes problemas, é comum o uso de paredes não paralelas, de absorvedores acústicos e de difusores acústicos, que realizam a reflexão da frente de onda de forma a não privilegiar uma direção de propagação. A resposta em frequência da mesma sala real é exibida na Figura 2.8, onde é possível observar os fenômenos citados.

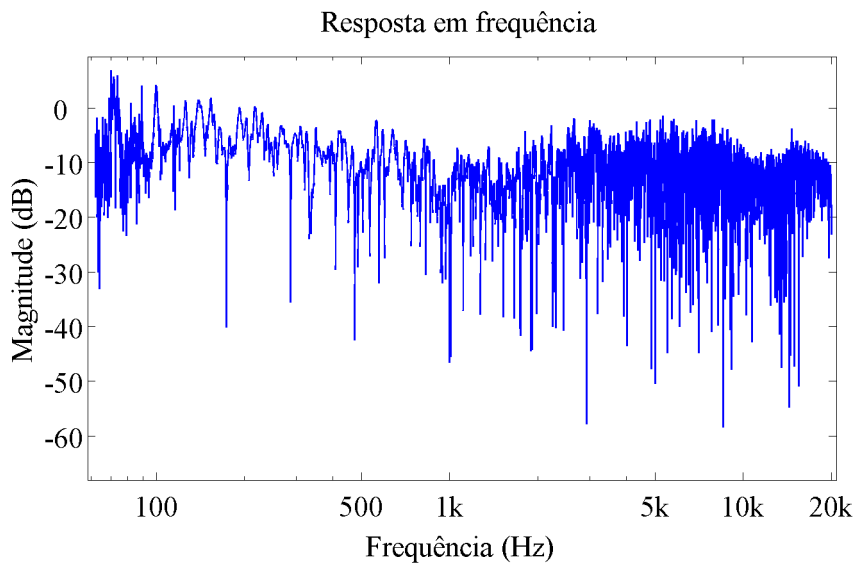


Figura 2.8: Resposta em frequência medida de uma sala real.

2.3 Alto-falantes: Atuadores Acústicos

Alto-falantes são dispositivos cuja função é reproduzir um sinal acústico a partir de um sinal elétrico. A implementação mais comum de alto-falantes faz uso de uma bobina condutora imersa em um campo magnético. Há também outros tipos de alto-falantes, como os dispositivos piezoelétricos, muito utilizados para reprodução em altas frequências, e os capacitivos.

Os alto-falantes de bobina têm como princípio de funcionamento a força gerada em uma carga elétrica quando se movimenta em um campo magnético. A Figura 2.9 ilustra os princípios básicos de funcionamento desses alto-falantes. O sinal que se deseja reproduzir é aplicado à bobina, gerando uma corrente que passa com o sentido indicado pela ilustração. Como a figura mostra um corte em seção, a

bobina está representada pelos círculos. Dentro dos círculos, um ponto representa a corrente saindo do plano, um “x” representa a corrente entrando no plano. Essa bobina, por sua vez, é acoplada à superfície conhecida como “cone”, que será responsável por interagir com o ar. Para que a membrana seja empurrada ou puxada, colocam-se ímãs permanentes afixados ao redor da bobina e em seu interior, de maneira a obter-se um campo magnético como o indicado pelas setas na figura.

Temos que, pelos princípios do Eletromagnetismo [11], para uma corrente i em um condutor imerso em um campo magnético de intensidade \mathbf{B} , a força \mathbf{F} é dada por $\mathbf{F} = i\mathbf{l} \times \mathbf{B}$. Como \mathbf{B} e \mathbf{l} são fixos, esta força gerada será proporcional à corrente e, conseqüentemente, proporcional à tensão aplicada. Quando a corrente é aplicada, é gerada a força \mathbf{F} , conforme indica a seta, realizando o efeito de variação da pressão explicado na Seção 2.1. A estrutura da bobina poderá se deslocar quase sem atrito no interior do ímã, devido a peças mecânicas de alinhamento que evitam que haja contato físico entre eles. Com isso, temos o alto-falante como o atuador desejado, convertendo a tensão aplicada em sinal acústico.

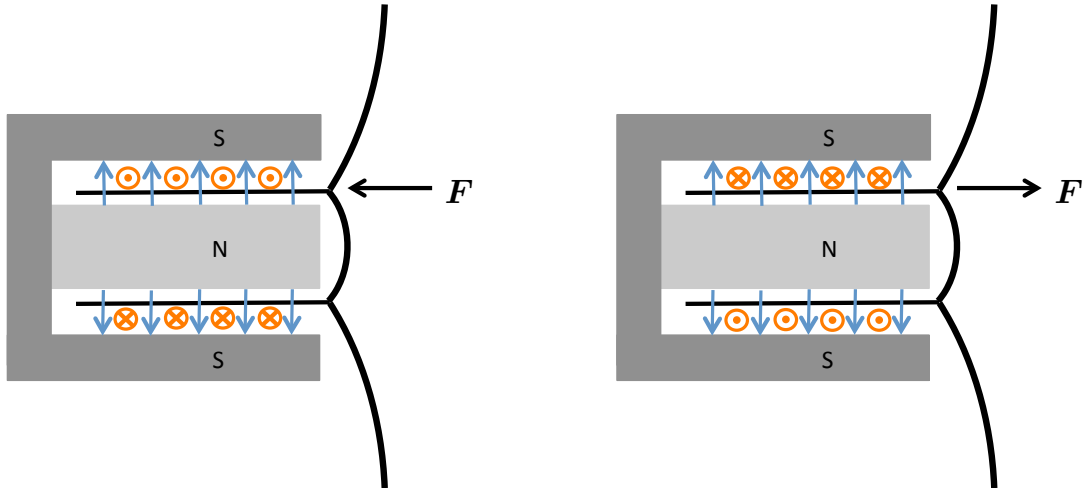


Figura 2.9: Ilustração do princípio de funcionamento de um alto-falante.

Todo alto-falante precisa estar montado em uma caixa, de forma que após a borda da membrana não haja caminho direto para passagem de ar. Caso contrário, haveria um grande deslocamento de ar de um lado da membrana para o outro, pois,

se em um lado a pressão aumenta com o ar sendo empurrado, na face oposta a pressão diminui. Isso modificaria consideravelmente a resposta em frequência dos alto-falantes, principalmente em frequências baixas. As dimensões e configurações espaciais da caixa também irão influenciar na resposta em frequência, o que deve ser levado em consideração em seu projeto.

Para se caracterizar um alto-falante, há três importantes fatores a serem considerados. O primeiro deles é a sua resposta em frequência. Devido a suas características físicas, sempre haverá um comportamento “passa-faixa”, muitas vezes causando a necessidade de se utilizar diferentes alto-falantes simultaneamente, cada um atuando em uma faixa de frequências diferente. A segunda característica é a impedância elétrica, que deve ser compatível com a do amplificador utilizado. Com isso, é possível utilizar alto-falantes combinados em série ou em paralelo, formando a carga necessária para o amplificador e distribuindo a potência pelos alto-falantes combinados. Valores comuns de impedância de alto-falantes têm cerca de unidades ou poucas dezenas de ohms. A terceira característica é a potência elétrica máxima que pode ser aplicada. Caso seja excedida, a membrana irá oscilar fora de sua região linear, causando distorção considerável no sinal reproduzido e podendo romper o filamento da bobina.

Com foco no problema de localização de sensores, é preciso escolher os alto-falantes de modo que atendam a suas especificidades. A primeira delas é a necessidade de tornar emissores e receptores os menores possíveis. Quanto melhor determinadas suas posições espaciais, menor será a incerteza nas medidas. Outro fator importante é a largura de banda reproduzível, o que vai de encontro às dimensões reduzidas. Normalmente, quanto menor o alto-falante, mais estreita é a largura do espectro reproduzido. Como os sinais utilizados para os experimentos têm espectro completo, como no caso de ruído branco aproximado por sequências pseudoaleatórias, quanto menor a largura de banda dos alto-falantes, maior será a degradação do sinal reproduzido, aumentando a correlação entre suas amostras.

Uma característica comum em alto-falantes é a direcionalidade da propagação variável ao longo do espectro. Como a reprodução se dá através do cone, não é possível obter-se uma perfeita propagação esférica. Com isso, a resposta em frequência se modifica ao longo dos ângulos de inclinação do receptor em relação ao eixo de

deslocamento do cone, conforme ilustrado na Figura 2.10. O microfone m_1 recebe o sinal emitido com inclinação nula, enquanto m_2 o está recebendo com um ângulo θ . Por efeito de difração, há uma tendência a ocorrer uma filtragem tipo “passa-baixas” à medida que se aumenta o ângulo θ . Dependendo da posição dos microfones e dos alto-falantes no experimento, o efeito da diretividade dos alto-falantes irá influenciar de forma variada as aquisições obtidas. Este, porém, não é um problema que possa ser facilmente contornado com a escolha do alto-falante.

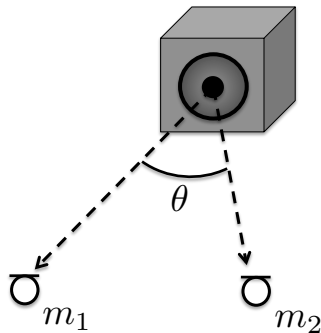


Figura 2.10: Ilustração da direcionalidade de um alto-falante.

2.4 Microfones: Sensores Acústicos

2.4.1 Tipos e Princípios Básicos

Microfones são transdutores construídos com o objetivo de se converter sinais acústicos em sinais elétricos correspondentes, possibilitando o processamento, registro ou transmissão destes sinais. Há várias formas de se construir um microfone. Dentre os tipos mais comuns, podemos citar os microfones dinâmicos, condensadores (capacitores), condensadores de eletreto e *Ribbon* como os mais utilizados na indústria fonográfica [12].

Em linhas gerais, microfones dinâmicos são largamente utilizados pela capacidade de suportarem altas pressões sonoras, terem alta durabilidade e, muitas vezes, baixo custo. Tendo a função dual dos alto-falantes, utilizam o mesmo princípio eletromagnético para gerar um sinal elétrico a partir da variação da posição da bobina imersa em um campo magnético [12]. Por ter seu funcionamento baseado no

movimento de uma membrana acoplada a uma bobina, esse sistema normalmente não é capaz de ter resposta em frequência plana e com banda de passagem larga, tornando-os inadequados para determinadas aplicações. A Figura 2.9 também ilustra o princípio desse tipo de microfone, bastando considerar que a força é gerada pela pressão sonora e a corrente induzida gera uma tensão nos terminais do microfone, sendo este agora o sinal desejado.

Os microfones condensadores, ou de capacitores, proporcionam melhor resposta em frequência e podem ser construídos com dimensões bem reduzidas mantendo sua qualidade de captação, o que justifica seu uso neste projeto. Em de sua qualidade, terão seus princípios de funcionamento detalhados.

Estes microfones são construídos com membranas feitas de um material plástico fino coberto com uma camada de metal, normalmente ouro, com alguns micrômetros de espessura, próxima a uma placa metálica rígida, de forma a configurar um capacitor. O princípio utilizado é o da capacitância, com a seguinte equação:

$$Q = CE, \quad (2.4)$$

onde Q é a carga elétrica acumulada nas placas, C é a capacitância e E é a tensão aplicada aos terminais. A Figura 2.11 ilustra a cápsula deste tipo de microfone. A placa mais fina será exposta a um sinal acústico, enquanto a outra será fixa e disposta atrás desta. Conforme a distância entre as placas varia, há uma variação na capacitância. Conforme as placas se aproximam, a capacitância aumenta, e vice-versa. Com a carga sendo mantida, ocorre uma variação na tensão. Com isso, flui uma corrente pelo resistor no circuito, gerando o sinal elétrico desejado. Posteriormente, haverá um processo de amplificação ainda no circuito interno do microfone.

Este tipo de microfone é consideravelmente mais sensível e acurado que os dinâmicos. Pelo baixo peso da membrana, a resposta em frequência tende a ser mais ampla e plana. Por outro lado, o que o torna mais sensível também o torna mais frágil e caro. Além disso, demanda o uso de um circuito eletrônico e de alimentação externa, de forma a possibilitar a carga da cápsula e a pré-amplificação, necessária devida à baixa corrente obtida.

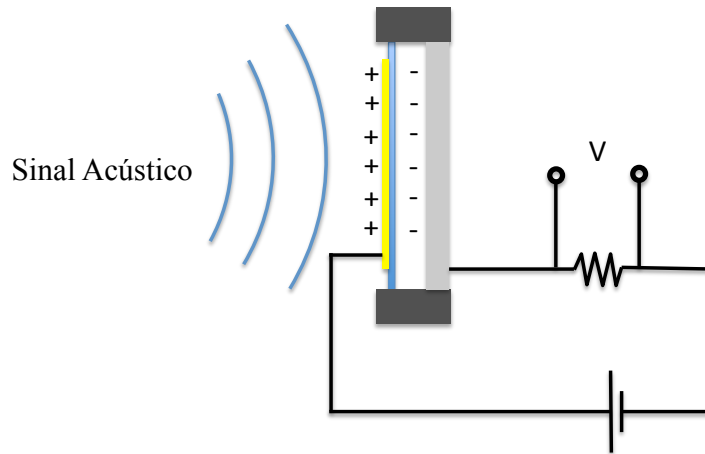


Figura 2.11: Ilustração do princípio de funcionamento de um microfone condensador.

2.4.2 Padrões Polares Básicos

Dentre as características de microfones, podemos destacar duas como principais: a resposta em frequência e o padrão polar. A resposta em frequência é medida com a membrana da cápsula apontada diretamente para a fonte sonora, de forma que a frente de onda interaja frontalmente. Desta forma, tem-se uma estimativa do comportamento do microfone quando captando fontes por incidência direta. O padrão polar mede a resposta em amplitude medida para diversos ângulos de incidência, podendo ser medida por faixas de frequência [12].

Dependendo da aplicação, é desejável que o microfone rejeite sons provenientes de outras direções, como situações onde há grande ruído ambiente, excesso de reverberação ou outras fontes indesejadas. Há diversos padrões polares úteis a estes casos, chamados de padrões cardioides. Para captação de ambientes, onde se deseja captar os sons de forma realista sem privilegiar uma direção específica, ou quando se realizam medições acústicas, é necessário o uso de microfones que capturem da melhor forma possível os sons provindos de qualquer direção. Estes microfones são chamados omnidirecionais ou apenas omni.

Tanto para microfones dinâmicos como para condensadores, os padrões polares podem ser formados pela soma ou subtração dos sinais de duas cápsulas ou pelo projeto do formato da cápsula. Havendo comunicação de ar, portanto acústica, entre a parte interna da membrana e o exterior da cápsula, haverá influência da

frente de onda nos dois lados da membrana. Como este percurso impõe impedância acústica⁴ variável com os ângulos de incidência, haverá, conseqüentemente, um diferente cancelamento do sinal na cápsula. É possível, por exemplo, construir-se a cápsula de forma que o som originado da parte traseira do microfone chegue nos dois lados da membrana com iguais atraso e atenuação. Com isso, esta cápsula terá uma grande rejeição na captação de sinais vindos desta direção. Configurando-se diferentemente a atenuação e o atraso do sinal de cancelamento, formam-se diversos padrões cardioides, como os ilustrados na Figura 2.12.

Microfones condensadores podem ser construídos com duas membranas dispostas na mesma cápsula, possibilitando a captação direta de fontes sonoras dispostas nas partes dianteira e traseira. Se os sinais gerados nestas placas forem somados, a resposta obtida irá contemplar mais direções de chegada, podendo configurar o padrão omnidirecional. Quando há uma frente de onda incidindo lateralmente, a pressão será igualmente exercida nas duas membranas. Se o sinal obtido de uma das membranas for invertido, haverá um cancelamento dos sinais igualmente captados por ambas, privilegiando as incidências frontal e traseira, configurando o padrão bidirecional ou “figura oito”.

A Tabela 2.1 resume algumas das características dos padrões. Nela, estão definidas as equações polares correspondentes; o arco em que o som é captado com atenuação menor que 3 dB, que pode ser interpretado como banda de passagem angular; a eficiência aleatória; e o índice de diretividade. A equação polar dos padrões é definida como $\rho = a + b \cos \theta$, onde $\rho \in \mathbb{R}$ é o ganho de captação e os coeficientes $a \in \mathbb{R}_+$ e $b \in \mathbb{R}_+$ somam 1.

A eficiência aleatória (EA) é definida como a medida de diretividade do eixo frontal do microfone em relação à sua resposta a sons originários de todas as direções, ou seja, o nível de rejeição médio a sons com incidência indireta. Por exemplo, supondo que haja um sinal de voz sendo captado frontalmente pelo microfone, o sinal reverberado pela sala incidirá no microfone por ângulos aleatórios. Se o microfone tiver uma eficiência aleatória de 1/3, o sinal da reverberação será captado com 1/3 da sensibilidade dos sons provenientes do eixo principal, que é o sinal de voz. A

⁴Impedância acústica está relacionada à perda de energia do som ao se propagar no meio. Quanto maior a impedância acústica do meio, mais rapidamente o sinal acústico é atenuado.

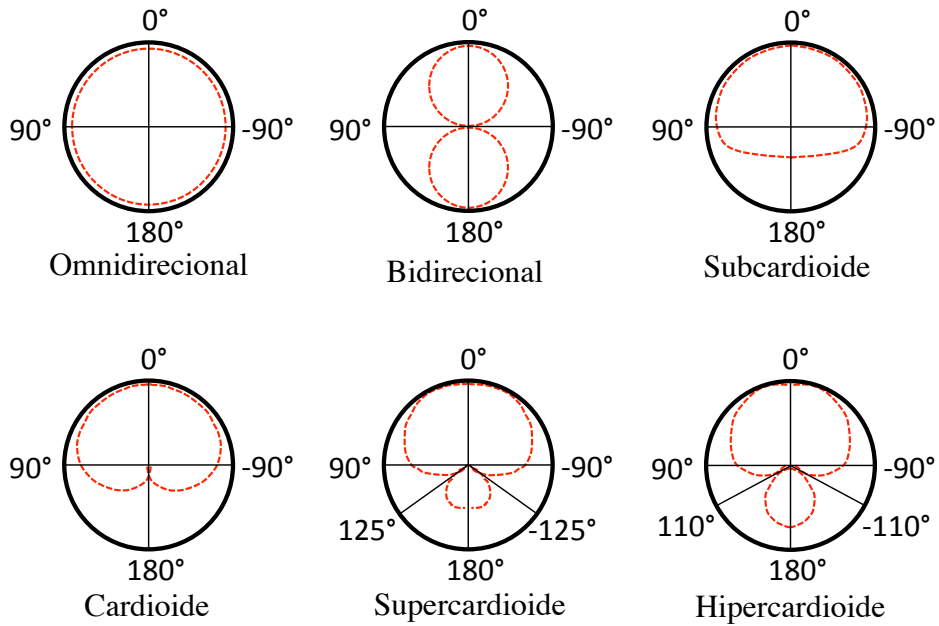


Figura 2.12: Padrões polares mais comuns de microfones.

eficiência aleatória é definida como $EA = 1 - 2b + 4b^2/3$ [13].

O índice de diretividade (ID) é a forma logarítmica da eficiência aleatória, $ID = 10 \log_{10} EA$. A partir desta tabela, é possível observar as razões para os nomes dos padrões. O padrão supercardioide tem maior rejeição lateral que o cardioide, enquanto o hipercardioide maximiza esta característica, sendo o padrão com maior ID [12].

Tabela 2.1: Resumo dos padrões polares [12].

Padrão polar	Equação	Arco de captação (3 dB)	EA	ID
Omnidirecional	$1 + 0 \cos \theta$	360°	1,00	0,0 dB
Bidirecional	$0 + 1 \cos \theta$	90°	0,33	4,8 dB
Subcardioide	$0,7 + 3 \cos \theta$	180°	0,55	2,5 dB
Cardioide	$0,5 + 0,5 \cos \theta$	131°	0,33	4,8 dB
Supercardioide	$0,37 + 0,63 \cos \theta$	115°	0,27	5,7 dB
Hipercardioide	$0,25 + 0,75 \cos \theta$	105°	0,25	6,0 dB

Apropriando-se dos conceitos abordados neste capítulo, o leitor será capaz

de compreender melhor os métodos apresentados e contextualizá-los nas aplicações práticas. Estes conceitos também serão de fundamental importância para a escolha dos equipamentos a serem utilizados e do posicionamento dos alto-falantes no ambiente de teste.

Capítulo 3

Localização de Sensores Acústicos

Neste capítulo, serão discutidos os métodos de localização de sensores acústicos presentes na literatura utilizada como referência para o trabalho, assim como um novo método proposto [8]. Nesse sentido, a implementação realizada é uma adaptação da literatura.

3.1 Definições Preliminares

Considerando o ar como meio homogêneo e linear, pode-se considerar que o som se propaga com velocidade constante $c \in \mathbb{R}_+$. Esta velocidade depende de condições intrínsecas do ar, como temperatura, umidade e pressão. Neste trabalho, assumiremos que a velocidade de propagação c é conhecida. Com isso, teremos uma relação direta entre a distância e o tempo de viagem (TOF) do estímulo sonoro do emissor para o receptor.

Matematicamente, definindo a posição do m -ésimo microfone no espaço 3-D como $\mathbf{m}_m \in \mathbb{R}^{3 \times 1}$, com $m \in \mathcal{M} \triangleq \{1, 2, \dots, M\}$, onde $M \in \mathbb{N}$ denota o número de microfones, e considerando a posição da s -ésima fonte sonora no espaço 3-D como $\mathbf{s}_s \in \mathbb{R}^{3 \times 1}$, com $s \in \mathcal{S} \triangleq \{1, 2, \dots, S\}$, onde $S \in \mathbb{N}$ denota o número de fontes sonoras, podemos definir o TOF, $t_{ms} \in \mathbb{R}_+$, entre o microfone m e a fonte s como

$$t_{ms} \triangleq \frac{\|\mathbf{m}_m - \mathbf{s}_s\|}{c}, \quad (3.1)$$

onde $\|\cdot\| : \mathbb{R}^{3 \times 1} \rightarrow \mathbb{R}_+$ representa a norma euclidiana de vetores.

Note que, na situação em que as posições dos alto-falantes são perfeitamente determinadas — fontes pontuais — e conhecidas, o valor do TOF irá determinar uma

esfera de raio ct_{ms} , centrada em \mathbf{s}_s e de cuja superfície todos os pontos são possíveis candidatos à posição \mathbf{m}_m . Este fato é decorrente da propagação esférica da onda sonora no ar. Como há infinitos possíveis valores, determina-se \mathbf{m}_m utilizando-se mais alto-falantes, como será visto na Seção 3.2.

3.2 Localização por Sistemas Ideais

Um sistema de localização ideal é aquele em que todos os TOFs e todas as posições dos alto-falantes são perfeitamente conhecidas. Assim, dados os TOFs t_{ms_1} e t_{ms_2} entre o m -ésimo microfone e dois alto-falantes distintos de índices $s_1, s_2 \in \mathcal{S}$, respectivamente, podemos relacioná-los da seguinte forma (veja a Eq. (3.1)):

$$\begin{aligned} (t_{ms_1}c)^2 - (t_{ms_2}c)^2 &= \|\mathbf{m}_m - \mathbf{s}_{s_1}\|^2 - \|\mathbf{m}_m - \mathbf{s}_{s_2}\|^2 \\ &= (\|\mathbf{m}_m\|^2 - 2\mathbf{s}_{s_1}^T \mathbf{m}_m + \|\mathbf{s}_{s_1}\|^2) - (\|\mathbf{m}_m\|^2 - 2\mathbf{s}_{s_2}^T \mathbf{m}_m + \|\mathbf{s}_{s_2}\|^2) \\ &= 2(\mathbf{s}_{s_2} - \mathbf{s}_{s_1})^T \mathbf{m}_m + (\|\mathbf{s}_{s_1}\|^2 - \|\mathbf{s}_{s_2}\|^2), \end{aligned} \quad (3.2)$$

chegando em

$$\begin{aligned} (\mathbf{s}_{s_2} - \mathbf{s}_{s_1})^T \mathbf{m}_m &= \frac{\|\mathbf{s}_{s_2}\|^2 - \|\mathbf{s}_{s_1}\|^2 - (t_{ms_2}^2 - t_{ms_1}^2)c^2}{2} \\ &\triangleq b_{s_2, s_1}^m. \end{aligned} \quad (3.3)$$

Com isso, para um determinado microfone indexado por m , cada par de alto-falantes fornecerá uma equação contendo três incógnitas, sendo estas as componentes do vetor \mathbf{m}_m . Como há S alto-falantes, então existirão $\frac{S(S-1)}{2}$ diferentes pares, o que significa que haverá $\frac{S(S-1)}{2}$ diferentes equações similares à Eq. (3.3). Agrupando tais equações numa única forma matricial, tem-se

$$\underbrace{\begin{bmatrix} (\mathbf{s}_1 - \mathbf{s}_2)^T \\ (\mathbf{s}_1 - \mathbf{s}_3)^T \\ \vdots \\ (\mathbf{s}_{S-1} - \mathbf{s}_S)^T \end{bmatrix}}_{\triangleq \mathbf{S} \in \mathbb{R}^{[S(S-1)/2] \times 3}} \mathbf{m}_m = \underbrace{\begin{bmatrix} b_{1,2}^m \\ b_{1,3}^m \\ \vdots \\ b_{S-1,S}^m \end{bmatrix}}_{\triangleq \mathbf{b}_m \in \mathbb{R}^{[S(S-1)/2] \times 1}} \quad (3.4)$$

onde \mathbf{S} é uma matriz conhecida contendo, em cada linha, a diferença entre as posições das fontes que constituem cada par e \mathbf{b}_m é um vetor também conhecido que agrega

a informação dos TOFs correspondentes ao m -ésimo microfone e respectivos alto-falantes.

Para que \mathbf{S} tenha posto completo, é necessário haver, pelo menos, 3 linhas ou colunas da matriz linearmente independentes. Como há $\frac{S(S-1)}{2}$ diferentes pares de alto-falantes, será necessário um mínimo de 4 alto-falantes para satisfazer a condição de posto completo de \mathbf{S} . Desta forma, \mathbf{S} terá 6 linhas e 3 colunas, sendo que das 6 linhas, apenas 3 serão linearmente independente das demais. Note que, ainda com $S = 4$, é possível que \mathbf{S} não tenha posto completo, o que dependerá da escolha das posições dos alto-falantes.

Com essa condição sendo atendida, é possível calcular a posição do m -ésimo microfone perfeitamente utilizando-se a pseudo-inversa $\mathbf{S}^\dagger \in \mathbb{R}^{3 \times [S(S-1)/2]}$, da seguinte forma:

$$\begin{aligned} \mathbf{m}_m &= (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T \\ &\triangleq \mathbf{S}^\dagger \mathbf{b}_m. \end{aligned} \tag{3.5}$$

Note que, pela Eq. (3.4), de fato temos que $\mathbf{S}^\dagger \mathbf{b}_m = \mathbf{S}^\dagger (\mathbf{S} \mathbf{m}_m) = (\mathbf{S}^T \mathbf{S})^{-1} (\mathbf{S}^T \mathbf{S}) \mathbf{m}_m = \mathbf{I}_3 \mathbf{m}_m = \mathbf{m}_m$.

Apesar de não haver a necessidade de mais do que 4 alto-falantes para que \mathbf{m}_m seja determinado no modelo ideal, temos de considerar que, na prática, haverá erros em todas as medidas. As fontes não serão pontuais e com propagação esférica, os sensores também não serão pontuais e os TOFs não serão perfeitamente conhecidos. Nessas condições, havendo um número maior de alto-falantes, o sistema terá mais graus de liberdade para realizar a estimação de \mathbf{m}_m .

Para a análise do sistema em condições reais, iremos considerar que as equações são aproximações, não havendo, portanto, igualdade estrita. Para que o equacionamento traduza tais condições, será preciso substituir esses valores pelas suas estimativas correspondentes. Essas adaptações serão descritas na Seção 3.3

3.3 Localização em Sistemas Reais

Quando implementado na prática, o sistema estará sujeito aos diversos fatores acústicos descritos na Seção 2. Conforme discutido na Subseção 2.2.2, a resposta ao impulso de uma sala é constituída por um primeiro pico, relativo ao sinal de

visada direta, e por subseqüentes réplicas suas atrasadas, atenuadas e deformadas, provenientes das reflexões da frente de onda nos diversos anteparos do ambiente. O efeito da reverberação das salas é um dos principais responsáveis por erros nas estimativas dos TOFs, juntamente com o ruído ambiente.

As características dos alto-falantes e microfones utilizados também podem surtir grande efeito na qualidade da estimação dos TOFs. Por exemplo, um microfone excessivamente direcional pode privilegiar uma frente de onda refletida em detrimento da onda em visada direta, ou alto-falantes podem ter uma resposta em frequência não plana, deteriorando o sinal reproduzido.

Conforme discutido anteriormente, a acurácia na estimação da posição dos microfones está diretamente ligada à acurácia na estimação dos TOFs. Os efeitos destes fatores serão discutidos com mais detalhes a seguir na Subseção 3.3.1, onde será descrito o processo de estimação dos TOFs.

3.3.1 Estimação de TOF

Para que sejam estimados os TOFs, assume-se que o sinal emitido pelos alto-falantes é conhecido a priori. Caso sejam utilizados dispositivos independentes do nó central, tais dispositivos já terão esse sinal pré-armazenado. Com isso, é possível comparar o sinal original com o capturado pelo microfone, de forma a buscar um atraso relativo que melhor os alinhe no tempo. A este atraso, será atribuído o TOF.

Um método computacionalmente simples e eficiente para realizar esta tarefa é o método de correlação-cruzada (CC). Este método irá fornecer a projeção do sinal gravado no sinal emitido deslocado no tempo. Quanto melhor for o alinhamento, mais alto será o valor resultante da projeção para aquele atraso. Considerando-se que o sinal originalmente emitido tenha uma autocorrelação baixa para atrasos não-nulos, então é natural que associemos o principal candidato a TOF ao valor máximo da função de correlação-cruzada entre os sinais emitido e gravado.

Como todo o processamento será realizado no domínio digital, consideremos que $F \in \mathbb{R}_+$ seja a taxa de amostragem do sistema, $x_s[n]$ seja o sinal emitido pelo s -ésimo alto-falante e $y_m[n]$ seja o sinal capturado pelo m -ésimo microfone. A função

de CC correspondente $\mathcal{R}_{ms}[\cdot] : \mathbb{N} \rightarrow \mathbb{R}_+$ será definida por

$$\mathcal{R}_{ms}[k] \triangleq \sum_{n \in \mathbb{N}} y_m[n] x_s[n+k]. \quad (3.6)$$

Note que, para k positivo, a CC calculada é relativa a um atraso de $x_s[n]$, enquanto que, para valores de k negativos, a CC será relativa a um sinal $x_s[n]$ sendo adiantado. Como os TOFs são sempre atrasos, apenas serão calculados valores positivos de k nas CCs em estimações de TOFs.

A CC pode ser calculada de forma mais eficiente no domínio da frequência utilizando-se algoritmos de transformada rápida de Fourier (FFT). Considerando que $y_m[n]$ e $x_s[n]$ tenham, respectivamente, N e M amostras, deverá ser realizado um *zero padding* de forma que o comprimento de ambos se torne $M + N - 1$. Com o *zero padding*, garante-se a equivalência entre as correlações realizadas de forma circular e de forma linear. O processo é dado como

$$\mathcal{R}_{ms}[k] = \text{iFFT}(\text{FFT}(\tilde{y}_m[n]) \circ \text{FFT}(\tilde{x}_s[n])^*), \quad (3.7)$$

onde \circ denota o produto de Hadamard¹ dos vetores e $\tilde{y}_m[n]$ e $\tilde{x}_s[n]$ denotam os sinais $y_m[n]$ e $x_s[n]$ com *zero padding*.

Com isso, pode-se definir a estimativa de t_{ms} como

$$\hat{t}_{ms} \triangleq \left(\arg \max_{k \in \mathbb{Z}} \{ \mathcal{R}_{ms}[k] \} \right) / F. \quad (3.8)$$

Convém notar que, mesmo que não haja erro por interferências sonoras ambientais ou de características inerentes aos equipamentos, aos TOFs serão atribuídos valores discretos de tempo, uma vez que são calculados a partir de sinais também discretos. Quanto maior a taxa de amostragem utilizada, menor será o erro de quantização temporal; porém, maior será o custo computacional.

¹O produto de Hadamard de duas matrizes do mesmo tipo, $\mathbf{a}_{m \times n}$ e $\mathbf{b}_{m \times n}$, é definido por uma matriz $\mathbf{c}_{m \times n}$, onde cada elemento $\mathbf{c}_{i \times j}$ é o produto dos elementos $\mathbf{a}_{i \times j}$ e $\mathbf{b}_{i \times j}$.

3.3.2 Método de Mínimos Quadrados

Considerando agora que os TOFs medidos são estimativas possivelmente ruidosas dos TOFs verdadeiros, devemos adaptar a Eq. (3.4) da seguinte forma:

$$\begin{aligned} \mathbf{S}\mathbf{m}_m &\approx \underbrace{\begin{bmatrix} \hat{b}_{1,2}^m \\ \hat{b}_{1,3}^m \\ \vdots \\ \hat{b}_{S-1,S}^m \end{bmatrix}}_{\triangleq \hat{\mathbf{b}}_m \in \mathbb{R}^{[S(S-1)/2] \times 1}} \\ &\approx \hat{\mathbf{b}}_m, \end{aligned} \quad (3.9)$$

onde, para cada par $(s_2, s_1) \in \mathcal{S}^2$ tal que $s_2 \neq s_1$, o elemento \hat{b}_{s_2,s_1}^m é dado por

$$\hat{b}_{s_2,s_1}^m \triangleq \frac{\|\mathbf{s}_{s_2}\|^2 - \|\mathbf{s}_{s_1}\|^2 - (\hat{t}_{ms_2}^2 - \hat{t}_{ms_1}^2)c^2}{2}. \quad (3.10)$$

Note que, adaptando a expressão para os TOFs estimados, devemos considerar as relações como aproximações, não havendo mais garantia de igualdade. Por isso, utiliza-se “ \approx ” em (3.9).

Para determinar uma posição de microfone que melhor aproxime a expressão em (3.9), adota-se uma solução conhecida como método de mínimos quadrados (do inglês, *least-squares*) dada, nesse caso, por

$$\hat{\mathbf{m}}_m^{\text{LS}} \triangleq \arg \min_{\mathbf{m} \in \mathbb{R}^{3 \times 1}} \{\|\mathbf{S}\mathbf{m} - \hat{\mathbf{b}}_m\|\}, \quad (3.11)$$

cuja solução é

$$\hat{\mathbf{m}}_m^{\text{LS}} = \mathbf{S}^\dagger \hat{\mathbf{b}}_m. \quad (3.12)$$

Com essa solução em mãos, iremos considerar os erros de estimação como sendo

$$\Delta t_{ms} \triangleq \hat{t}_{ms} - t_{ms} \quad (3.13)$$

$$\begin{aligned} \Delta \mathbf{b}_m &\triangleq \hat{\mathbf{b}}_m - \mathbf{b}_m \\ &= [\Delta b_{1,2}^m \ \Delta b_{1,2}^m \ \cdots \ \Delta b_{S-1,S}^m]^T, \end{aligned} \quad (3.14)$$

possibilitando escrever a expressão (3.9) como

$$\mathbf{S}\mathbf{m}_m \approx \mathbf{b}_m + \Delta \mathbf{b}_m, \quad (3.15)$$

onde

$$\Delta b_{s_2, s_1}^m \triangleq -\frac{c^2}{2} [(\Delta t_{ms_2}^2 - \Delta t_{ms_1}^2) + 2(t_{ms_2} \Delta t_{ms_2} - t_{ms_1} \Delta t_{ms_1})]. \quad (3.16)$$

Assim, temos que (ver Eq. (3.12))

$$\begin{aligned} \hat{\mathbf{m}}_m^{\text{LS}} &= \mathbf{S}^\dagger \left(\underbrace{\mathbf{b}_m}_{=\mathbf{S}\mathbf{m}_m} + \Delta \mathbf{b}_m \right) \\ &= \mathbf{m}_m + \mathbf{S}^\dagger \Delta \mathbf{b}_m \\ &\triangleq \mathbf{m}_m + \Delta \mathbf{m}_m^{\text{LS}}. \end{aligned} \quad (3.17)$$

É importante destacar que, por não haver conhecimento de \mathbf{m}_m , não há como se ter acesso direto ao vetor $\Delta \mathbf{b}_m$. Portanto, só nos resta definir adequadamente as posições das caixas de forma a obter uma matriz \mathbf{S}^\dagger que não amplifique em demasia os erros modelados por $\Delta \mathbf{b}_m$.

Até este ponto, as posições dos alto-falantes haviam sido consideradas perfeitamente conhecidas, o que não é verdade, uma vez que são medidas manualmente e alto-falantes não são propriamente fontes pontuais.

Conforme descrito no artigo original [2], uma boa forma de se posicionar os alto-falantes é distribuí-los ao redor dos microfones, de forma a circunscrevê-los. Com isso, os erros de medição das posições dos alto-falantes tendem a ser menores.

3.3.3 Método de Otimização Não-Linear

Na implementação original do sistema [2], após ter sido obtida uma estimativa inicial das posições dos microfones utilizando o método LS, há uma etapa de refinamento das posições por meio da minimização de uma função-custo onde há dois termos principais: os TOFs medidos com o método de CC, já definidos como \hat{t}_{ms} , e as expressões teóricas dos TOFs em função das posições \mathbf{m}_m e \mathbf{s}_s . Com esses parâmetros, é possível refinar as posições calculadas resolvendo o seguinte problema de otimização:

$$(\hat{\mathbf{m}}_1^{\text{NL}}, \dots, \hat{\mathbf{m}}_M^{\text{NL}}, \hat{\mathbf{s}}_1^{\text{NL}}, \dots, \hat{\mathbf{s}}_S^{\text{NL}}) \triangleq \arg \min_{\substack{(\mathbf{m}_m, \mathbf{s}_s) \in \mathbb{R}^{3 \times 1} \times \mathbb{R}^{3 \times 1} \\ (m, s) \in \mathcal{M} \times \mathcal{S}}} \left\{ \sum_{m \in \mathcal{M}} \sum_{s \in \mathcal{S}} \left(\hat{t}_{ms} - \frac{\|\mathbf{m}_m - \mathbf{s}_s\|}{c} \right)^2 \right\}. \quad (3.18)$$

As posições $\hat{\mathbf{m}}_1^{\text{LS}}, \dots, \hat{\mathbf{m}}_M^{\text{LS}}$, calculadas na primeira etapa, são utilizadas como inicialização no método de otimização que utilizará a função-custo (3.18) como base para o cálculo.

Já $\mathbf{s}_1, \dots, \mathbf{s}_S$, na implementação original do sistema, não são conhecidas a priori. Por esta razão, é necessário localizar os alto-falantes numa primeira etapa, para, posteriormente, localizar os microfones. Incluir as posições dos alto-falantes na função-custo tem como intuito melhorar essas estimativas. Assim, a etapa de otimização não-linear cumpre um papel imprescindível no sistema proposto em [2]. Porém, note que a função-custo parte da ideia de que \hat{t}_{ms} foi estimado com boa precisão a ponto de se poder tomá-lo como referência.

De uma forma geral, foi observado que, quando há conhecimento a priori de $\mathbf{s}_1, \dots, \mathbf{s}_S$ com boa precisão e as condições ambientais — ruído e reverberação — não são severas, a etapa de estimação pelo método LS já é capaz de fornecer bons resultados. Mas, considerando um caso em que haja um par microfone-caixa (m, s) com \hat{t}_{ms} mal estimado, o correspondente erro de estimação irá se refletir em $\hat{\mathbf{m}}_m^{\text{LS}}$.² Com isso, possivelmente, os tempos de percurso relativos às distâncias de $\hat{\mathbf{m}}_m^{\text{LS}}$ até os outros alto-falantes serão diferentes dos TOFs estimados. Neste caso, a função de otimização não-linear, provavelmente, traria melhoras às estimativas; porém, os métodos de otimização não-linear, em geral, têm um elevado custo computacional, se comparados com o método LS, o que pode tornar o sistema inviável em situações onde há baixa capacidade de processamento. Além disso, a função-custo utiliza a estimativa de posição de todos os microfones, o que pode ser um fator complicador em casos onde são utilizados dispositivos independentes de uma central de processamento.

Como em nossa implementação o método LS fornece resultados satisfatórios para o caso onde há estimativas precisas de TOF, identificar e corrigir as estimativas problemáticas utilizando rotinas que demandem baixo custo computacional pode ser a chave para contornar a necessidade do uso de uma etapa de otimização não-linear. Com isso, são eliminadas as limitações associadas a essa etapa, como a grande complexidade computacional e a dependência de todas as estimativas. A próxima

²O impacto do erro de estimação de um TOF na posição $\hat{\mathbf{m}}_m^{\text{LS}}$ será função não somente deste erro, mas, também, do número de alto-falantes envolvidos na estimação.

seção descreve um método proposto com esse intuito.

3.3.4 Método de Seleção de TOFs [8]

Vimos na Seção 2.2 que no fenômeno da reflexão a potência da onda incidente é sempre maior que a potência da onda refletida, pois parte da energia incidente será convertida em energia térmica e outra parte será transmitida através do material. Além disso, quanto maior a distância percorrida pelo sinal acústico, menor será sua intensidade, que decai com o quadrado da distância conforme descrito na Eq. (2.2). Como decorrência desses fatos, a resposta ao impulso de uma sala será, em geral, caracterizada por um primeiro pico relacionado com o sinal de visada direta, cujo atraso é associado ao TOF, e por outros picos mais atrasados e atenuados, relacionados com reflexões na sala, conforme ilustrado na Figura 2.7.

Entretanto, há situações em que esse padrão não é observado. Dentre as principais causas para isso, podemos citar: (i) a interferência de sinais refletidos no sinal de visada direta, o que pode gerar picos secundários ligeiramente maiores que o principal; (ii) a diretividade dos alto-falantes e microfones fazendo com que sejam privilegiados sinais refletidos em detrimento do sinal direto; (iii) perda de potência no sinal de visada direta pela existência de obstáculos e (iv) SNR excessivamente baixa.

Como o método original de estimação do TOF se resume a associar ao TOF o atraso relacionado ao maior pico da CC, este processo resultará em erros quando ocorrer alguma das situações descritas. Com o intuito de tornar o processo de estimação mais robusto a esses possíveis problemas, foi proposto um novo método [8] não-linear que identifica as medidas de TOF com erros e as corrige automaticamente.

Para isso, levamos em conta o fato de que os fatores citados (i), (ii) e (iii) dizem respeito apenas à amplitude dos picos observados na CC, não aos atrasos correspondentes. Em outras palavras, por mais que estes efeitos tornem o pico principal menor que outros, este ainda continuará sendo o primeiro pico, sob o ponto de vista da ordenação temporal.

Porém, isso não continua sendo verdade para baixas SNRs, em que o pico principal se confunde com o chão de ruído. A Figura 3.1 ilustra a CC entre um sinal pseudoaleatório com 1000 amostras e o mesmo sinal somado com ruído. Observando

as figuras, constata-se que, conforme a SNR diminui, a diferença entre o pico principal e o chão de ruído também diminui, ou seja, mais difícil se torna determinar o TOF nesta situação. Contudo, melhorar a imunidade do sistema ao ruído é uma tarefa relativamente mais simples. Isso ficará mais claro no Capítulo 4, onde serão discutidas as características dos sinais utilizados.

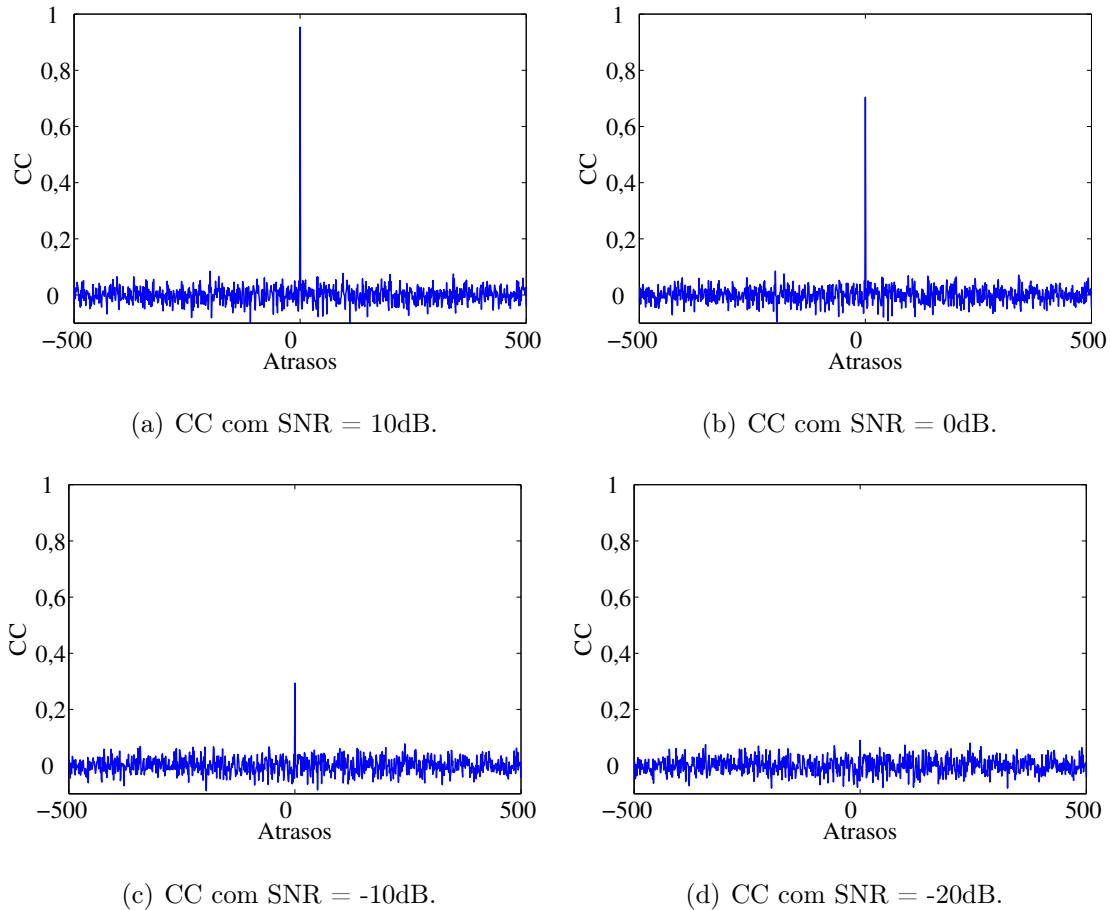
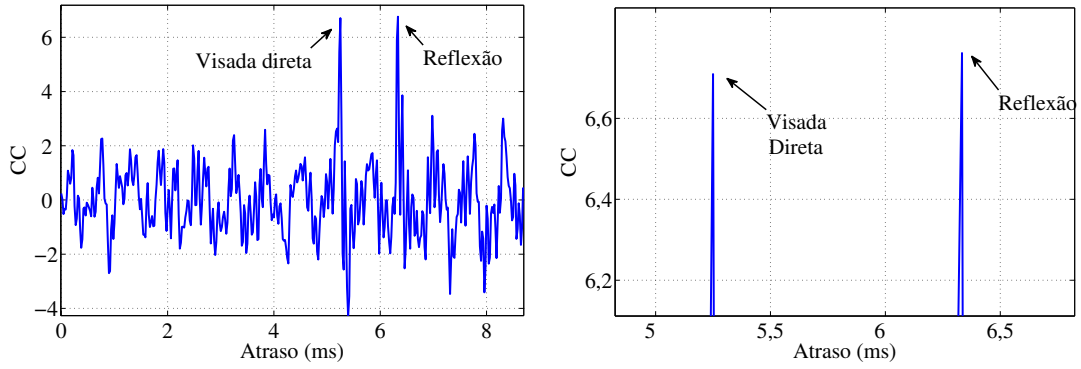


Figura 3.1: Exemplo da influência do ruído aditivo na correlação cruzada.

Considerando o caso em que a SNR é suficientemente alta, tomemos como exemplo a CC ilustrada na Figura 3.2. As imagens evidenciam um caso em que o pico referente a uma reflexão é representado com valor maior que o pico referente ao sinal de visada direta. Neste exemplo, podemos observar que a diferença entre os valores dos picos é consideravelmente menor que a diferença entre um deles e qualquer outro pico da CC. Observando outras ocorrências deste fenômeno, podemos constatar que este fato é recorrente, nos permitindo afirmar que, se há um erro de estimação na CC em questão, o pico candidato de maior relevância será, com grande probabilidade, o que tiver a amplitude mais próxima da do maior pico.



(a) Exemplo de CC de um sinal gravado.

(b) Detalhe dos picos.

Figura 3.2: Exemplo de CC de um sinal gravado evidenciando uma situação onde o maior pico não está relacionado ao sinal de visada direta. Estas figuras foram adaptadas de [8].

Sendo assim, a primeira etapa do método de seleção de TOFs proposto consiste em calcular os TOFs da maneira descrita na Seção 3.3.1, buscando os atrasos relacionados com os valores máximos das CCs. Estes TOFs serão denominados \hat{t}_s e utilizados para formar o vetor $\hat{\mathbf{b}}$, como na Eq. (3.10). Em seguida, é feita uma estimação inicial da posição dos microfones pelo método de LS, resultando em $\hat{\mathbf{m}}^{(0)\text{LS}}$, como na Eq. (3.12). Se denominarmos \bar{t}_s como sendo o TOF referente à posição $\hat{\mathbf{m}}^{(0)\text{LS}}$,

$$\bar{t}_s \triangleq \frac{\|\hat{\mathbf{m}}^{(0)\text{LS}} - \mathbf{s}_s\|}{c}, \quad (3.19)$$

o erro entre a estimativa \hat{t}_s e \bar{t}_s será

$$e_s \triangleq \bar{t}_s - \hat{t}_s. \quad (3.20)$$

Se tivermos os TOFs estimados corretamente, $e_s \approx 0$ para todo $s \in \mathcal{S}$. Caso um ou mais TOFs estejam incorretos, todos os \bar{t}_s serão influenciados pelo erro. Com isso, será necessário realizar uma busca por todos os resultados das CCs, uma vez que não é possível associar os \hat{t}_s mal estimados aos maiores e_s . Portanto, será avaliado o valor médio dos erros $\bar{e} = \frac{1}{S} \sum_{s \in \mathcal{S}} |e_s|$. O algoritmo iniciará o processo de correção dos TOFs se \bar{e} ultrapassar um limiar de erro previamente escolhido $\gamma \in \mathbb{R}_+$. Em suma, se os TOFs estimados não forem suficientemente próximos dos TOFs calculados a partir de $\hat{\mathbf{m}}^{\text{LS}}$ em média, considera-se que há algum TOF mal estimado.

Os TOFs serão trocados um por vez e será calculado \bar{e} para avaliar se a mudança realizada no TOF bastou para corrigir $\hat{\mathbf{m}}^{\text{LS}}$. Se há um \hat{t}_s errado, é razoável afirmar que o TOF verdadeiro é mais curto. Portanto, para corrigir o \hat{t}_s , a primeira etapa consiste em buscar, para todos os alto-falantes, o maior pico dentre os que precedem o pico referente ao \hat{t}_s . Este pico será o candidato a TOF para aquele alto-falante.

Dentre esses picos, o que apresentar maior relevância será eleito como novo \hat{t}_s , permitindo o cálculo de um novo $\hat{\mathbf{m}}^{\text{LS}}$. A métrica utilizada para avaliar a relevância dos picos é a razão entre sua amplitude e a amplitude do pico previamente relacionado ao \hat{t}_s . O procedimento será realizado desta forma enquanto $\bar{e} > \gamma$. Caso não haja mais picos a serem escolhidos, o procedimento é abortado e o valor inicial do LS é mantido. Este, provavelmente, será o resultado no caso de a SNR estar abaixo de um nível crítico.

Note que este processo será feito para cada microfone individualmente. Com isso, uma posição de microfone mal estimada não influenciará nas outras e não haverá a necessidade de um dispositivo com um microfone se comunicar com outro, bastando conhecer o instante em que o sinal foi emitido pelo alto-falante. É importante destacar também que, apesar de este método calcular $\hat{\mathbf{m}}^{\text{LS}}$ diversas vezes, a complexidade computacional para este cálculo é desprezível diante do cálculo das CCs. Assim, obtemos um método robusto a efeitos da reverberação de ambientes com baixa complexidade computacional. O método proposto está resumido em forma de pseudolinguagem no Algoritmo 1 [8].

Algoritmo 1 Seleção de TOF.

Defina um limiar $\gamma \in \mathbb{R}_+$

para todo $s \in \mathcal{S}$ **faça**

$$\hat{k}_s \leftarrow \arg \max_{k \in \mathbb{N}} \{R_s[k]\}$$

$$\hat{t}_s \leftarrow R_s[\hat{k}_s]/F$$

fim para

Defina \mathbf{S} e \mathbf{b} usando a Eq. (3.4)

$$\hat{\mathbf{m}}^{(0)\text{LS}} \leftarrow \mathbf{S}^\dagger \hat{\mathbf{b}}$$

$$\hat{\mathbf{m}}^{\text{LS}} \leftarrow \hat{\mathbf{m}}^{(0)\text{LS}}$$

Calcule e_s usando Eq. (3.19) e Eq. (3.20) para todo $s \in \mathcal{S}$

$$\bar{e} \leftarrow \frac{1}{S} \sum_{s=1}^S |e_s|$$

enquanto $\bar{e} > \gamma$ **faça**

$$s^* \leftarrow \emptyset \text{ and } r^* \leftarrow 0$$

para todo $s \in \mathcal{S}$ **faça**

$$\mathcal{P}_s \leftarrow \left\{ k \in \mathbb{N} \mid k < \hat{k}_s, R_s[k] \text{ é um pico} \right\}$$

$$k_s \leftarrow \arg \max_{k \in \mathcal{P}_s} R_s[k]$$

$$r_s \leftarrow R_s[k_s]/R_s[\hat{k}_s]$$

se $r_s > r^*$ **então**

$$r^* \leftarrow r_s$$

$$s^* \leftarrow s$$

fim se

fim para

se $s^* \neq \emptyset$ **então**

$$\hat{t}_{s^*} \leftarrow k_{s^*}/F$$

Atualize $\hat{\mathbf{b}}$ usando Eq. (3.10)

$$\hat{\mathbf{m}}^{\text{LS}} \leftarrow \mathbf{S}^\dagger \hat{\mathbf{b}}$$

Atualize e_{s^*} usando Eq. (3.19) e Eq. (3.20)

$$\bar{e} \leftarrow \frac{1}{S} \sum_{s=1}^S |e_s|$$

senão

Abortar procedimento

$$\hat{\mathbf{m}}^{\text{LS}} \leftarrow \hat{\mathbf{m}}^{(0)\text{LS}}$$

fim se

fim enquanto

Capítulo 4

Experimentos: Procedimentos, Resultados e Discussões

Este capítulo descreve os resultados simulados e experimentais obtidos utilizando-se os métodos de localização de sensores apresentados no Capítulo 3. De forma a abranger diversos possíveis cenários reais, foram realizadas simulações com grande variação dos parâmetros acústicos e espaciais, tais como nível de ruído ambiente, reverberação e tamanho das salas. Foi realizado também um experimento prático, de maneira a avaliar a robustez do método quando submetido a situações reais, em que há influência de características não-ideais intrínsecas aos equipamentos reais.

4.1 Estudo da Ortogonalidade dos Sinais Emitidos

Antes de nos debruçarmos sobre os procedimentos práticos, convém realizarmos um estudo dos sinais emitidos a serem utilizados pelo sistema, de forma a evidenciar as características que os tornam mais ou menos adequados a este serviço.

O método de CC apresentado na Subseção 3.3.1 baseia-se no cálculo da projeção de dois sinais com diferentes valores de atraso relativo. Com base na Eq. (3.6), cada amostra $\mathcal{R}_{ms}[k]$ é a representação em magnitude da projeção de um sinal $y_m[n]$ em $x_s[n+k]$. Quando o valor da projeção é nulo, diz-se que $y_m[n]$ é ortogonal a $x_s[n+k]$.

O sinal ideal para um sistema de localização de sensores é ortogonal a ele mesmo para todo atraso $k \neq 0$. Claramente, quando $k = 0$, o que temos é a

energia do sinal em questão. Utilizaremos o termo “ortogonalidade” para dizer o quão próximo um sinal está de ser ortogonal. Quanto maior é o pico da CC em $k = 0$ em relação aos demais picos, dizemos que o sinal é “mais ortogonal”. Note que as outras amostras podem ter grande magnitude e, ainda assim, não formarem picos, bastando estarem próximas do pico em $k = 0$. Com isso, há um aumento na largura desse pico, piorando a definição do mesmo.

Vale ressaltar novamente que, conforme visto na Subseção 2.2.2, a resposta ao impulso típica de uma sala contém, além do pico referente à visada direta, picos secundários atenuados e atrasados, relativos às reflexões nas superfícies da sala, indesejados para o processo de estimação do TOF. Com isso, utilizando-se sinais ortogonais, cada pico seria otimamente representado, de forma a minimizar a chance de erros pelo chamado “chão de ruído” inerente ao próprio sinal na CC.

O primeiro sinal avaliado será um *chirp* linear, conforme utilizado em [2]. O *chirp* é um sinal senoidal cuja frequência varia linearmente ao longo do tempo, ou seja, a derivada de sua fase não é constante. Sua frequência instantânea é definida como

$$f(t) = f_0 + \left(\frac{f_1 - f_0}{t_1} \right) t, \quad (4.1)$$

onde f_0 é a frequência inicial do *chirp* e f_1 é a frequência no instante t_1 . A fase da senoide é obtida a partir da integração da frequência da seguinte forma:

$$\phi(t) = \phi_0 + 2\pi \int_0^t \left[f_0 + \left(\frac{f_1 - f_0}{t_1} \right) \tau \right] d\tau = \phi_0 + 2\pi \left[f_0 t + \left(\frac{f_1 - f_0}{2t_1} \right) t^2 \right]. \quad (4.2)$$

Assim, pode-se definir o *chirp* linear no tempo como sendo

$$x(t) = \sin \left[\phi_0 + 2\pi \left(f_0 t + \frac{f_1 - f_0}{2t_1} t^2 \right) \right], \quad (4.3)$$

onde ϕ_0 é a fase inicial do *chirp*. Analogamente, pode-se definir o *chirp* linear no domínio do tempo discreto como

$$x[n] = \sin \left[\phi_0 + \frac{2\pi}{F} \left(f_0 n + \frac{f_M - f_0}{2M} n^2 \right) \right], \quad (4.4)$$

onde f_M é a frequência na amostra M e F é a taxa de amostragem utilizada. Portanto, se o sinal tiver comprimento M , o *chirp* irá variar da frequência f_0 à frequência f_M .

A Figura 4.1 exibe seis funções de autocorrelação de diferentes *chirps* com taxa de amostragem $F = 48$ kHz. Mantendo-se constante o número de amostras dos sinais, variamos a largura de banda do sinal, obtendo 5 kHz, 10 kHz e 15 kHz, respectivamente. As Figuras 4.1(a), 4.1(b) e 4.1(c) ilustram os resultados das autocorrelações de *chirps* com $f_0 = 5$ kHz e as larguras de banda citadas. As Figuras 4.1(d), 4.1(e) e 4.1(f) diferem das primeiras por conterem $f_0 = 1$ kHz, porém, com as mesmas larguras de banda, de forma que os resultados dos sinais de mesma banda estão dispostos um acima do outro.

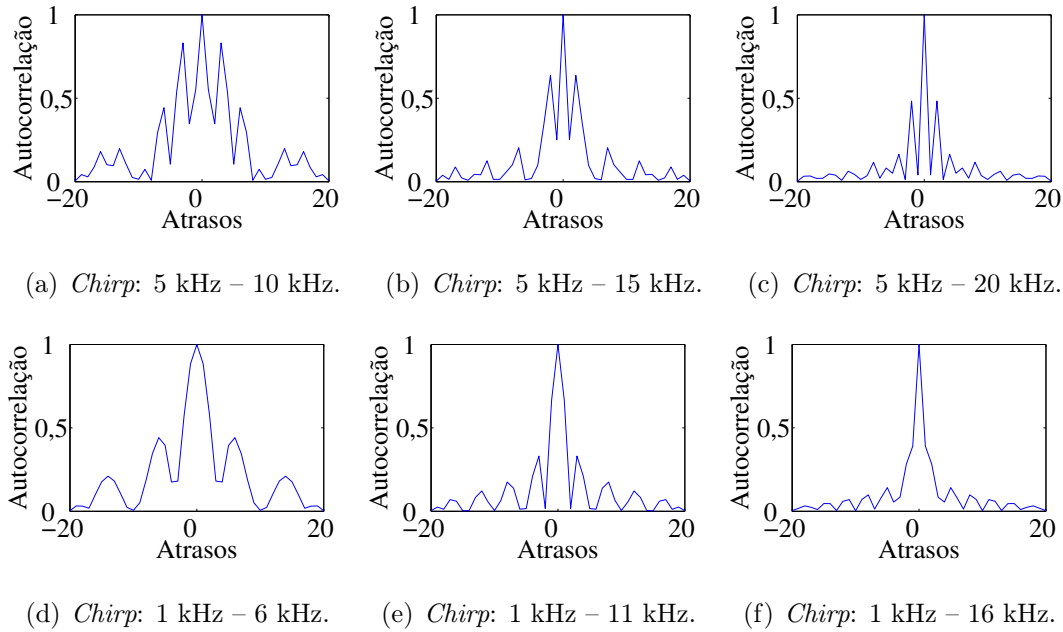


Figura 4.1: Comparação entre autocorrelações de *chirps* variando largura de banda e faixa de frequências.

Pela observação atenta da Figura 4.1, podemos notar que: (i) o aumento da largura de banda do *chirp* torna-o mais ortogonal; e (ii) a magnitude dos picos secundários nas Figuras 4.1(a), 4.1(b) e 4.1(c) são, aproximadamente, as mesmas magnitudes das amostras correspondentes a atrasos próximos de 0 nas Figuras 4.1(d), 4.1(e) e 4.1(f), apesar da aparente melhora de ortogonalidade. Portanto, podemos afirmar que o principal parâmetro responsável pela ortogonalidade dos *chirps* será a largura de banda.

A Figura 4.2 mostra a autocorrelação de *chirps* com largura de banda de 10 kHz, o mesmo sinal utilizado na Figura 4.1(b), porém, variando-se o número de amostras. Pela observação das autocorrelações obtidas, é possível constatar que a

sensibilidade da ortogonalidade dos *chirps* à quantidade de amostras é muito baixa, o que nos leva a concluir que pode-se utilizar sinais curtos.

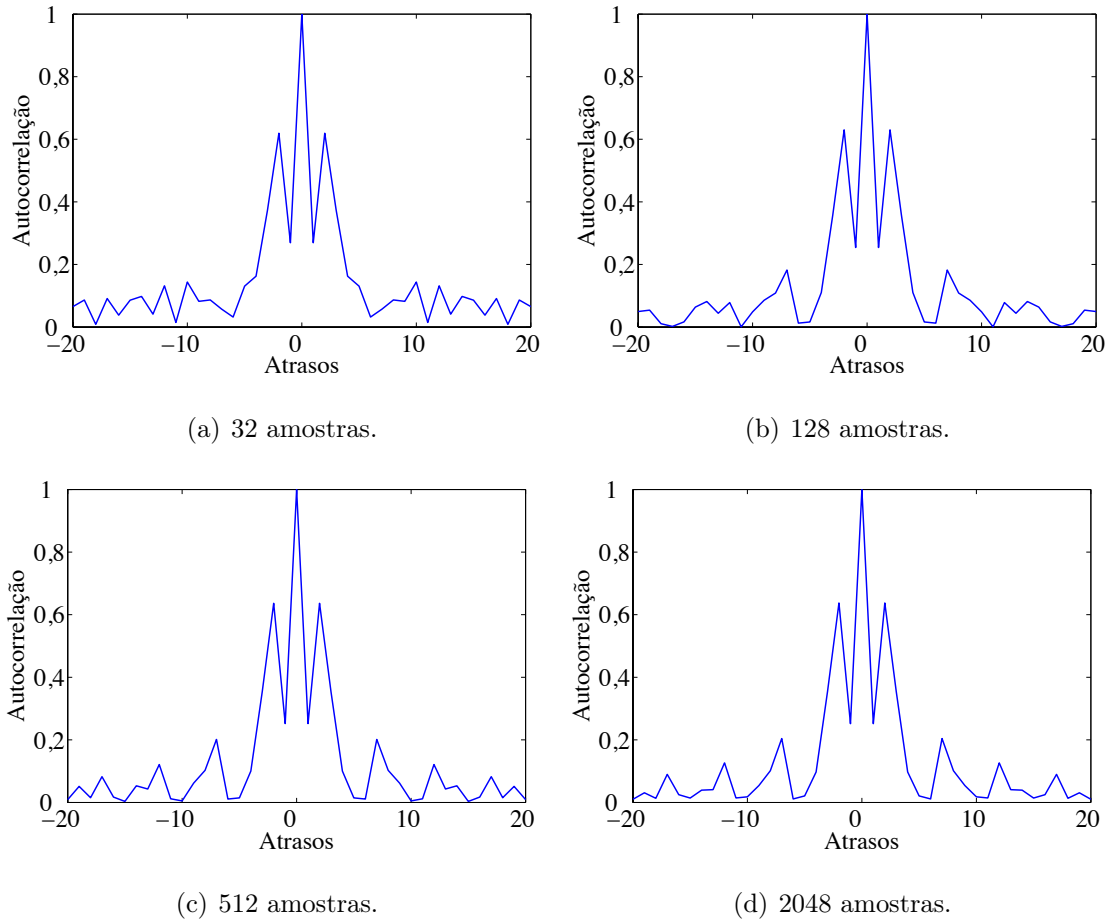


Figura 4.2: Comparação entre autocorrelações de *chirps* com diferentes números de amostras.

Na busca de sinais que fossem mais ortogonais, chegamos aos sinais pseudoaleatórios de banda plana e distribuição gaussiana (do inglês *white Gaussian noise* - WGN). A Figura 4.3 mostra a autocorrelação de um *chip*¹ dessa natureza variando-se o número de amostras. Note que, quanto maior o número de amostras, mais ortogonal é o sinal. Como comparação, mesmo o *chip* com 32 amostras da Figura 4.4(a) já se mostra consideravelmente mais ortogonal que qualquer um dos *chirps* apresentados na Figura 4.1.

A Figura 4.4 ilustra a CC de um *chip* pseudoaleatório com ele próprio contaminado por ruído. Para isso, bastou somá-lo com outro sinal pseudoaleatório. A

¹O *chip* é definido como um sinal gerado por seqüências pseudoaleatórias, de forma a transmitir informações distribuindo o sinal no espectro de frequências.

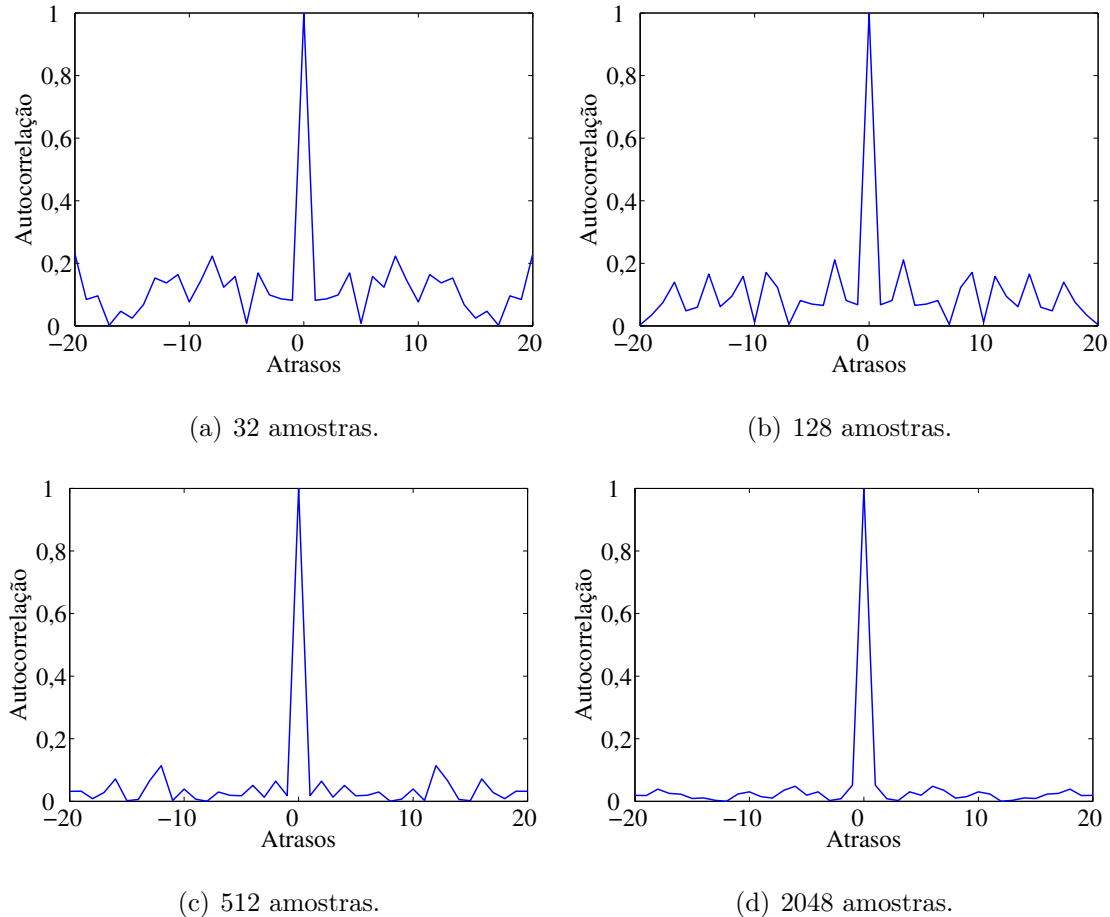


Figura 4.3: Comparação entre autocorrelações de sinais pseudoaleatórios com diferentes números de amostras.

razão sinal-ruído, ou SNR (do inglês *signal-to-noise ratio*), que mede a razão entre a potência do sinal e a potência do ruído em decibéis, é de 0 dB nesse caso, pois a potência dos sinais somados era a mesma. Observe que, aumentando-se o número de amostras dos sinais, o chão de ruído é reduzido, enquanto que o pico em $k = 0$ não sofre grandes modificações.

Outra característica interessante dos sinais pseudoaleatórios é que podem ser obtidos por diversas realizações de forma a serem ortogonais entre si. Isso permite que seja utilizado um *chip* único para cada alto-falante. Se o mesmo sinal for utilizado em todos os alto-falantes, o intervalo de reprodução dos sinais deve ser tal que um não interfira no outro, o que requer que se considere o RT-60 da sala para o cálculo desse intervalo de reprodução entre as caixas.

Caso a correlação entre os sinais de diferentes alto-falantes seja baixa, pode-se considerar intervalos de reprodução pequenos em relação ao RT-60 do ambiente, de

forma que as réplicas geradas pelas reflexões de um sinal emitido por um alto-falante não gerem picos consideráveis na CC que utiliza o sinal emitido por outro alto-falante. No entanto, a presença de outro sinal pseudoaleatório somado ao desejado diminui a SNR obtida.

Com isso, podemos reduzir o tempo de intervalo entre *chips* — sinais reproduzidos — de centenas de ms para unidades de ms, obtendo resultados semelhantes. Com a redução dos intervalos, o tempo de reprodução de todos os *chips* será drasticamente reduzido, possibilitando, por exemplo, localizar microfones em movimento.

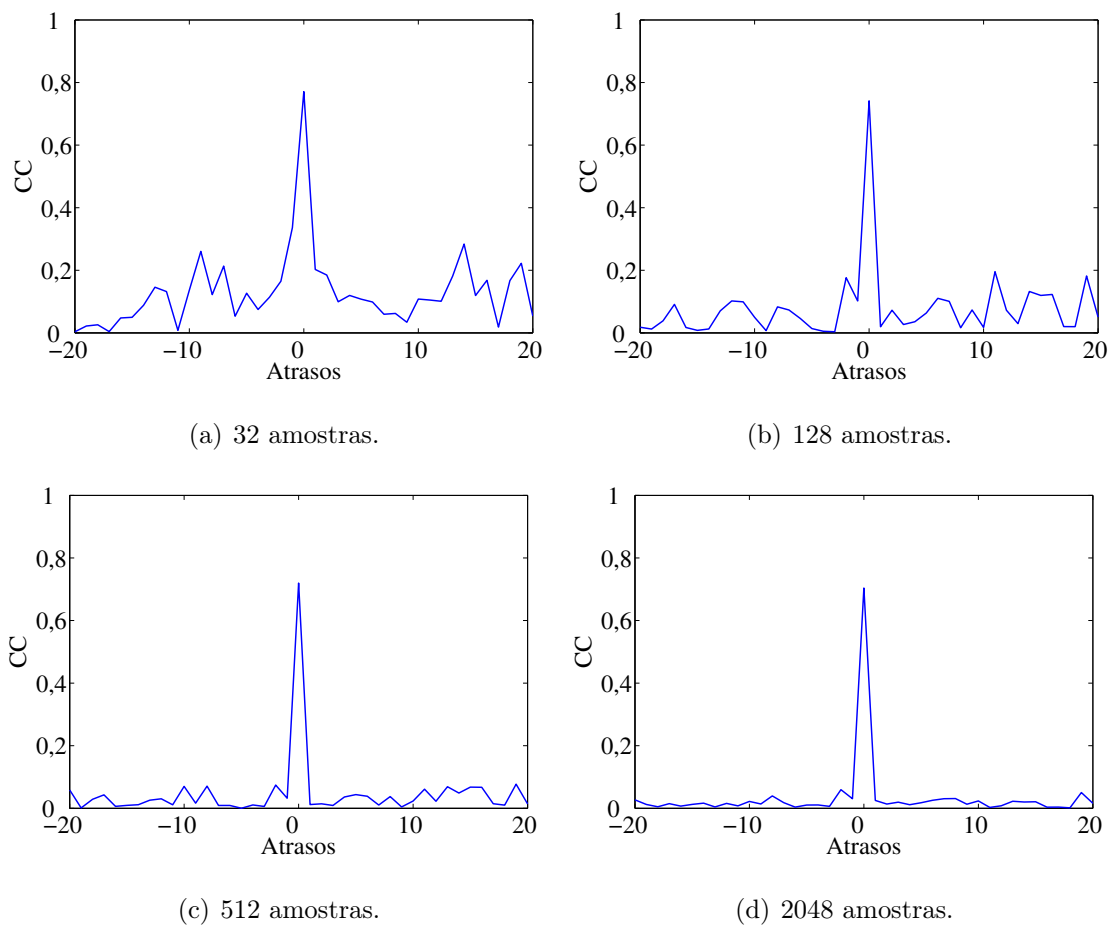


Figura 4.4: Comparação entre auto-correlações de sinais pseudo-aleatórios corrompidos por ruído e com diferentes números de amostras (SNR = 0 dB).

Os sinais utilizados no sistema original [2] são *chirps* de 10 ms de duração com frequências variando de 5 kHz a 10 kHz, e com a autocorrelação ilustrada na Figura 4.1(a). Pelos motivos citados, iremos utilizar sinais pseudoaleatórios, de forma a aumentarmos a aplicabilidade do sistema em situações reais. As simulações realizadas serão descritas na próxima seção.

4.2 Simulações

4.2.1 Metodologia

Como forma de avaliar o desempenho dos métodos de localização de sensores acústicos em situações reais, foram realizados experimentos em ambientes de simulação e, posteriormente, em ambientes reais controlados. As simulações serão de grande utilidade para obtenção de dados estatísticos sobre os métodos, já que as mesmas podem ser realizadas um grande número de vezes, se dispusermos do poder computacional necessário. Tal procedimento seria inviável em experimentos práticos, fato devido ao alto custo de tempo de suas realizações. Em nosso caso, desejamos fazer um apanhado geral do desempenho dos métodos, valendo-nos de diferentes dimensões de salas, tempos de reverberação e níveis de ruído ambiente.

Adicionando Ruído

A incorporação de ruído branco gaussiano aditivo aos sinais capturados de cada microfone pode fornecer um bom indicativo da robustez do método à presença de sinais indesejados de banda larga, como ventiladores e aparelhos de ar-condicionado. A métrica utilizada para avaliação de qualidade é a SNR. Os ruídos utilizados nas simulações são sequências pseudoaleatórias somadas ao sinal reverberado que chega a cada microfone de forma independente, não havendo correlação significativa entre os sinais resultantes deste processo.

Simulando Reverberação de Salas

Nestas simulações, foram utilizadas rotinas em Matlab[®] disponibilizadas em [14] de emulação de resposta ao impulso de uma sala para dadas dimensões, coeficiente de reflexão $\bar{\alpha}_i = 1 - \alpha_i$, referente à i -ésima superfície, com fator de absorção α_i , e posição do emissor e do receptor. As rotinas simulam as réplicas do sinal emitido formadas pelas superfícies de uma sala paralelepípedica modelando os efeitos da propagação no ar e das reflexões nas superfícies por filtragens e atenuações. Todos os sinais são gerados com essa rotina. Para conseguir uma sala anecoica, bastará configurar todos os coeficientes de reflexão para $\bar{\alpha}_i = 0$.

Para as simulações, foi utilizado o mesmo coeficiente $\bar{\alpha}$ de reflexão para todas

as superfícies da sala. Para seu cálculo, foi utilizada a seguinte equação:

$$\bar{\alpha} = e^{\left(-\frac{kV}{AT_{60}}\right)}, \quad (4.5)$$

onde A é o somatório das áreas das superfícies da sala em m^2 , $k = 0,161$ s/m para o ar com temperatura de 20°C , V é o volume da sala em m^3 e T_{60} é o RT-60 desejado em s. Esta equação foi obtida por manipulação da equação de Eyring [15] para o cálculo de RT-60 dada por

$$T_{60} = -\frac{kV}{A \ln(\bar{\alpha})}. \quad (4.6)$$

Note que esta equação é uma adaptação da Equação (2.3) de Sabine [10], tornando o modelo mais preciso, principalmente, para salas com tempo de reverberação curto.

Método de Monte Carlo

Para todas as simulações, foi utilizado o método de Monte Carlo a fim de obtermos resultados com significado estatísticos dos métodos de localização de sensores acústicos. O método de Monte Carlo consiste em realizar o mesmo experimento diversas vezes sorteando-se o valor das variáveis aleatórias intrínsecas ao sistema. O processo foi realizado 100 vezes para cada cenário de teste, de forma a nos proporcionar uma gama rica de resultados.

Segundo aponta o autor [2], o método de localização de sensores oferece resultados mais satisfatórios quando os microfones estão dispostos no interior de uma região do espaço circunscrita pelos alto-falantes utilizados. Por esse motivo, os alto-falantes foram dispostos de tal forma que houvesse um em cada superfície da sala, totalizando 6 alto-falantes. Para que fossem obtidos resultados mais genéricos, que não privilegiassem posições dentro da sala, as simulações foram feitas gerando posições dos microfones de forma aleatória com distribuição uniforme ao longo da sala. Como forma de facilitar os modelos matemáticos, todos os alto-falantes e microfones serão omnidirecionais.

A métrica utilizada para medir o erro de estimação em cada simulação é a norma euclidiana:

$$e_m \triangleq \|\mathbf{m}_m - \hat{\mathbf{m}}_m\|, \quad (4.7)$$

onde \mathbf{m}_m é a posição conhecida do m -ésimo microfone e $\hat{\mathbf{m}}_m$ é a posição estimada do m -ésimo microfone com o método de localização em questão. Com isso, teremos o erro médio calculado da seguinte forma:

$$\bar{e} \triangleq \frac{1}{M} \sum_{m \in \mathcal{M}} e_m. \quad (4.8)$$

Para as simulações, foram escolhidos sinais pseudoaleatórios com duração de 10 ms (960 amostras) e amostrados a 48 kHz, foi utilizado o valor $c = 340$ m/s para a velocidade do som, e o limiar de erro médio γ , utilizado no algoritmo de seleção de TOFs (vide Subseção 3.3.4), foi de 2 cm.

4.2.2 Resultados

Ruído

A Figura 4.5 mostra o erro médio obtido das 100 realizações para cada valor de SNR. Os resultados em azul são referentes ao método LS (vide Subseção 3.3.2); os resultados em vermelho tracejado são referentes ao método de seleção de TOF (vide Subseção 3.3.4). Nesta simulação, foi utilizada uma sala de dimensões 5,2 m \times 7,5 m \times 2,6 m e nenhuma reverberação, de forma a avaliarmos apenas a interferência causada por ruído. Foram calculados resultados para SNRs variando de -30 dB a 15 dB.

É possível observar que a presença de ruído no ambiente não causa impacto algum nos resultados até se atingir uma SNR em torno de -10 dB, quando o pico referente ao TOF nas CCs começa a se confundir com o chão de ruído. Os resultados obtidos com o método de seleção de TOF foram iguais aos obtidos pelo método LS. Note que este resultado está relacionado com o número de amostras escolhido para os sinais. Utilizando-se sinais mais longos, o limiar de SNR para que os métodos de localização comecem a errar a posição dos microfones se reduzirá.

Reverberação

A Figura 4.6 mostra o erro médio obtido das 100 realizações para cada valor de RT-60, com intervalos de 200 ms. Novamente, a curva em azul é referente aos resultados obtidos com o método LS e a vermelha tracejada, ao método de seleção de TOF. Como, nesta simulação, não é somado ruído aos sinais, a SNR estará

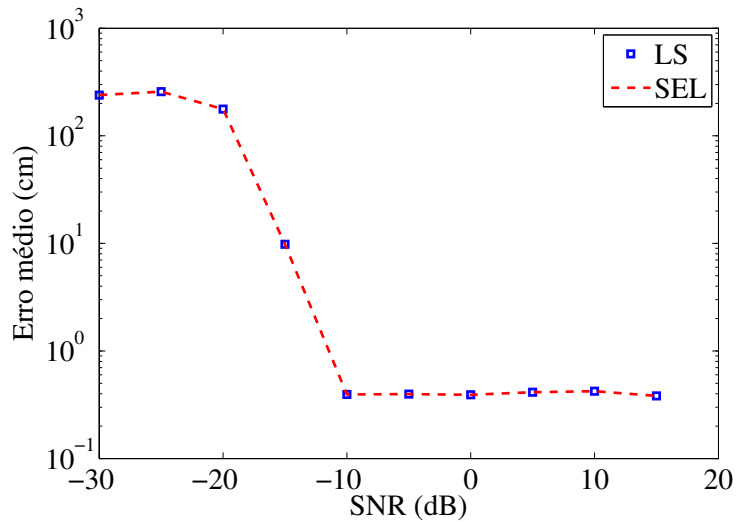


Figura 4.5: Resultado para LS e seleção de TOFs das simulações variando a SNR em uma sala anecoica de dimensões $5,2 \text{ m} \times 7,5 \text{ m} \times 2,6 \text{ m}$.

associada ao ruído de quantização do sinal, que é desprezível para fins práticos. A sala utilizada também tem dimensões $5,2 \text{ m} \times 7,5 \text{ m} \times 2,6 \text{ m}$. Observando a Figura 4.6, é possível constatar que o método de seleção de TOF tornou o sistema robusto aos efeitos da reverberação, o que não seria possível sem realizar correções diretamente nas estimativas de TOF.

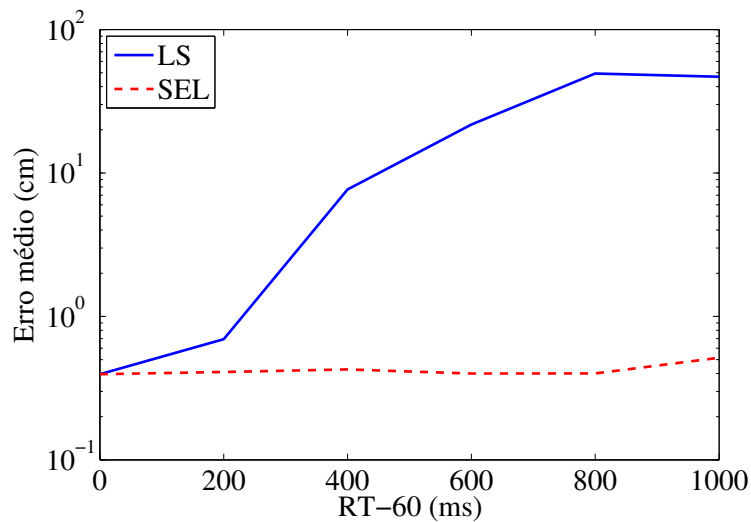


Figura 4.6: Resultado para LS e seleção de TOFs das simulações variando o RT-60 de uma sala de dimensões $5,2 \text{ m} \times 7,5 \text{ m} \times 2,6 \text{ m}$ com SNR idealmente infinita.

Salas – Ruído e Reverberação

Para simular situações reais, foram elencados cenários que representassem ambientes práticos. Foram escolhidos três diferentes tamanhos de sala e, para cada sala, duas diferentes SNRs e dois RT-60 compatíveis com as dimensões da sala em questão. Os cenários escolhidos são:

- Sala pequena: 3 m × 2 m × 2,5 m
 - SNR: -10 dB e 10 dB
 - RT-60: 300 ms e 600 ms

- Sala média: 5,2 m × 7,5 m × 2,6 m
 - SNR: -10 dB e 10 dB
 - RT-60: 400 ms e 800 ms

- Sala grande: 20 m × 15 m × 6 m
 - SNR: -10 dB e 10 dB
 - RT-60: 700 ms e 1400 ms

Como forma de ilustrar os resultados obtidos, foram utilizados não somente o erro médio de cada simulação nas tabelas como também histogramas, possibilitando uma visualização mais detalhada da distribuição dos erros de estimação de posição para cada cenário simulado.

A sala pequena pode ser comparada a um pequeno quarto com objetos absorvedores, como, por exemplo, uma cama ou cortinas, no caso em que RT-60 = 300 ms. Para o caso com RT-60 = 600 ms, esta sala pode ser comparada a um banheiro ou algum outro ambiente tipicamente mais “vivo”, ou reverberante. Os resultados referentes a estas salas estão ilustrados na forma de histogramas nas Figuras 4.7 e 4.8 e resumidos na Tabela 4.1 com média e desvio-padrão relativos a cada cenário.

As Figuras 4.7 e 4.8, ilustram os resultados para RT-60 = 300 ms e 600 ms para os métodos LS e com seleção de TOF. Os histogramas apresentam o aumento drástico da quantidade de estimativas com erro próximo de zero, apesar de ainda restarem *outliers* — erros de estimação — para SNR = -10 dB. Para a configuração

com 10 dB de SNR, todos os *outliers* foram corrigidos, passando o erro médio de $14,9 \pm 37,3$ cm com LS para $0,4 \pm 0,3$ cm com seleção de TOF.

A configuração com $\text{SNR} = -10$ dB e $\text{RT-60} = 600$ ms se mostrou ser a situação mais severa dentre todas as simulações. Talvez isso se deva à proximidade maior das paredes, fazendo com que as reflexões na sala cheguem mais rapidamente aos microfones, com isso causando interferência maior nos sinais de visada direta. Essas interferências contribuem para subir o chão de ruído e gerar outros picos altos na CC, pois todo sinal refletido é uma réplica do primeiro.

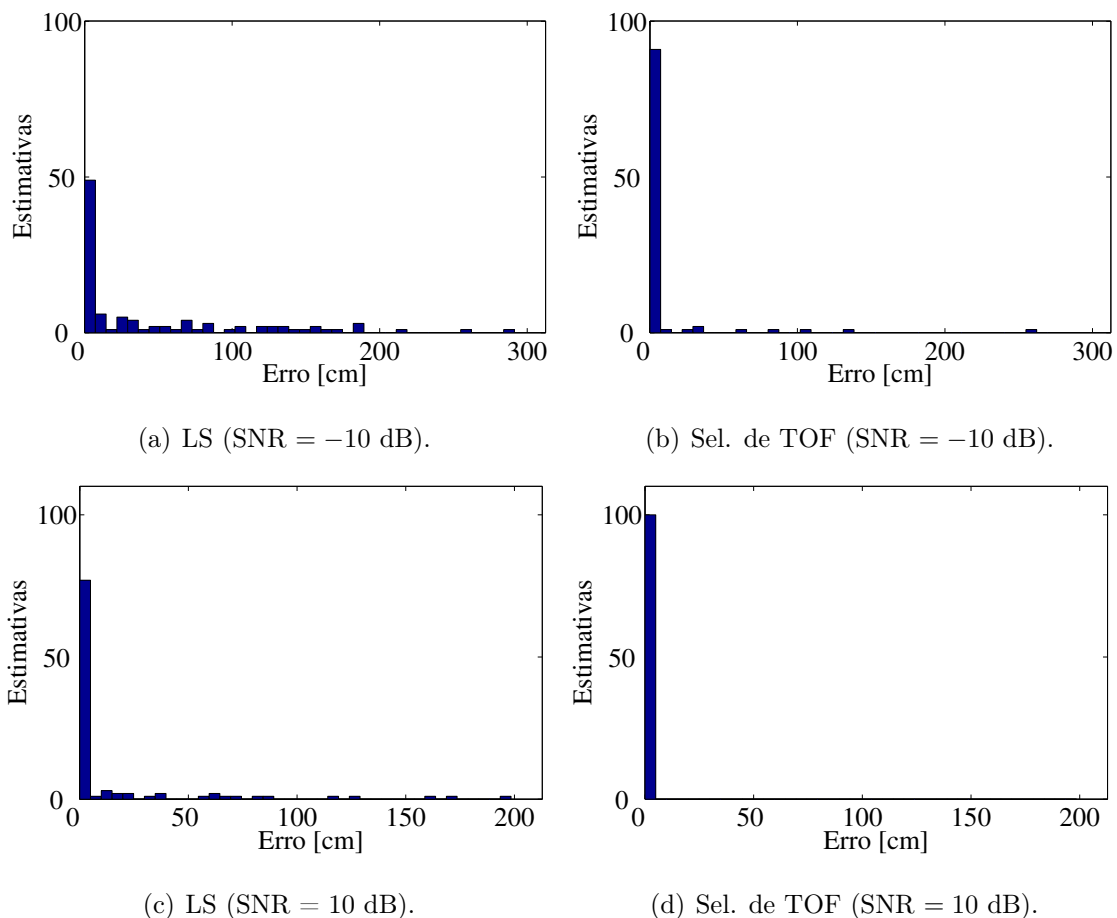


Figura 4.7: Histogramas das simulações de uma sala pequena com $\text{RT-60} = 300$ ms.

Quanto à sala de tamanho médio, esta pode ser comparada a uma sala de aula ou de estar. Se houver cortinas, tapetes, sofás ou outros objetos absorvedores, esta sala pode ser comparada ao caso da simulação com RT-60 de 400 ms, sendo esta uma sala relativamente “seca” ou “morta”. Não havendo muitos objetos dessa natureza, a sala se torna mais reverberante, sendo caracterizada pela simulação com RT-60 de 800 ms. Os resultados referentes a estas salas estão ilustrados nas

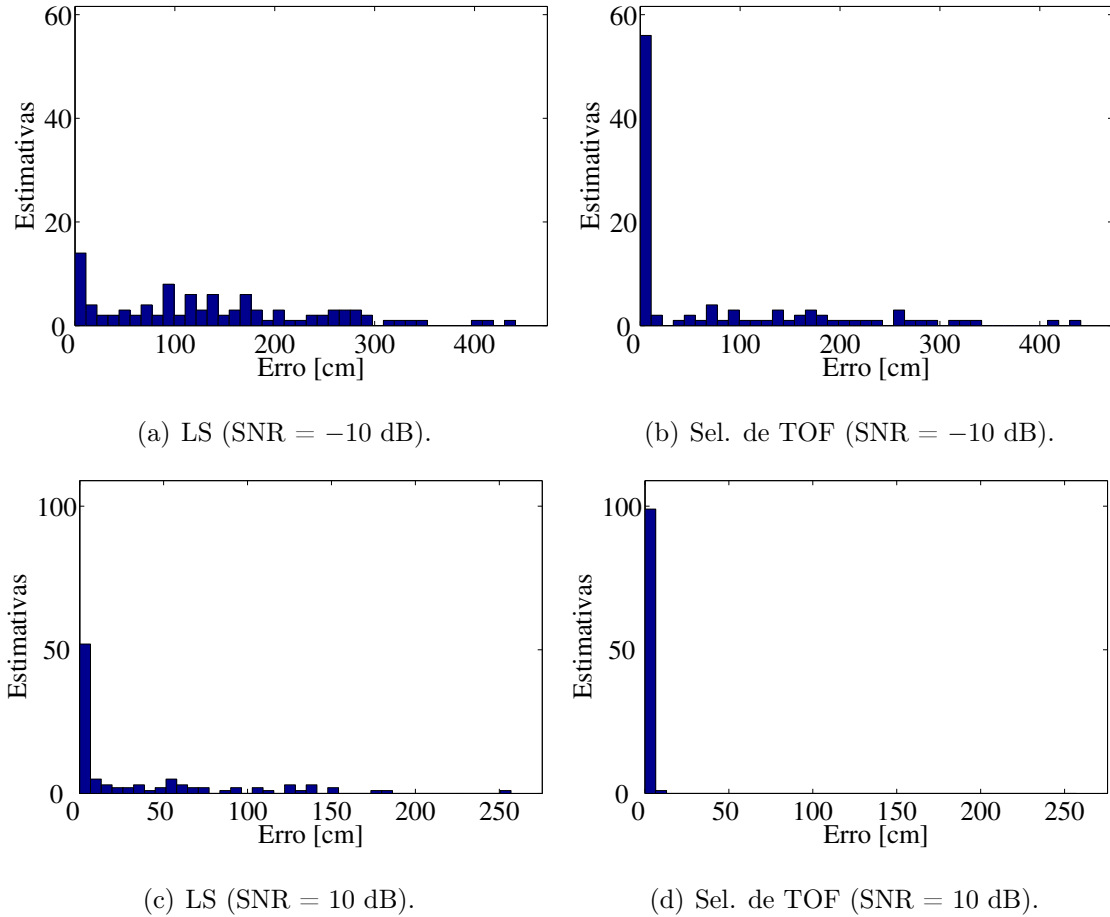


Figura 4.8: Histogramas das simulações de uma sala pequena com $RT-60 = 600$ ms.

Tabela 4.1: Resumo dos resultados da simulação da sala pequena.

RT-60	300 ms		600 ms	
SNR	-10 dB	10 dB	-10 dB	10 dB
Método	Erro médio [cm] (desvio-padrão [cm])			
LS	46,1 (66,2)	14,9 (37,3)	139,8 (105,7)	36,5 (53,7)
Sel. de TOF	7,9 (32,4)	0,4 (0,3)	75,5 (108,2)	0,4 (0,7)

Figuras 4.9 e 4.10 e também resumidos na Tabela 4.2 com média e desvio-padrão relativos a cada cenário.

Para esta sala, todos os resultados utilizando o método de seleção de picos foram satisfatórios, apenas apresentando 4 *outliers* dentre 100 estimativas na configuração com mais ruído e reverberação, o que resultou em um erro médio de $6,6 \pm 33,4$ cm.

Neste ponto, importa destacar novamente que o modelo utilizado para medida de SNR leva em conta a potência do sinal que chega aos microfones, ou seja, sinais com a reverberação da sala. Com isso, é difícil definir com exatidão qual a real SNR em relação ao sinal de visada direta puro, mais ainda sem levar em conta possíveis réplicas somadas a este sinal, o que vai variar de acordo com a posição do microfone e do alto-falante.

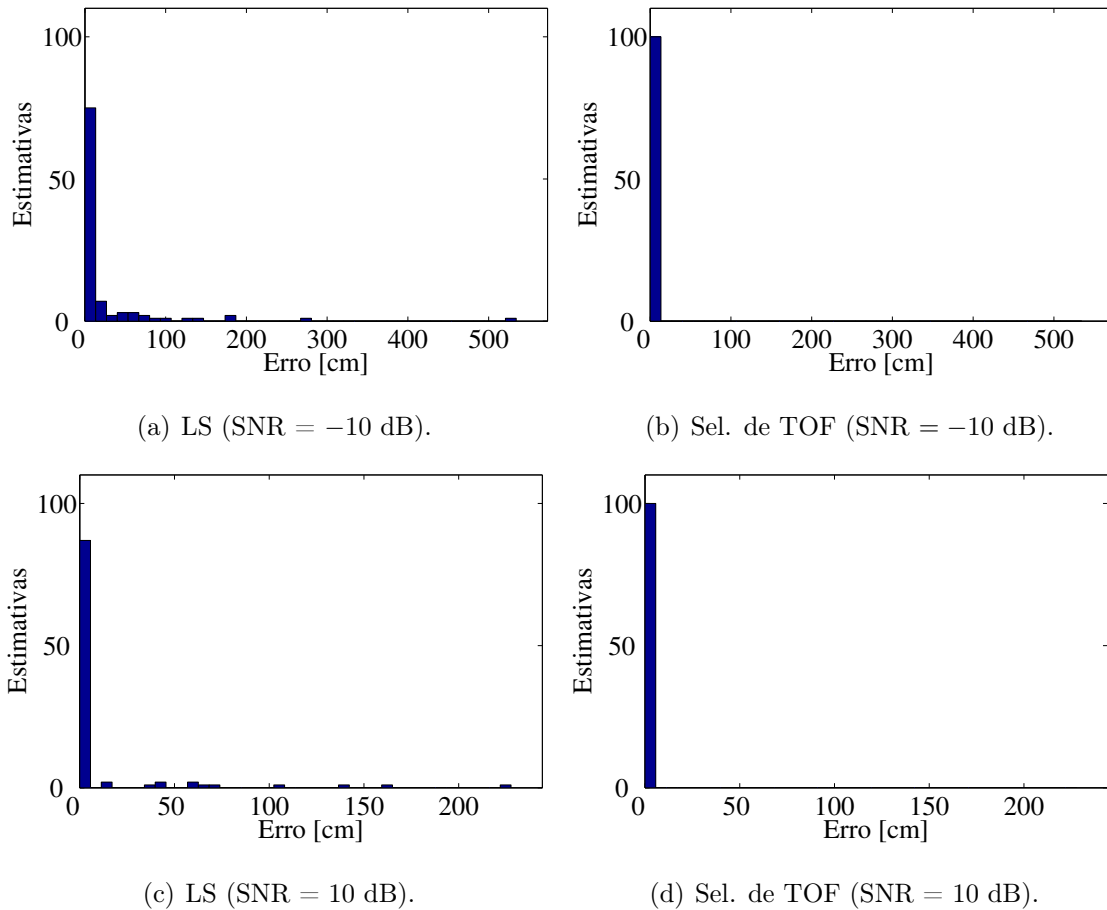


Figura 4.9: Histogramas das simulações de uma sala média com $RT_{60} = 400$ ms.

Tabela 4.2: Resumo dos resultados da simulação da sala média.

RT-60	400 ms		800 ms	
SNR	-10 dB	10 dB	-10 dB	10 dB
Método	Erro médio [cm] (desvio-padrão [cm])			
LS	23,1 (66,9)	10,8 (34,5)	93,6 (132,5)	30,1 (67,5)
Sel. de TOF	0,4 (0,2)	0,4 (0,2)	6,6 (33,4)	0,5 (0,6)

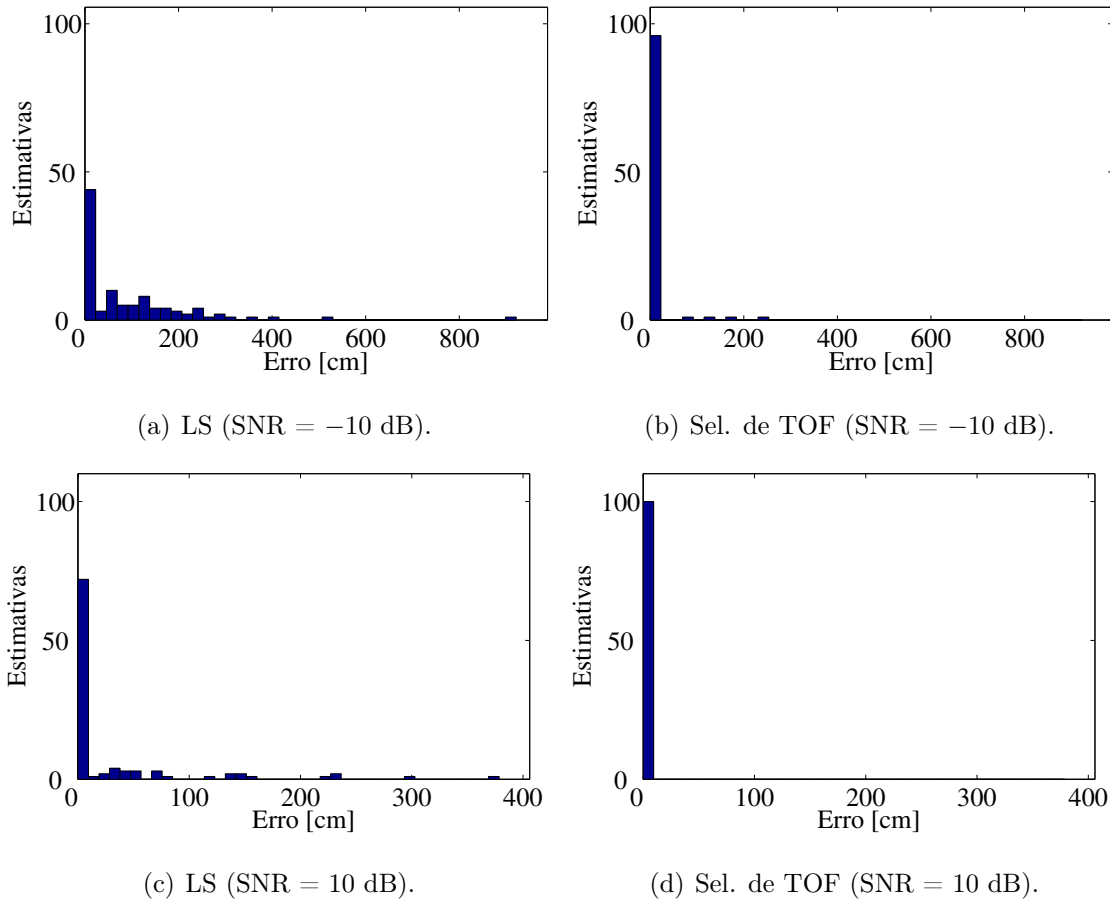


Figura 4.10: Histogramas das simulações de uma sala média com $RT-60 = 800$ ms.

A sala grande representa grandes salas de aula ou auditórios. Nesse tipo de sala, é comum o uso de muitas cadeiras acolchoadas, piso de madeira ou carpete, etc. As simulações com $RT-60$ de 700 ms são uma tentativa de caracterizar esses casos, enquanto uma situação onde haja menos objetos absorvedores é caracterizada no caso de $RT-60$ de 1400 ms.

Os resultados referentes a estas salas estão ilustrados nas Figuras 4.11 e 4.12 e resumidos na Tabela 4.3 com média e desvio-padrão relativos a cada cenário. Em geral, estes cenários não configuraram situações especialmente desafiadoras para os métodos, de forma que foram obtidos baixos erros médios em todos os resultados para o método de seleção de TOF.

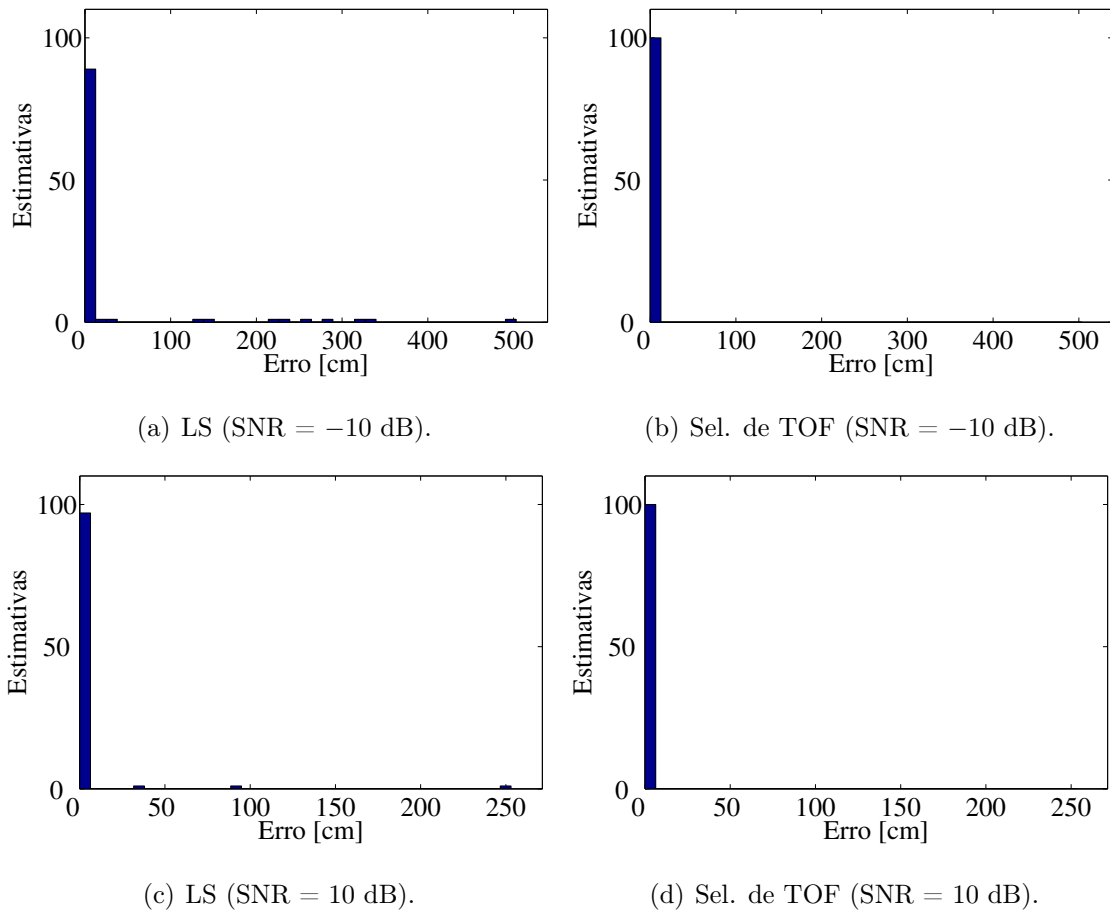
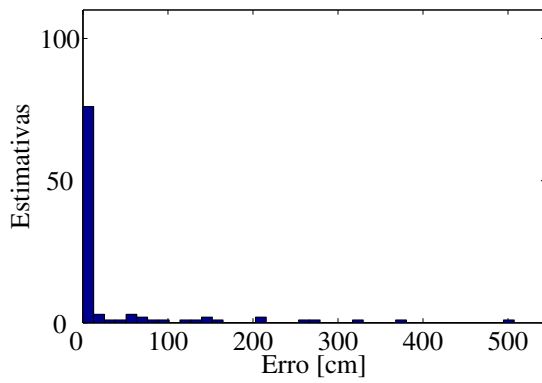


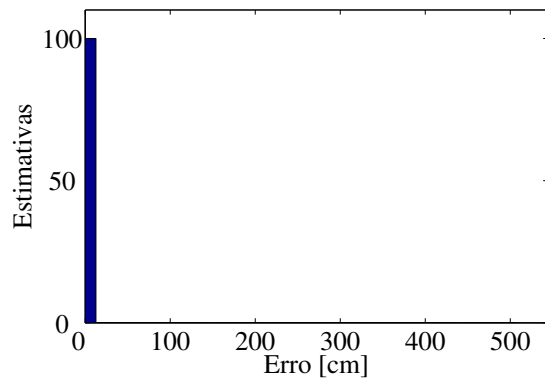
Figura 4.11: Histogramas das simulações de uma sala grande com $RT-60 = 700$ ms.

Tabela 4.3: Resumo dos resultados da simulação da sala grande.

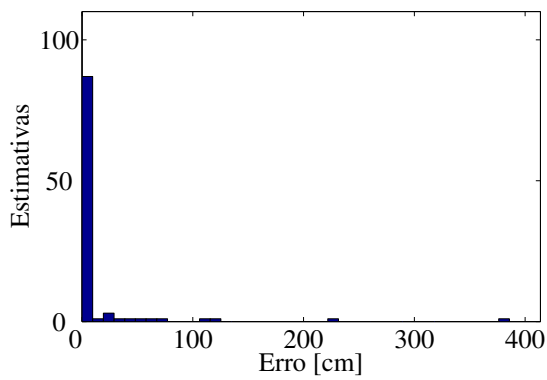
RT-60	700 ms		1400 ms	
SNR	-10 dB	10 dB	-10 dB	10 dB
Método	Erro médio [cm] (desvio-padrão [cm])			
LS	25,0 (83,0)	4,2 (26,3)	35,1 (86,1)	12,0 (47,0)
Sel. de TOF	0,4 (0,2)	0,4 (0,4)	0,5 (0,2)	0,4 (0,2)



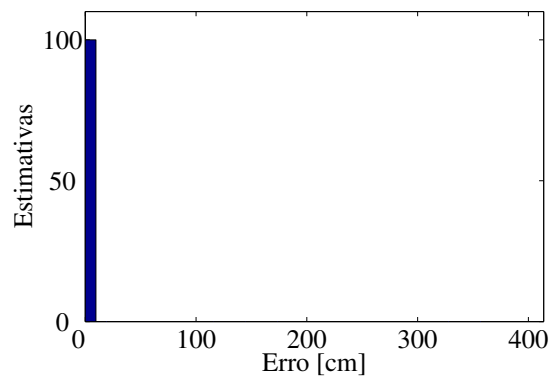
(a) LS (SNR = -10 dB).



(b) Sel. de TOF (SNR = -10 dB).



(c) LS (SNR = 10 dB).



(d) Sel. de TOF (SNR = 10 dB).

Figura 4.12: Histogramas das simulações de uma sala grande com $RT-60 = 1400$ ms.

4.3 Experimento Prático

4.3.1 *Setup* Experimental

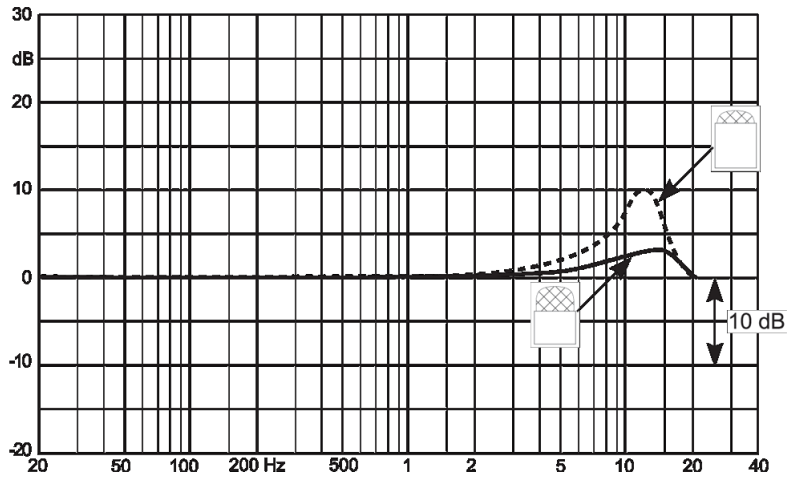
Como forma de aferição experimental do desempenho dos métodos, foi realizado um experimento prático utilizando equipamentos reais em uma sala cujas características foram medidas. A sala utilizada tem dimensões $5,2\text{ m} \times 7,5\text{ m} \times 2,6\text{ m}$ e $RT-60$ de, aproximadamente, 550 ms , cuja resposta ao impulso está ilustrada na Figura 2.7 e cuja resposta em frequência na Figura 2.8. A velocidade de propagação c medida indiretamente pela temperatura de 23°C é de $344,8\text{ m/s}$. A SNR medida durante os experimentos oscilava em torno de 35 dB .

Nesse experimento, foram utilizados 16 microfones da marca DPA e modelo SMK4060. Estes microfones são condensadores omnidirecionais e suas cápsulas têm dimensão aproximada de $0,5\text{ cm}$. A Figura 4.13 exibe uma fotografia do microfone sem a proteção externa da capsula. Suas pequenas dimensões nos são favoráveis no sentido de proporcionar medidas mais precisas de sua posição espacial e de os tornarem mais próximos de sensores pontuais. Por serem condensadores, estes microfones necessitam de alimentação externa de 48 V .

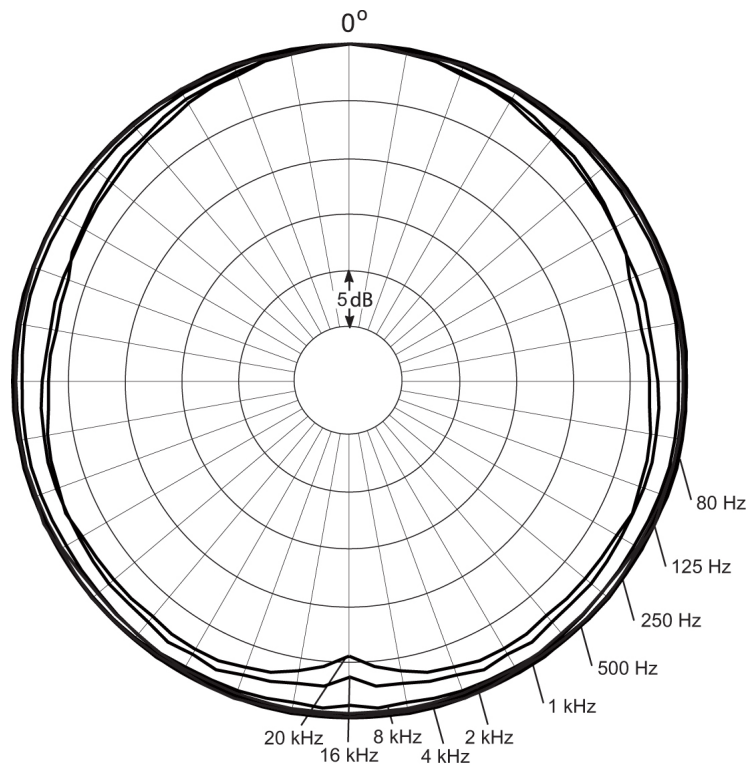


Figura 4.13: Microfone condensador omnidirecional [16].

Na Figura 4.14 estão dispostas as especificações técnicas mais relevantes dos microfones utilizados: sua resposta em frequência e seu padrão polar. Sua característica extremamente omnidirecional é um fator que favorece sua localização, conforme discutido na Subseção 2.4.2. A curva da resposta em frequência utilizada é a fornecida em traço cheio. Note que estes microfones têm resposta em frequência bastante plana, alcançando um ganho máximo de aproximadamente 3 dB em torno de 15 kHz . As informações foram fornecidas pelo *site* [16] do fabricante.



(a) Resposta em frequência [16].



(b) Padrão polar [16].

Figura 4.14: Especificações técnicas do microfone.

Os alto-falantes [17] utilizados estão ilustrados na Figura 4.15. Estes alto-falantes proporcionam boa resposta em frequência de, aproximadamente, 120 Hz até 18 kHz, podendo ser excitados com, no máximo, 100 W_{RMS} . Com aproximadamente 7,6 cm de diâmetro, os alto-falantes impõem certa dificuldade para medir sua posição espacial.

Todas as medidas de posição foram realizadas utilizando-se uma trena óptica,



Figura 4.15: Fotografia do alto-falante utilizado.

o que facilita medidas de distâncias maiores que 1 m. Porém, os alto-falantes são construídos com uma grade de proteção frontal, o que impossibilita o acesso direto ao seu cone para medição precisa. Por isso, utilizamos a borda dessa grade como referência para a medição, considerando que o ponto eleito para considerar a posição do alto-falante será o centro do cone, conforme ilustrado na Figura 4.16.

Como os alto-falantes são todos presos em hastes, as posições são medidas de forma relativa às hastes em que os alto-falantes se encontram fixados. As cruzes brancas na Figura 4.16 mostram dois pontos utilizados como referência para medição. O ponto 1 é utilizado para medições horizontais, por estar na mesma linha vertical do centro do cone, enquanto o ponto 2 é utilizado para medidas verticais, como medida da altura do alto-falante, por motivo análogo ao anterior.

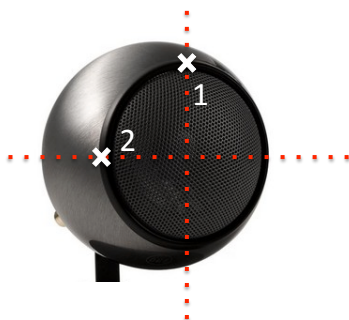
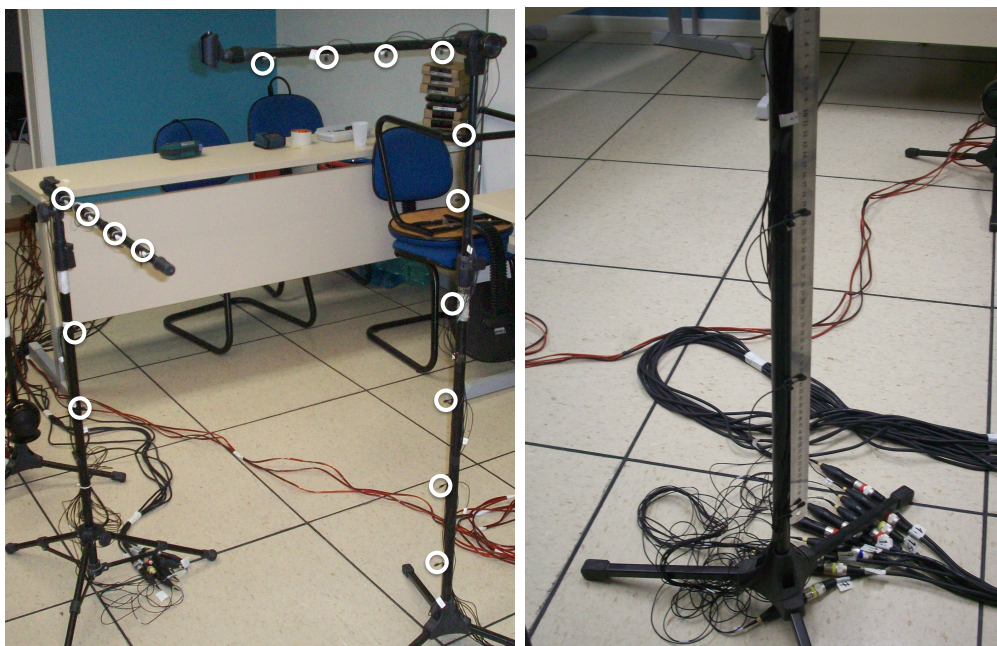


Figura 4.16: Referência para medição da posição dos alto-falantes.

A Figura 4.17 mostra os pequenos microfones montados em dois pedestais metálicos, do mesmo tipo dos utilizados para afixar os alto-falantes. Os círculos brancos evidenciam suas posições, para facilitar a visualização. Todas as medidas foram realizadas relativas aos pedestais, para que houvesse certa flexibilidade do

arranjo em relação à posição do pedestal no chão. A Figura 4.17(b) ilustra a medição de pequenas distâncias realizada com uma régua metálica.



(a) Microfones montados.

(b) Detalhe de medição.

Figura 4.17: Microfones utilizados para o experimento prático montados no pedestal. Os círculos brancos indicam a localização dos pequenos microfones.

Na Figura 4.18 está uma foto de todo o *setup* de gravação, com os microfones e alto-falantes posicionados da forma como foi realizado o experimento. Todas as medidas de posição tomaram como referência o eixo de coordenadas ilustrado no canto inferior direito da imagem. As marcações do piso foram utilizadas para medir as coordenadas x e y dos pedestais. Estimamos que, pela forma como foram realizadas as medidas, o erro associado às medições é de, aproximadamente, 1 cm para os microfones, o que deve ser levado em consideração na avaliação dos resultados.

Os alto-falantes foram dispostos, assim como nas simulações, de forma a circunscreverem os microfones. Isso favorecerá, também, o posicionamento dos microfones em relação à direção preferencial de emissão dos alto-falantes, uma vez que estes não são omnidirecionais.

Para realizarmos as gravações e reproduções dos sinais, dispomos de equipamentos de alta fidelidade. Foram utilizadas as interfaces de áudio Fireface UFX e Fireface UCX da marca RME [18]. Aliadas a elas, dois módulos de pré-amplificadores de oito canais modelo OctaMic D, também da RME, e um amplificador para os



Figura 4.18: *Setup* utilizado no experimento prático. Há 12 alto-falantes (esferas pretas na imagem), e 16 microfones.

alto-falantes da marca Denon [19], modelo POA-3012CI, que possibilita o uso de 12 canais independentes, um para cada alto-falante.

A Figura 4.19 ilustra a forma como foram realizadas as interconexões dos equipamentos. A interface Fireface UFX — Interface 1 na imagem — recebe dos dois módulos de pré-amplificadores os sinais dos 16 microfones digitalizados na taxa desejada e realiza a conversão DA de 6 dos 12 canais de saída, enviando-os para o amplificador. Os outros 6 canais são convertidos pela Fireface UCX — Interface 2 na imagem —, enviando-os igualmente para o mesmo amplificador. Do amplificador, os sinais são enviados aos alto-falantes, que os reproduzem na sala. As duas interfaces estão conectadas via USB com um computador, que envia e recebe os sinais citados, armazenando os dados adquiridos para processamento *off-line*. Os pré-amplificadores e interfaces estão sincronizados por um sinal de *clock* gerado pela interface Fireface UFX.

Ao realizarmos um primeiro teste, notamos que os erros obtidos estavam muito acima do esperado, o que nos levou a analisarmos os TOFs obtidos. Após uma análise cuidadosa, foi observado que, em todos os TOFs medidos, havia uma

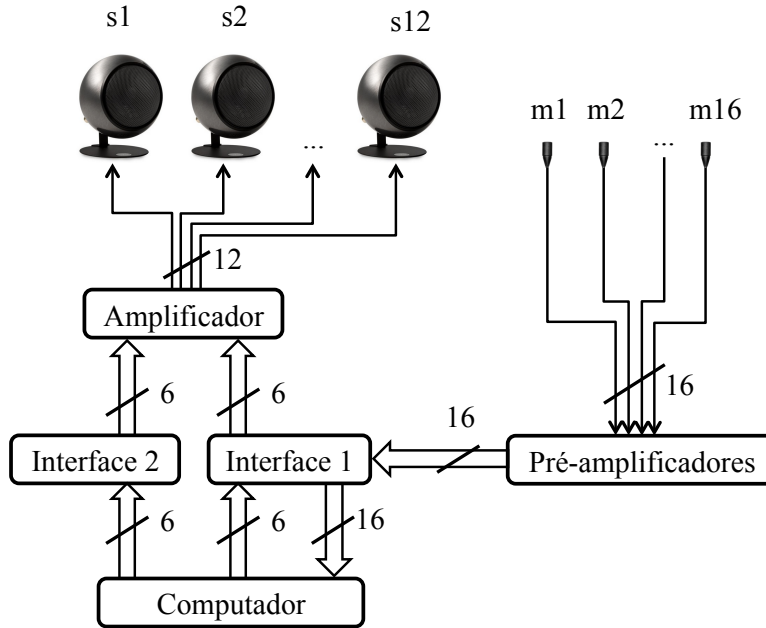


Figura 4.19: Esquemático do sistema de gravação e reprodução do experimento.

polarização nos resultados por conta de um atraso constante causado pelos equipamentos.

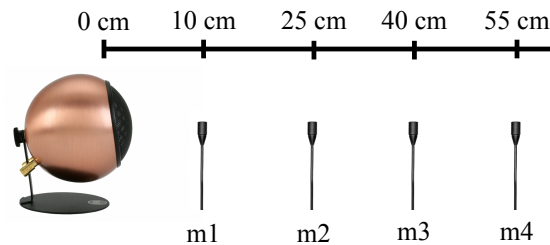
Para calibrarmos o equipamento de forma a compensarmos o atraso nas rotinas dos métodos, a primeira etapa foi medir o atraso causado pelas interfaces de áudio. Assim, foi reproduzido um *chirp* de banda completa por uma das saídas e, interligando esta saída com uma entrada, gravamos o sinal reproduzido. Foi calculada a CC entre os sinais e não foi possível observar atraso algum. Apenas após interpolarmos os sinais aumentando 10 vezes a taxa foi possível observar algum atraso. Em suma, o atraso causado por estes equipamentos é desprezível, estando em torno de $2 \mu\text{s}$, ou 10% do período de amostragem.

Diante disso, restou-nos medir o atraso causado no sistema amplificador + alto-falantes, de forma a garantirmos que qualquer atraso entre o sinal convertido e o sinal acústico reproduzido será contemplado. Dessa forma, montamos um pequeno *setup* de gravação com um arranjo linear alinhado com um alto-falante, conforme ilustrado na Figura 4.20(a). Foram utilizados 6 *chirps* lineares, de forma a reproduzirem 6 oitavas do espectro entre 312,5 Hz e 20 kHz. Desta forma, poderíamos observar algum comportamento não-linear da fase.

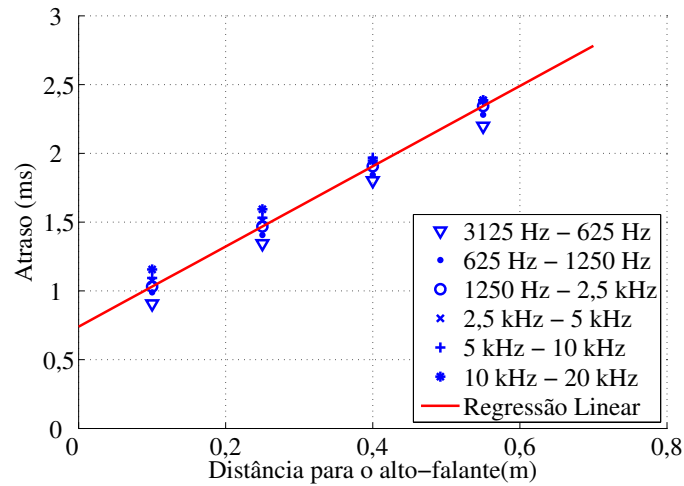
Após gravados os sinais, foi realizado o processo de CC para medirmos os

atrasos em 4 diferentes distâncias. Com isso, obtivemos o gráfico ilustrado na Figura 4.20(b), onde há quatro diferentes pontos para cada oitava reproduzida. Como forma de obtermos um atraso que mais bem representasse os atrasos medidos em todas as oitavas, foi realizada uma regressão linear com todas as medidas, resultando no atraso de 0,76 ms.

Para fins ilustrativos, considerando $c = 340$ m/s, teríamos um erro de TOF associado a uma distância percorrida de aproximadamente 26 cm, o que causaria erro inadmissível para o cálculo correto das posições dos microfones. Com essa mesma velocidade do som, o erro gerado no TOF considerando o atraso causado pelas interfaces seria equivalente a uma distância de 0,7 mm, dimensão que é desprezível para fins práticos.



(a) Esquema de gravação.



(b) Regressão linear.

Figura 4.20: *Setup* utilizado para medir o atraso causado pelos equipamentos.

4.3.2 Resultados

Para o teste prático, foram utilizados dois sinais distintos: um *chirp* de 10 ms com frequência variando de 1 kHz a 15 kHz e um sinal pseudoaleatório, também de 10 ms. Para o primeiro, os sinais de cada caixa foram reproduzidos com um intervalo de 1 s — aproximadamente, duas vezes o RT-60 da sala —, de forma a podermos considerar que os sinais chegam aos microfones individualmente, ou sem interferência dos anteriores. Para o segundo, foi utilizado este mesmo intervalo para um conjunto de realizações e, em outro conjunto, um intervalo de 3 ms, de forma que o tempo para que todos os alto-falantes emitissem seus sinais fosse de 153 ms. Essa seria uma situação mais favorável para o caso de termos microfones em movimento.

Para todas as configurações, foram gravadas três realizações, de forma a obtermos confirmações dos resultados. Após realizadas as gravações no cenário real descrito, os sinais foram processados utilizando-se as mesmas rotinas das simulações, levando-se em conta a polarização dos TOFs de 0,76 ms.

Os resultados utilizando os *chirps* estão resumidos na Tabela 4.4 e os histogramas referentes às realizações do experimento se encontram ilustrados na Figura 4.21. Os histogramas mostram que, mesmo com as realizações sendo medidas uma seguida da outra, houve mudanças nos erros de estimação. Os erros médios mostram que, apesar das mudanças, o método LS já foi suficiente para obtermos bons resultados, enquanto os resultados reproduzidos com o método de seleção de TOF apresentaram pequena melhora.

Note que, os erros obtidos estão um pouco acima do erro estimado da medição manual das posições (≈ 1 cm), devendo-se considerar o fato de os alto-falantes terem diâmetro de mais de 7 cm.

Os resultados referentes ao experimento com sinais pseudoaleatórios estão resumidos na Tabela 4.5 e nos histogramas da Figura 4.22. Curiosamente, neste experimento foram obtidos exatamente os mesmos resultados para os dois métodos, e entre as realizações os resultados foram muito próximos.

Na Figura 4.23 estão os histogramas relativos ao experimento com sinais pseudoaleatórios e intervalo de 3 ms. É possível observar que houve dois *outliers* nas três realizações para o método LS. Utilizando-se a seleção de TOFs, os dois *outliers* foram corrigidos para as duas primeiras realizações, enquanto só um deles

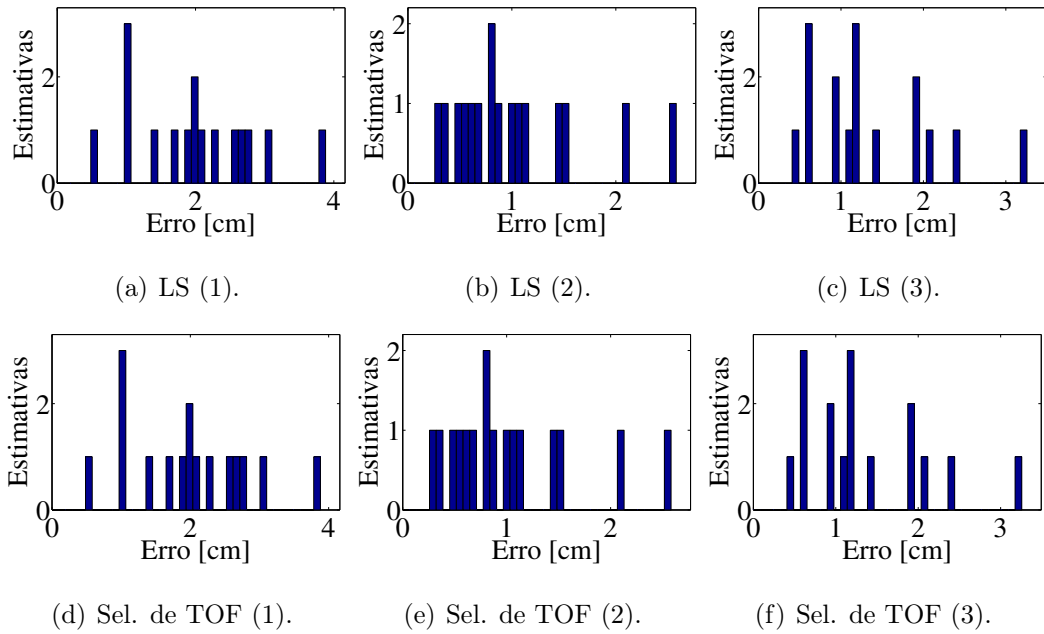


Figura 4.21: Histogramas das realizações do experimento com *chirps* reproduzidos individualmente.

Tabela 4.4: Resumo dos resultados do experimento com *chirps* reproduzidos individualmente.

	Erro médio [cm] (desvio-padrão [cm])	
	LS	Sel. de TOF
Realização 1	2,0 (0,9)	2,0 (0,9)
Realização 2	1,4 (0,6)	1,0 (0,6)
Realização 3	1,4 (0,8)	1,0 (0,8)

Tabela 4.5: Resumo dos resultados do experimento com *chirps* pseudoaleatórios reproduzidos individualmente.

	Erro médio [cm] (desvio-padrão [cm])	
	LS	Sel. de TOF
Realização 1	1,4 (0,9)	1,4 (0,9)
Realização 2	1,4 (0,8)	1,4 (0,8)
Realização 3	1,4 (0,8)	1,4 (0,8)

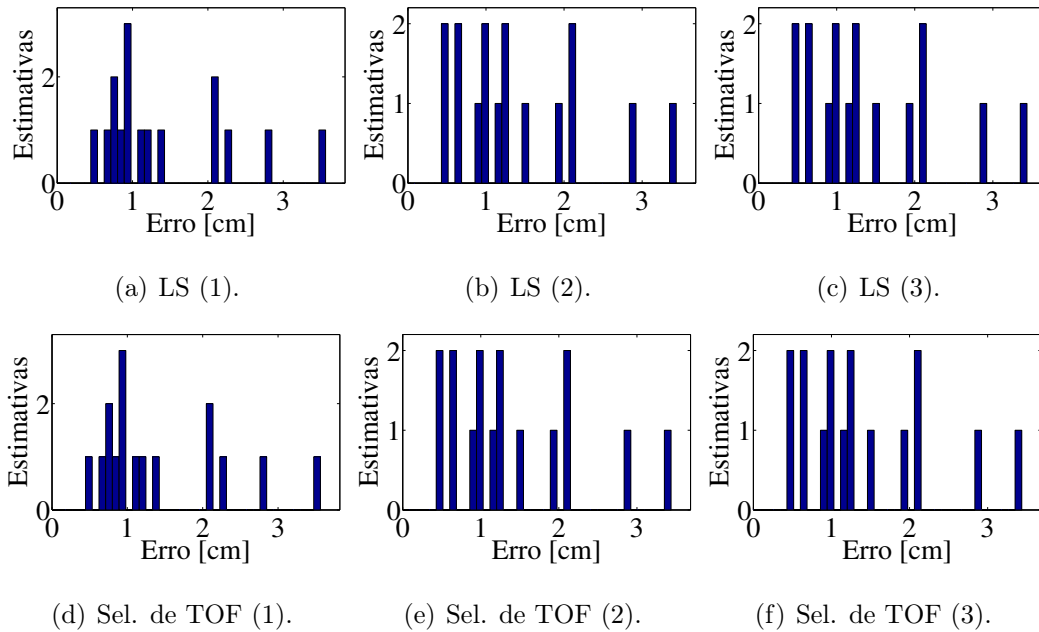


Figura 4.22: Histogramas das realizações do experimento com *chips* pseudoaleatórios reproduzidos individualmente.

foi corrigido na terceira realização. Os erros médios estão dispostos na Tabela 4.6, onde é possível observar que quando os *outliers* foram corrigidos, os erros médios obtidos foram semelhantes aos obtidos nos outros experimentos.

Esses resultados mostram que é possível utilizar intervalos curtos entre os sinais reproduzidos, o que só é viável utilizando-se sinais distintos entre as caixas. Com isso, o uso de sinais pseudoaleatórios traz grandes vantagens para o sistema de localização de sensores acústicos implementado e testado.

Tabela 4.6: Resumo dos resultados do experimento com *chips* pseudoaleatórios reproduzidos com intervalo de 3 ms.

	Erro médio [cm] (desvio-padrão [cm])	
	LS	Sel. de TOF
Realização 1	127,1 (342,5)	1,5 (0,9)
Realização 2	127,3 (343,0)	1,4 (0,9)
Realização 3	127,3 (343,0)	72,4 (284,8)

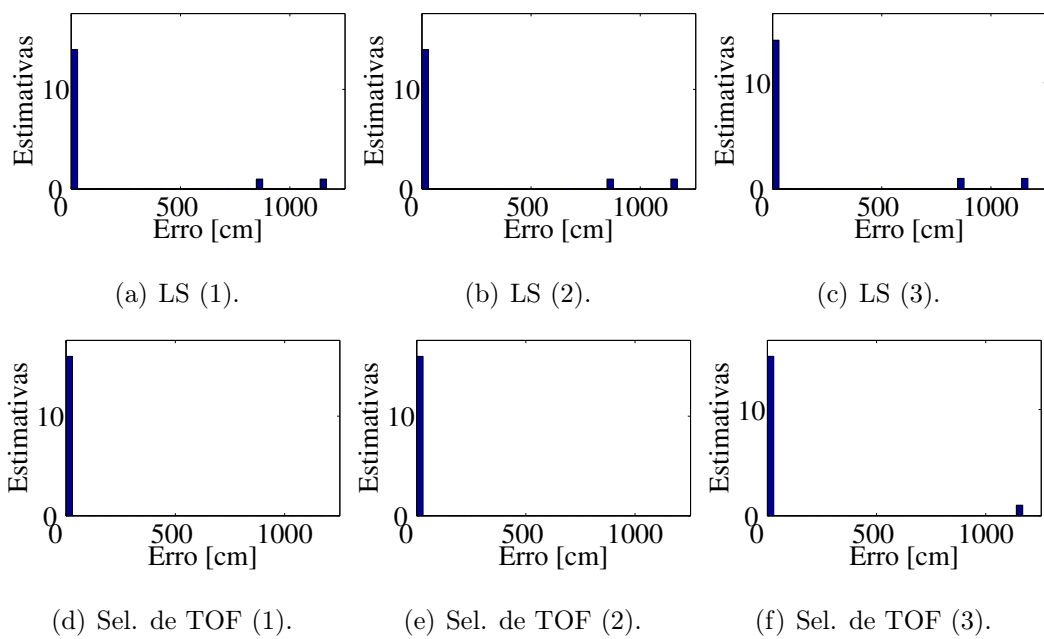


Figura 4.23: Histogramas das realizações do experimento com *chips* pseudoaleatórios reproduzidos com intervalo de 3 ms.

Capítulo 5

Conclusões e Trabalhos Futuros

5.1 Conclusões

Durante o período dedicado a este projeto, foi possível realizar o estudo teórico e a experimentação prática de um método da literatura para localização de sensores acústicos. Na abordagem adotada, foram feitas adaptações do método descrito em [2], de forma a utilizá-lo apenas para localização de microfones e a nos permitir propor um novo método [8] que possibilita corrigir estimativas de TOF. Os métodos foram implementados e simulados, de maneira a obterem-se dados do comportamento do sistema quando submetido a diversos ambientes, com diferentes dimensões, tempos de reverberação e níveis de ruído de fundo. Após a realização das simulações, foram feitos experimentos práticos que permitiram avaliar o funcionamento do sistema quando submetido a condições mais realísticas. Os resultados obtidos nos permitem afirmar que o método proposto se mostrou eficaz, tornando o sistema mais robusto a reverberação severa e reduzindo sua complexidade computacional por eliminar a necessidade do uso do método de otimização não-linear, o que favorece seu uso em aplicações práticas. O uso de *chips* pseudoaleatórios no lugar de *chirps* também contribuiu para ampliar o uso do sistema, reduzindo drasticamente o tempo total de gravação dos sinais por possibilitar o uso de sinais distintos entre os alto-falantes.

5.2 Trabalhos Futuros

Como continuação imediata do trabalho realizado, podem ser feitos experimentos em outros ambientes reais com tempo de reverberação maior, de forma a testar a eficácia do método em relação a esta característica ambiental. Podem ser feitos também experimentos com diferentes sinais interferidores localizados, como sinais de fala, música etc. Dessa forma, pode-se averiguar a aplicabilidade do sistema em locais onde haja a presença destes sons ou pode-se buscar alguma forma de tornar o uso dos sinais emitidos mais agradável, acrescentando outros sinais propositalmente para mascarar o som dos sinais utilizados na localização.

Originalmente [2], para que fosse estimada a posição de um microfone, havia a necessidade de se conhecer a posição estimada de todos os microfones, como forma de possibilitar o uso do método de otimização não-linear descrito em 3.3.3. O método de seleção de TOFs 3.3.4 proposto neste projeto tornou possível a correção das estimativas de TOF sem que houvesse essa necessidade, de maneira a tornar os nós — dispositivos com microfones a serem localizados — mais independentes. Contudo, como todos os métodos de estimação de TOF consideram que o instante em que foram emitidos os sinais pelos alto-falantes é conhecido, ainda há a necessidade de comunicação dos nós com a central responsável pela emissão dos sinais.

Dessa forma, como trabalhos futuros, podem ser implementados métodos que utilizem a diferença dos tempos de vôo (do inglês *time-difference-of-flight* — TDOF) no lugar do TOF, tornando a estimação da posição dos microfones totalmente independente de qualquer comunicação. Com essa modificação, abre-se uma nova gama de possibilidades de experimentos, possibilitando o uso do sistema em diversos dispositivos, bastando terem armazenados os sinais emitidos e as posições dos alto-falantes previamente medidas.

Referências Bibliográficas

- [1] BENESTY, J., CHEN, J., HUANG, Y., *Microphone Array Signal Processing*. Berlin, Germany, Springer, 2010.
- [2] RAYKAR, V., DURAISWAMI, R., “Automatic position calibration of multiple microphones”. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'04)*, v. 4, pp. 69–72, May 2004.
- [3] BENESTY, J., SONDHI, M. M., HUANG, Y., *Springer Handbook of Speech Processing*. Berlin, Germany, Springer, 2007.
- [4] COBOS, M., MARTI, A., LOPEZ, J. J., “A Modified SRP-PHAT Functional for Robust Real-Time Sound Source Localization With Scalable Spatial Sampling”, *IEEE Signal Processing Letters*, v. 18, n. 1, pp. 71–74, January 2011.
- [5] DMOCHOWSKI, J. P., BENESTY, J., AFFES, S., “A Generalized Steered Response Power Method for Computationally Viable Source Localization”, *IEEE Transactions on Audio, Speech, and Language Processing*, v. 15, n. 8, pp. 2510–2526, November 2007.
- [6] BENESTY, J., MAKINO, S., CHEN, J., *Speech Enhancement*. Berlin, Germany, Springer, 2010.
- [7] HENNECKE, M. H., FINK, G. A., “Towards acoustic self-localization of ad hoc smartphone arrays”. In: *Proceedings of the Joint Workshop on Hands-free Speech Communication and Microphone Arrays*, pp. 127–132, Edinburgh, UK, June 2011.
- [8] MARTINS, W., NUNES, L., HADDAD, D., *et al.*, “Time-of-Flight Selection for Improved Acoustic Sensor Localization Using Multiple Loudspeakers”. In:

Anais do XXXI Simpósio Brasileiro de Telecomunicações (SBrT 2013), Fortaleza, Setembro 2013.

- [9] KUTTRUFF, H., *Acoustics*. Abingdon, UK, Taylor & Francis, 2007.
- [10] SABINE, W. C., *Collected Papers on Acoustics*. Harvard University Press, 1922.
- [11] NOTAROS, B. M., *Eletromagnetismo*. São Paulo, Brasil, Pearson, 2012.
- [12] EARGLE, J., *Handbook of Recording Engineering*. Los Angeles, USA, Springer, 2006.
- [13] EARGLE, J., *The Microphone Book*. Burlington, USA, Focal Press, 2005.
- [14] DONOHUE, K. D., “<http://www.engr.uky.edu/~donohue/>”.
- [15] EYRING, C. F., “Reverberation time in ‘dead’ rooms”, *Journal of the Acoustical Society of America*, v. 1, pp. 217–241, January 1930.
- [16] DPA Microphones, <http://www.dpamicrophones.com>, 2013, Visitado em 08/06/2013.
- [17] ORB Audio, <http://www.orbaudio.com>, 2013, Visitado em 08/06/2013.
- [18] RME, http://www.rme-audio.de/en_index.php, 2013, Visitado em 08/06/2013.
- [19] DENON, <http://usa.denon.com/us/pages/home.aspx>, 2013, Visitado em 08/06/2013.