

Ricardo A S Carvalho

# **Estimação da Direção de Chegada de Fontes Sonoras**

**Brasil**

**2014**



Ricardo A S Carvalho

## **Estimação da Direção de Chegada de Fontes Sonoras**

Projeto de Iniciação Científica para o curso  
de Engenharia Elétrica da Escola Politécnica  
da USP

Universidade de São Paulo – USP

Escola Politécnica

Orientador: Prof. Dr. Cristiano Panazio

Brasil

2014



# Lista de ilustrações

Figura 1 – Modelo Teórico do Arranjo de Microfones . . . . .	9
Figura 2 – Ângulos Máximos em Função da Distância entre os Microfones . . . . .	13
Figura 3 – (a) Ângulos observáveis com $D = 10,8\text{cm}$ (b) Ângulos observáveis com $D = 54\text{cm}$ . . . . .	14
Figura 4 – (a) Correlação Cruzada com resultado próximo do esperado. (b) Correlação Cruzada com resultado distante do esperado . . . . .	15
Figura 5 – (a) Histograma de simulação feita com janela de 441 amostras. (b) Histograma de simulação feita com janela de 1050 amostras. . . . .	16
Figura 6 – (a) Autocorrelação de um sinal branco (b) Autocorrelação de uma senóide de 8820hz . . . . .	17
Figura 7 – Correlação cruzada de sinais senoidais com diferentes frequências. Ângulo de chegada de $0^\circ$ e distância de 8cm entre os microfones. . . . .	18
Figura 8 – Correlação cruzada de sinais senoidais com diferentes frequências. Ângulo de chegada de $0^\circ$ e distância de 20cm entre os microfones. . . . .	19
Figura 9 – Gráfico de correlação cruzada para sinal colorido. . . . .	19
Figura 10 – Histograma de 4214 realizações para sinal colorido acrescido de ruído. . . . .	20
Figura 11 – (a) Gráfico de correlação cruzada com resultado próximo de $30^\circ$ . (b) Gráfico de correlação cruzada com resultado próximo de $60^\circ$ . . . . .	20
Figura 12 – Histogramas de teste com sinal de voz sem ruído, ângulo de chegada de $30^\circ$ e distância entre os microfones de 15cm. Janela utilizada de $2^8$ amostras. (a) Correlação Cruzada. (b) GCC-PHAT . . . . .	22
Figura 13 – Histogramas de teste com sinal de voz sem ruído, ângulo de chegada de $30^\circ$ e distância entre os microfones de 15cm. Janela utilizada de $2^{12}$ amostras. (a) Correlação Cruzada. (b) GCC-PHAT . . . . .	23
Figura 14 – Histogramas de teste com sinal de voz com ruído, ângulo de chegada de $30^\circ$ e distância entre os microfones de 15cm. Janela utilizada de $2^{12}$ amostras. SNR de 4.8dB. (a) Correlação Cruzada. (b) GCC-PHAT . . . . .	23
Figura 15 – Histogramas de teste com sinal branco sem ruído, ângulo de chegada de $30^\circ$ e distância entre os microfones de 15cm. Janela utilizada de $2^8$ amostras. (a) Correlação Cruzada. (b) GCC-PHAT . . . . .	24
Figura 16 – Divisor de Áudio . . . . .	25
Figura 17 – Arranjo de Microfones e Aplicativo de Celular . . . . .	25
Figura 18 – Esboço da sala utilizada nos testes . . . . .	26

Figura 19 – Histogramas de teste com sinal de voz, ângulo de chegada de $-15^\circ$ , distância ao arranjo de 1,5m e distância entre os microfones de 15cm. 330 janelas utilizadas para estimação. SNR média das janelas de 70,12dB. (a) Correlação Cruzada. (b) GCC-PHAT . . . . .	26
Figura 20 – Histogramas de teste com sinal de voz, ângulo de chegada de $-30^\circ$ , distância ao arranjo de 1,5m e distância entre os microfones de 15cm. 330 janelas utilizadas para estimação. SNR média das janelas de 70,27dB. (a) Correlação Cruzada. (b) GCC-PHAT . . . . .	27
Figura 21 – Histogramas de teste com sinal de voz, ângulo de chegada de $-60^\circ$ , distância ao arranjo de 1,5m e distância entre os microfones de 15cm. 327 janelas utilizadas para estimação. SNR média das janelas de 69,30dB. (a) Correlação Cruzada. (b) GCC-PHAT . . . . .	27
Figura 22 – Histogramas de teste com sinal de voz, ângulo de chegada de $-30^\circ$ , distância ao arranjo de 1,5m e distância entre os microfones de 30cm. 351 janelas utilizadas para estimação. SNR média das janelas de 68,54dB. (a) Correlação Cruzada. (b) GCC-PHAT . . . . .	28
Figura 23 – Histogramas de teste com sinal de voz, ângulo de chegada de $-30^\circ$ , distância ao arranjo de 1,5m e distância entre os microfones de 60cm. 336 janelas utilizadas para estimação. SNR média das janelas de 68,06dB. (a) Correlação Cruzada. (b) GCC-PHAT . . . . .	29
Figura 24 – Histogramas de teste com sinal de voz, ângulo de chegada de $-30^\circ$ , distância ao arranjo de 1.0m e distância entre os microfones de 15cm. 348 janelas utilizadas para estimação. SNR média das janelas de 71,97dB. (a) Correlação Cruzada. (b) GCC-PHAT . . . . .	30
Figura 25 – Histogramas de teste com sinal de voz, ângulo de chegada de $-30^\circ$ , distância ao arranjo de 2.0m e distância entre os microfones de 15cm. 331 janelas utilizadas para estimação. SNR média das janelas de 69,61dB. (a) Correlação Cruzada. (b) GCC-PHAT . . . . .	30
Figura 26 – Histogramas de teste com sinal branco, ângulo de chegada de $-30^\circ$ , distância ao arranjo de 1,5m e distância entre os microfones de 15cm. 525 janelas utilizadas para estimação. SNR média das janelas de 71,18dB. (a) Correlação Cruzada. (b) GCC-PHAT . . . . .	31

# Sumário

<b>1</b>	<b>INTRODUÇÃO, METODOLOGIA E OBJETIVOS</b>	<b>7</b>
<b>2</b>	<b>ESTIMAÇÃO DO TEMPO DE CHEGADA</b>	<b>9</b>
<b>2.1</b>	<b>Modelo Teórico</b>	<b>9</b>
<b>2.2</b>	<b>Correlação Cruzada</b>	<b>10</b>
2.2.1	Implementação em Tempo Discreto	11
<b>2.3</b>	<b>Análise de Limitações</b>	<b>12</b>
2.3.1	Ângulo de chegada e Distância entre os Microfones	12
2.3.2	Tamanho da janela de cálculo e Presença de Ruído	14
2.3.3	Densidade Espectral de Potência do Sinal	16
<b>2.4</b>	<b>GCC-Phat</b>	<b>21</b>
<b>2.5</b>	<b>Comparações Teóricas</b>	<b>22</b>
<b>2.6</b>	<b>Testes Práticos</b>	<b>24</b>
2.6.1	Ângulo de Chegada	26
2.6.2	Distância Entre os Microfones	28
2.6.3	Distância ao Arranjo	29
2.6.4	Considerações quanto a sinal branco	31
<b>2.7</b>	<b>Conclusões</b>	<b>32</b>
	<b>Referências</b>	<b>33</b>



# 1 Introdução, Metodologia e Objetivos

A estimação da direção de chegada tem diversas aplicações. Entre as várias, podem ser citadas o sonar (do inglês, SOund NAvigation and Ranging) [7], que visa a detecção de cardumes ou embarcações e até mesmo mapeamento do solo oceânico. Também é possível encontrar aplicações da técnica em aeroportos, onde ela é utilizada para identificar empresas aéreas cujas aeronaves ultrapassem o limite de ruído permitido [3]. E ainda, como um último exemplo, a técnica está presente em âmbitos militares, onde um aparelho pode ser utilizado para localizar atiradores [2]. No caso do tema desta iniciação científica, a técnica é utilizada para encontrar um interlocutor, visando apontar uma webcam para a sua posição, tendo como aplicações vídeo-conferências, gravação de aulas, entre outras. O objetivo é, portanto, descobrir a posição azimutal da fonte sonora, ou seja, o ângulo de chegada.

Para alcançar tal objetivo, se utiliza um arranjo de sensores, no caso, microfones, para a captação dos sinais, que posteriormente serão tratados com técnicas de processamento de sinais. Dependendo da configuração do arranjo, diversas informações acerca da fonte, como posição azimutal, elevação e distância em relação ao arranjo podem ser medidas. Quanto à técnica utilizada, como se deseja obter a posição da fonte, uma que se mostra mais acessível, tanto em termos de exigência computacional quanto em facilidade de implementação, além de entregar resultados satisfatórios para a aplicação, é a da estimação da diferença de tempo de chegada através da correlação cruzada que, posteriormente neste trabalho, foi comparada a um de seus aprimoramentos, o GCC-PHAT [1].

Visando a implementação e análise da técnica, primeiramente, neste trabalho, definiu-se um modelo teórico para fontes distantes, relacionando o ângulo de chegada com a diferença de tempo de chegada entre dois microfones, a distância entre eles e a velocidade do som no ar. Com o modelo pronto, foi feita uma descrição da correlação cruzada, da aplicação da ferramenta no domínio de tempo discreto e como ela será utilizada no processo de localização da fonte. Após isso, foi feito um estudo das variáveis limitantes do experimento, através de análises teóricas e simulações em MatLab. Por fim, testes práticos preliminares foram realizados, para que sua precisão fosse testada em um ambiente real.

Com os resultados dos testes preliminares em mãos, passamos à implementação do GCC-PHAT, que se baseou no mesmo modelo teórico da TDOA utilizado para o método padrão. O código do MatLab foi praticamente o mesmo, com algumas modificações para se trabalhar no domínio da frequência. Com o GCC implementado, comparações teóricas, por meio de simulações, foram feitas entre a técnica padrão e a modificada. Por fim,

realizaram-se testes práticos visando analisar ambas as técnicas quando expostas a sinais de voz e comparar os resultados, para que se pudesse chegar a conclusão de qual técnica é mais indicada para o objetivo proposto.

## 2 Estimação do Tempo de Chegada

### 2.1 Modelo Teórico

Neste trabalho, serão estabelecidas certas simplificações visando a criação de um modelo teórico para que, posteriormente, este possa ser utilizado para realizar simulações, e comparar com os resultados práticos. Posto isso, a fonte sonora será considerada como pontual, permitindo que a posição que se deseja encontrar seja perfeitamente apontável, e que a distância desta aos microfones é grande o suficiente para que as ondas por ela emitidas se comportem como planas ao atingir o arranjo. Com isso em mente, a figura 1 descreve o modelo geométrico de um arranjo de dois microfones.

Sabe-se que a distância percorrida pela onda sonora entre o instante que ela atinge o primeiro microfone e o instante em que ela atinge o segundo é dada por:

$$d = v \times \tau, \quad (2.1)$$

em que  $v$  é a velocidade do som no ar e  $\tau$  é a diferença de tempo entre o instante que o sinal chega no microfone 1 ( $\tau_1$ ) e no microfone 2 ( $\tau_2$ ), ou seja,  $\tau = \tau_1 - \tau_2$ .

Também, tem-se que essa mesma distância é obtida através da expressão

$$d = D \sin(\theta), \quad (2.2)$$

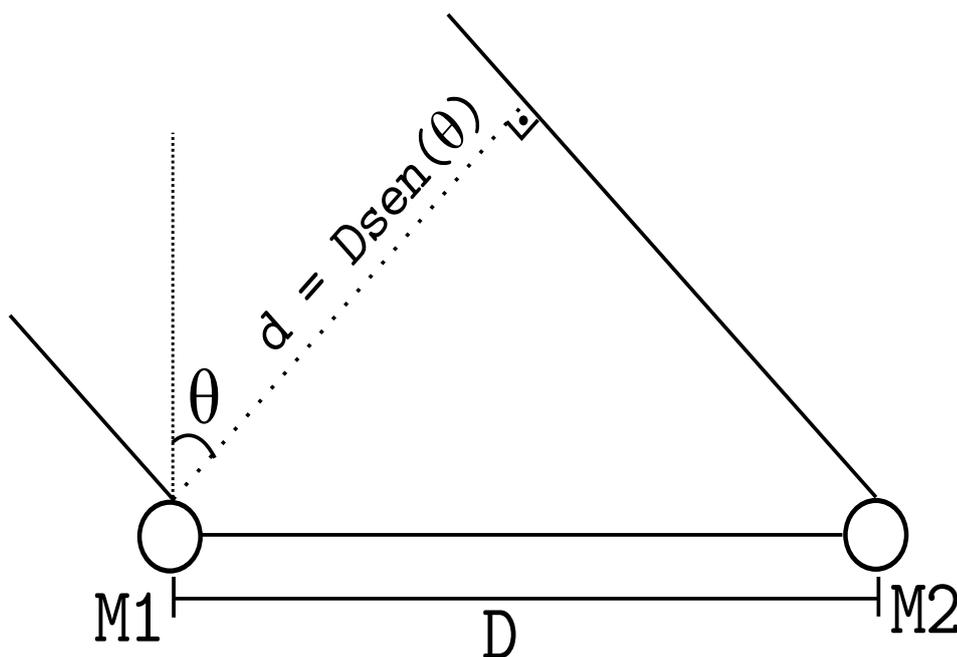


Figura 1 – Modelo Teórico do Arranjo de Microfones

na qual  $D$  é a distância entre os dois microfones e  $\theta$  é o ângulo formado entre a fonte e a normal em relação arranjo de microfones. Portanto, têm-se que

$$\text{sen}(\theta) = \frac{v \times \tau}{D}. \quad (2.3)$$

Logo, conclui-se que, para obter o ângulo  $\theta$  relacionado a posição horizontal da fonte sonora, deve-se utilizar a expressão

$$\theta = \arcsen\left(\frac{v \times \tau}{D}\right) \quad (2.4)$$

Sendo assim, para conseguir determinar a direção de chegada, é necessário estimar a diferença de tempo de chegada  $\tau$  entre os sinais recebidos pelo microfone 1 e pelo microfone 2. Pode-se fazer isso utilizando o conceito de correlação cruzada [1].

## 2.2 Correlação Cruzada

A correlação cruzada pode ser vista como o produto interno entre dois sinais. Considerando dois sinais reais de potência deterministicss,  $\check{x}_1(t)$  e  $\check{x}_2(t)$  têm-se a expressão do produto interno na equação

$$\langle \check{x}_1(t), \check{x}_2(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \check{x}_1(t) \check{x}_2(t) dt. \quad (2.5)$$

Já a correlação cruzada entre dois sinais aleatórios é definida como [1]

$$E[x_1(t)x_2(t)], \quad (2.6)$$

na qual  $E[\cdot]$  é o operador esperança e  $x_1(t)$  e  $x_2(t)$  são dois processos estocásticos, que, no caso, são os sinais capturados pelos microfones 1 e 2, respectivamente. Para se calcular a equação (2.6), uma das opções seria através das funções densidade de probabilidade de  $x_1(t)$  e de  $x_2(t)$ . Entretanto, caso  $x_1(t)$  e  $x_2(t)$  possam ser considerados processos ergódicos e estacionários, a equação (2.6) pode ser calculada tal como (2.5).

No caso do experimento, o sinal obtido no segundo microfone é uma versão atrasada no tempo do sinal do primeiro microfone, de modo que

$$x_2(t) = x_1(t - \tau). \quad (2.7)$$

Substituindo a expressão (2.7) em (2.5) obtém-se

$$\langle x_1(t), x_1(t - \tau - r) \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x_1(t) x_1(t - \tau) dt. \quad (2.8)$$

Considerando que os ruídos nos microfones são descorrelacionados entre si e entre os sinais recebidos, e observando a semelhança com a definição de produto interno, pode-se

deduzir, utilizando a desigualdade de Cauchy-Schwarz, que o pico da expressão (2.8) é obtido quando  $\tau = 0$  [5]. Entretanto, a única situação em que isso aconteceria é no caso em que a fonte estivesse numa posição perpendicular ao arranjo dos microfones. Para que se possa descobrir outras posições, introduz-se um atraso arbitrário  $r$ , de forma que o pico da correlação cruzada seja obtido quando  $\tau = -r$ . Portanto, obtém-se a correlação cruzada na forma de

$$R_{x_1x_1}(r) = \langle x_1(t), x_1(t - \tau) \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x_1(t)x_1(t - \tau - r)dt. \quad (2.9)$$

### 2.2.1 Implementação em Tempo Discreto

Para que seja possível realizar a análise dos sinais, primeiramente eles devem ser captados pelos microfones. A partir disso eles são amostrados com uma taxa  $fs$  fixa, passando de uma representação em tempo contínuo para tempo discreto. No domínio de tempo discreto, um intervalo de tempo  $I$  finito não contém infinitos pontos, mas sim um número  $N$  de pontos igual a

$$N = \left\lfloor \frac{I}{fs} \right\rfloor. \quad (2.10)$$

Tais pontos estão espaçados no tempo por um valor  $T$ , chamado de período de amostragem, que é igual ao inverso da frequência  $fs$ . Sendo assim, as funções discretas tem a forma  $x(nT)$  em que  $n$  é um número inteiro qualquer. Logo, a equação (2.9) é exibida no domínio do tempo discreto como

$$R_{x_1x_1}(k) = \langle x_1(nT), x_1(nT - \tau - kT) \rangle = \lim_{N \rightarrow \infty} \frac{1}{2N + 1} \sum_{n=-N}^N x_1(nT)x_1(nT - \tau - kT). \quad (2.11)$$

Sendo assim, o pico da equação (2.11) está no ponto onde  $kT$ , ou  $k/fs$ , é mais próximo de  $\tau$ , o que introduz uma possível fonte de erros no experimento, já que  $\tau$  é um número real, e  $k$  é um número inteiro. Deste modo, caso  $\tau$  não seja um múltiplo de  $T$ , o pico da equação (2.11) não coincidirá com o pico da equação (2.9), ocasionando um erro de estimação.

Considerando isso, deve-se reavaliar a expressão (2.4), de modo que ela esteja condizente com o domínio de tempo discreto. Seja  $k'$  o número inteiro que minimiza  $|\tau - kT|$ . Assim, com as devidas substituições, chega-se a expressão

$$\hat{\theta} \cong \arcsen\left(\frac{vk'}{Dfs}\right), \quad (2.12)$$

na qual  $\hat{\theta}$  só é igual ao ângulo real de chegada quando  $\tau$  é um múltiplo de  $k'/fs$ .

O possível erro devido a discretização do sinal é apenas uma das limitações que ocorrem na implementação de tal técnica. A seguir, serão analisadas outras importantes limitações que influenciam na precisão das estimativas do ângulo de chegada.

## 2.3 Análise de Limitações

Nas realizações da estimação de tempo de chegada, diversas variáveis podem afetar o desempenho do sistema como um todo, sendo, portanto, necessário verificar previamente os limites do modelo desenvolvido. Neste caso, as variáveis que afetam de forma predominante o resultado das estimações são: ângulo de chegada, distância entre os microfones, densidade espectral de potência do sinal, tamanho da janela e presença de ruído.

Nas subseções seguintes, as variáveis acima serão avaliadas, visando estimar sua influência no resultado final.

### 2.3.1 Ângulo de chegada e Distância entre os Microfones

A precisão da técnica empregada em determinar corretamente um ângulo é diretamente dependente da distância  $D$  entre os microfones. Considerando que para determinada distância  $D$  o máximo  $\tau$  detectável é obtido quando a fonte está posicionada a um ângulo de  $90^\circ$  em relação ao arranjo de microfones, têm-se que

$$\tau_{max} = \frac{D}{v}. \quad (2.13)$$

Como discutido na subseção 2.2.1, é necessário trabalhar com as equações no domínio do tempo discreto. Portanto, a equação (2.13) vira

$$k_{max} = \left\lfloor \frac{D \times fs}{v} \right\rfloor. \quad (2.14)$$

Neste caso, utiliza-se um arredondamento, já que  $k$  é inteiro, e o arredondamento é obrigatoriamente para baixo, já que utilizar na expressão (2.12) um valor maior que o obtido na expressão (2.14) faria com que o valor  $\hat{\theta}$  encontrado fosse complexo, e valores complexos não fazem sentido físico.

Por fim, ao substituir  $k_{max}$  da expressão (2.14) em (2.12) obtém-se o ângulo máximo possível de ser encontrado em função da distância  $D$  entre os microfones como sendo

$$\hat{\theta}_{max} = \arcsen \left( \frac{v}{D} \times \left\lfloor \frac{D \times fs}{v} \right\rfloor \times \frac{1}{fs} \right). \quad (2.15)$$

A expressão (2.15) serve de base para a figura 2, que apresenta um gráfico dos ângulos máximos em função da distância entre os microfones.

Analisando a figura 2, em conjunto com a expressão (2.15), duas conclusões são obtidas, ambas relacionadas ao fato da técnica operar no mundo discreto. A primeira é que apesar de teoricamente o ângulo máximo ser de  $90^\circ$ , em momento algum esse ângulo será alcançado para distâncias entre  $0m$  e  $1,5m$ . A segunda é o fato de o ângulo máximo apresentar discontinuidades, explicado pela equação (2.15). Isto se deve ao fato de que

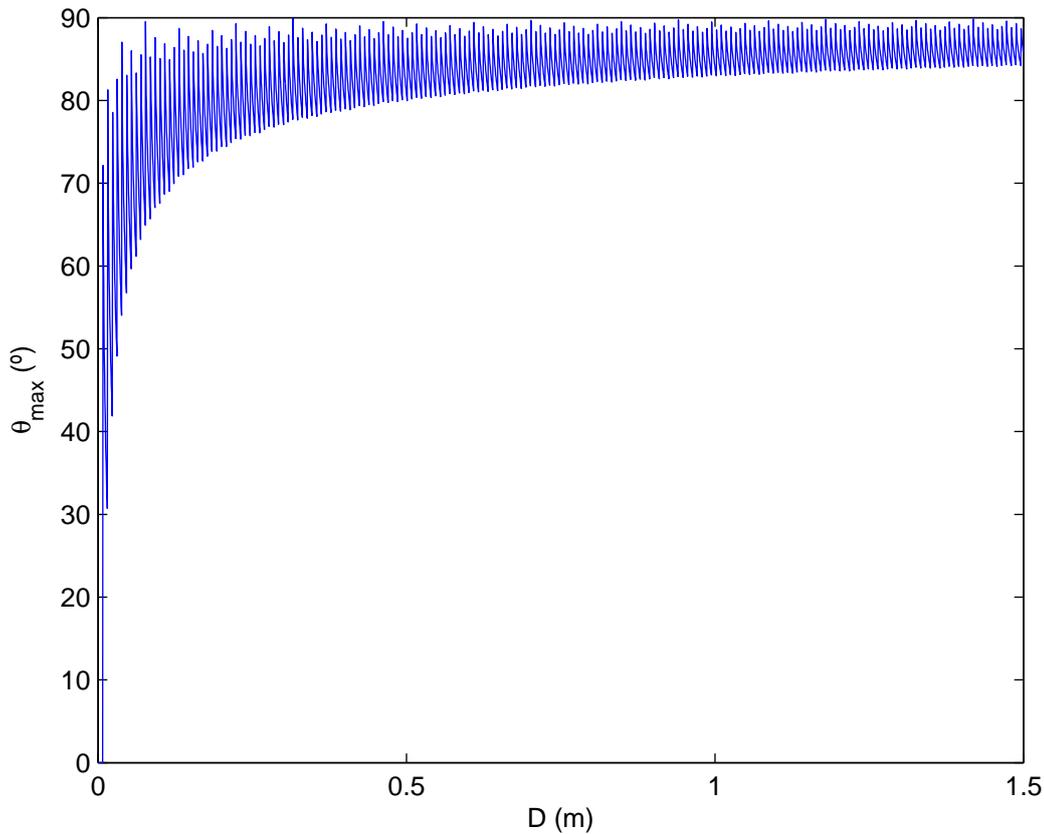


Figura 2 – Ângulos Máximos em Função da Distância entre os Microfones

quando aumenta-se a distância  $D$ , em intervalos definidos, o arredondamento da expressão (2.14) gerará o mesmo número e, portanto, o resultado de (2.15) diminuirá, e, a partir de determinadas distâncias, ocorre um salto quando o resultado da equação (2.14) aumenta em uma unidade. Na prática, pequenas variações de  $D$  na montagem experimental podem levar a erros consideráveis para fontes próximas do ângulo máximo detectável.

Também é preciso levar em consideração quais ângulos podem de fato ser achados. A expressão (2.12) relaciona os valores  $\hat{\theta}$  em função de  $k'$ , e a expressão (2.14) apresenta o  $k_{\max}$  em função de  $D$ . A partir disso, foi gerada a figura 3, que apresenta uma comparação entre dois sistemas, um com microfones espaçados por  $10,8\text{cm}$  e outro por  $54\text{cm}$ , ambas utilizando  $f_s$  como sendo igual a  $44100\text{Hz}$ , a taxa de amostragem mais comum e também a maior possível em uma grande parte das placas de som comerciais, de modo que ambas configurações tenham o mesmo  $\hat{\theta}_{\max}$ .

É possível perceber através da figura 3 que um maior espaçamento apresenta um número muito maior de ângulos no mesmo intervalo, e, por conseguinte, uma maior precisão. Por exemplo, considere uma fonte a um ângulo de  $63,5^\circ$ . Na primeira configuração,  $D = 10,8\text{cm}$ , os ângulos mais próximos são  $59^\circ$  e  $68^\circ$ , o que representa um erro de  $\pm 4,5^\circ$ . Já na segunda configuração, com  $D = 54\text{cm}$ , o mais próximo é  $64^\circ$ , o que representa um

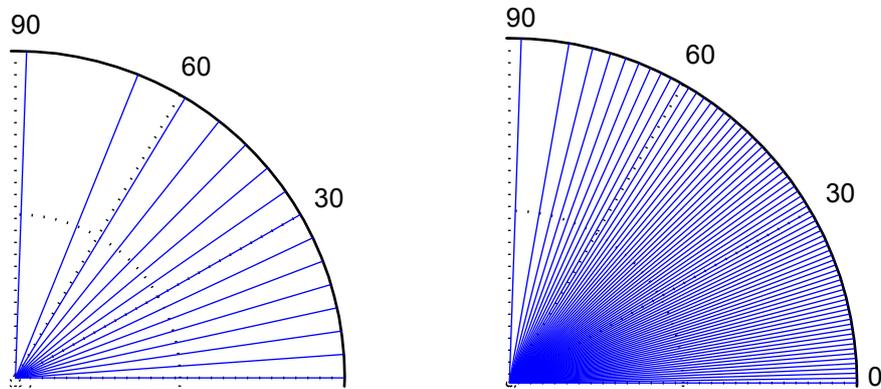


Figura 3 – (a) Ângulos observáveis com  $D = 10,8\text{cm}$  (b) Ângulos observáveis com  $D = 54\text{cm}$

erro de apenas  $0,5^\circ$ . Por fim, é também visível o fato de que quanto mais o ângulo se aproxima de  $90^\circ$ , mais espaçados são os ângulos observáveis, atrapalhando a precisão da técnica conforme a fonte se aproxima desse ângulo.

### 2.3.2 Tamanho da janela de cálculo e Presença de Ruído

Na correlação cruzada, o tamanho da janela determina quantas amostras serão levadas em conta em cada estimação do ângulo. Em geral, uma janela maior, com uma maior quantidade de amostras, exibe uma confiabilidade maior no resultado final da estimação. Para o experimento, o tamanho da janela também determina quanto tempo o sistema vai coletar sinais até que sejam feitos os cálculos da correlação.

O ruído, por sua vez, é considerado como qualquer som alheio ao experimento, ou distorção que advém dos equipamentos utilizados, *i.e.*, ruído térmico. O ruído é prejudicial ao experimento, no sentido em que a presença dele afeta os resultados dos cálculos da correlação cruzada fazendo com que haja uma possibilidade maior de que o pico da correlação não coincida com o verdadeiro pico.

Para exibir melhor a influência desses fatores no resultado final, foram feitas simulações que permitissem comparar situações. As simulações das figuras 4 e 5 foram

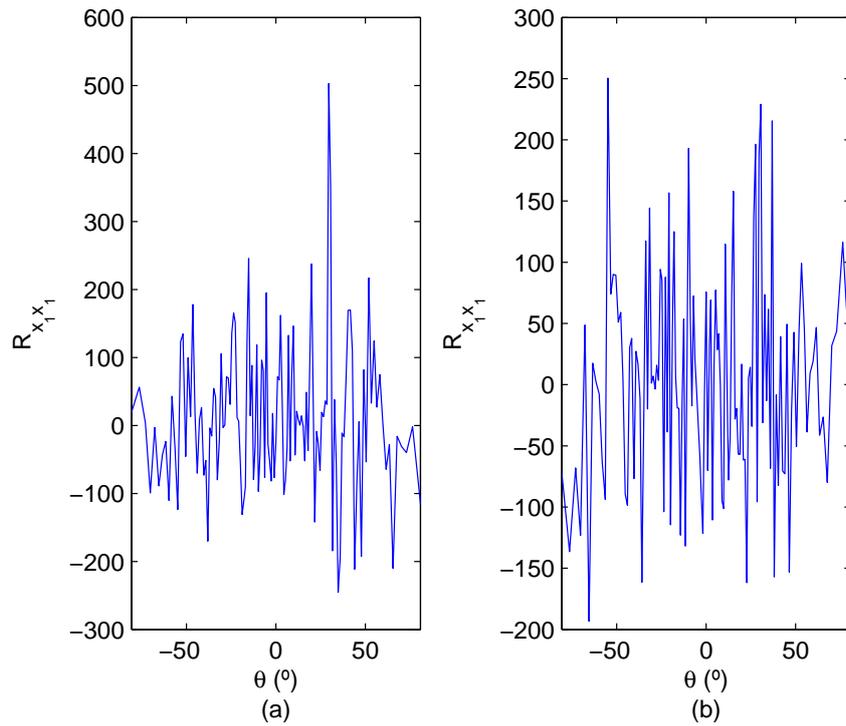


Figura 4 – (a) Correlação Cruzada com resultado próximo do esperado. (b) Correlação Cruzada com resultado distante do esperado

ambas feitas com sinais de entrada branco, ângulo de chegada de  $30^\circ$  e distância de  $50\text{cm}$  entre os microfones. O ruído adicionado em ambas é do tipo branco e a relação sinal-ruído é de aproximadamente  $4.8\text{dB}$ .

A figura 4 apresenta dois gráficos de correlação cruzada de diferentes estimações, ambas com janela de 441 amostras. Na figura 4(a) é exibido um gráfico cujo pico indicou um ângulo bem próximo do esperado, enquanto a figura 4(b) tem um pico distante, próximo de  $-55^\circ$ . Pode-se perceber que, ao contrário de quando se utiliza sinal de entrada sem ruído, como na figura 6(a), os resultados da correlação foram mais equilibrados, sem a existência de um pico tão predominante.

Já a figura 5 apresenta o histograma de duas simulações feitas com o mesmo nível de ruído, a primeira, figura 5(a), com janela de 441 amostras e a segunda, figura 5(b), com 1050 amostras. Na figura 5(a) em relação a figura 5(b) é perceptível a maior dispersão das estimativas.

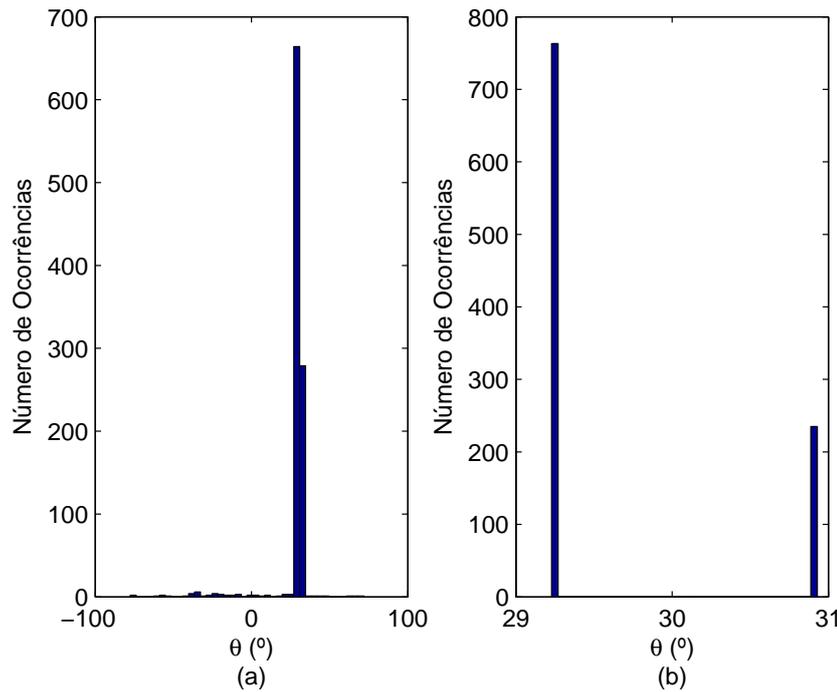


Figura 5 – (a) Histograma de simulação feita com janela de 441 amostras. (b) Histograma de simulação feita com janela de 1050 amostras.

### 2.3.3 Densidade Espectral de Potência do Sinal

A densidade espectral de potência de um sinal é a transformada de Fourier da função de autocorrelação (correlação cruzada entre o sinal e ele mesmo). Ela é expressa em  $W/Hz$ , determina a potência do sinal em cada faixa de frequência e ajuda a observar características periódicas dos sinais analisados. Um ruído branco, por exemplo, por ser aperiódico, tem uma densidade espectral de potência infinitamente constante. Já o oposto extremo, uma senóide, apresenta apenas um impulso no ponto de sua frequência.

A influência deste parâmetro tem relação direta com o resultado da correlação cruzada. Observando a figura 6, na qual figura 6(a) exibe a autocorrelação de um sinal branco, calculada a partir de uma janela de comprimento 441 amostras, e a figura 6(b) exibe a autocorrelação de uma senóide de  $8820Hz$ , percebe-se claramente que não há problema em determinar o único pico no caso do sinal branco. Já no caso senoidal, a função de autocorrelação apresenta múltiplos picos no intervalo considerado, de modo que haveria ambiguidades na determinação do ângulo de chegada.

A quantidade de picos da função de autocorrelação para funções senoidais varia com fatores como a distância entre os microfones, frequência de captação e frequência da senóide, como pode-se ver na figura 7, que apresenta a correlação cruzada para diferentes valores de frequência, utilizando uma onda com ângulo de chegada de  $0^\circ$  e distância de  $8cm$  entre os microfones.

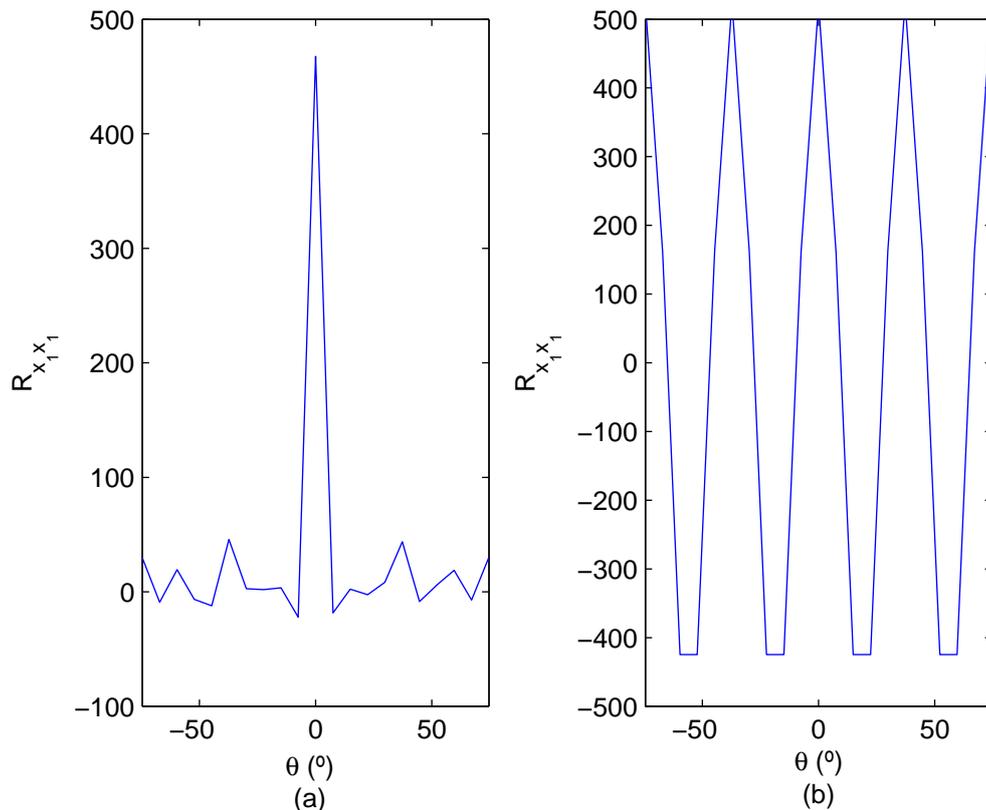


Figura 6 – (a) Autocorrelação de um sinal branco (b) Autocorrelação de uma senóide de 8820hz

Como forma de comparação, a figura 7 foi refeita, porém desta vez com uma distância de 20cm entre os microfones, configuração apresentada pela figura 8, de forma que o ângulo máximo fosse o mesmo. O que se pode observar é que, apesar de distâncias maiores, em geral, significarem uma precisão maior, elas também fazem com que, no caso de senóides, os múltiplos picos apareçam mais cedo. Na figura 7, picos múltiplos apareceram em torno de 8000Hz, e na figura 8, eles apareceram em torno de 2000Hz.

Em se tratando de voz, o sinal não é branco e pode conter componentes senoidais bastante pronunciadas, por isso é necessário que as situações intermediárias sejam consideradas. Para analisar o desempenho da técnica nessas situações, foi utilizado um filtro passa-faixas ideal com corte em 2500Hz e 3500Hz e que é usado para filtrar um sinal branco que chega a um Ângulo de 30° para um arranjo com separação de 30cm entre os microfones.

A figura 9 apresenta uma realização de correlação cruzada com janela de 1050 amostras para um sinal de entrada como o descrito acima. O que se pode observar na figura 9 é que para o sinal de entrada em questão, a autocorrelação se assemelha mais a uma senóide e menos a um impulso, e, portanto, qualquer perturbação (como, por exemplo, presença de ruído) pode fazer com que o verdadeiro pico seja mascarado. Além disso,

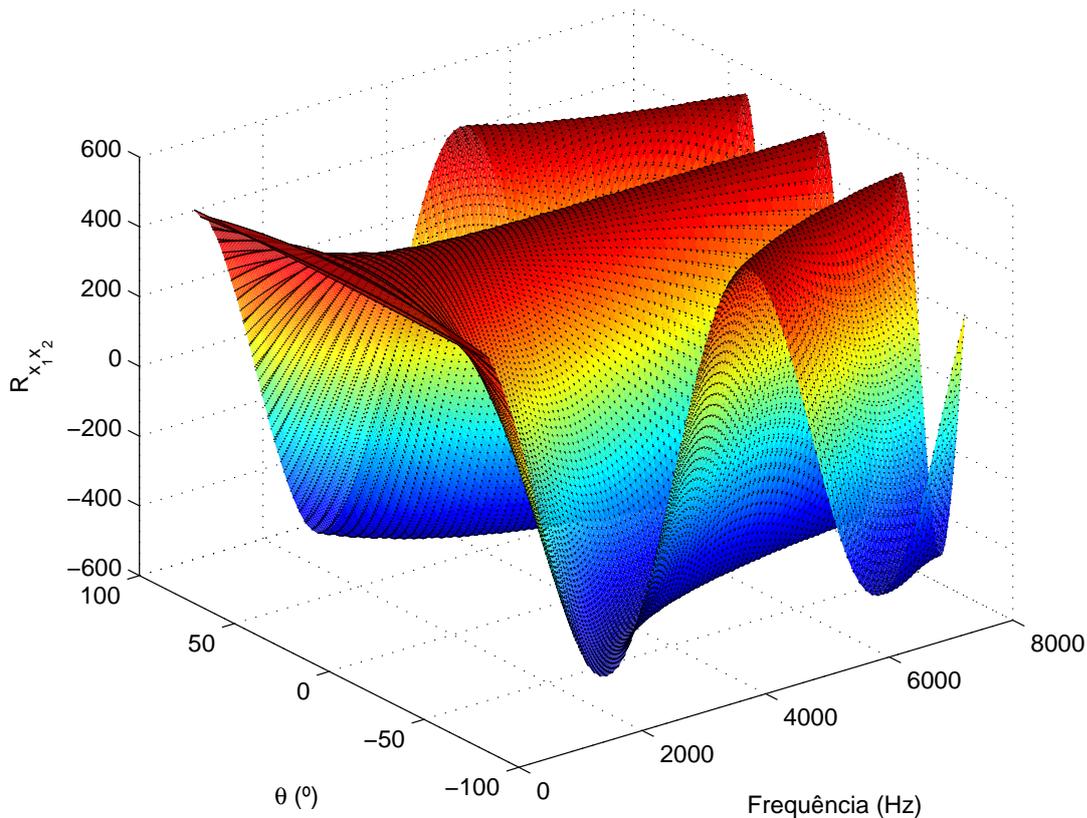


Figura 7 – Correlação cruzada de sinais senoidais com diferentes frequências. Ângulo de chegada de  $0^\circ$  e distância de 8cm entre os microfones.

observa-se a existência de picos próximos concentrados em torno de  $7^\circ$ ,  $60^\circ$  e em torno do ângulo original,  $30^\circ$ .

Sendo assim, para demonstrar o problema das possíveis perturbações, foi utilizado o mesmo sinal, porém desta vez acrescido de um ruído. O ruído tem as mesmas características do utilizado na subseção 2.3.2, ele é branco e a relação sinal-ruído é de -4.8dB. A figura 10 apresenta o histograma de 4214 realizações. É perceptível que o histograma apresenta valores próximos aos dos picos observados na figura 9.

Para exemplificar o efeito do ruído direto na correlação cruzada, a figura 11 apresenta dois gráficos de correlação cruzada resultantes da mesma configuração acima. A figura 11(a) exhibe o gráfico de uma realização cujo pico se deu em torno de  $30^\circ$ , enquanto a figura 11(b) exhibe um gráfico cuja realização apresentou um pico em torno de  $60^\circ$ .

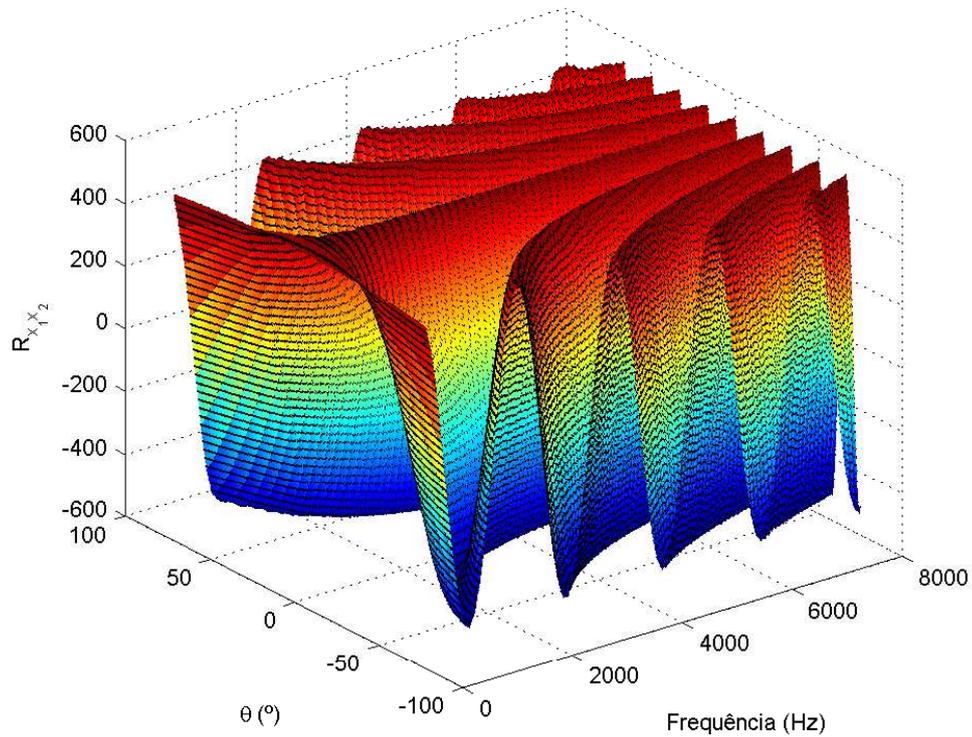


Figura 8 – Correlação cruzada de sinais senoidais com diferentes frequências. Ângulo de chegada de  $0^\circ$  e distância de 20cm entre os microfones.

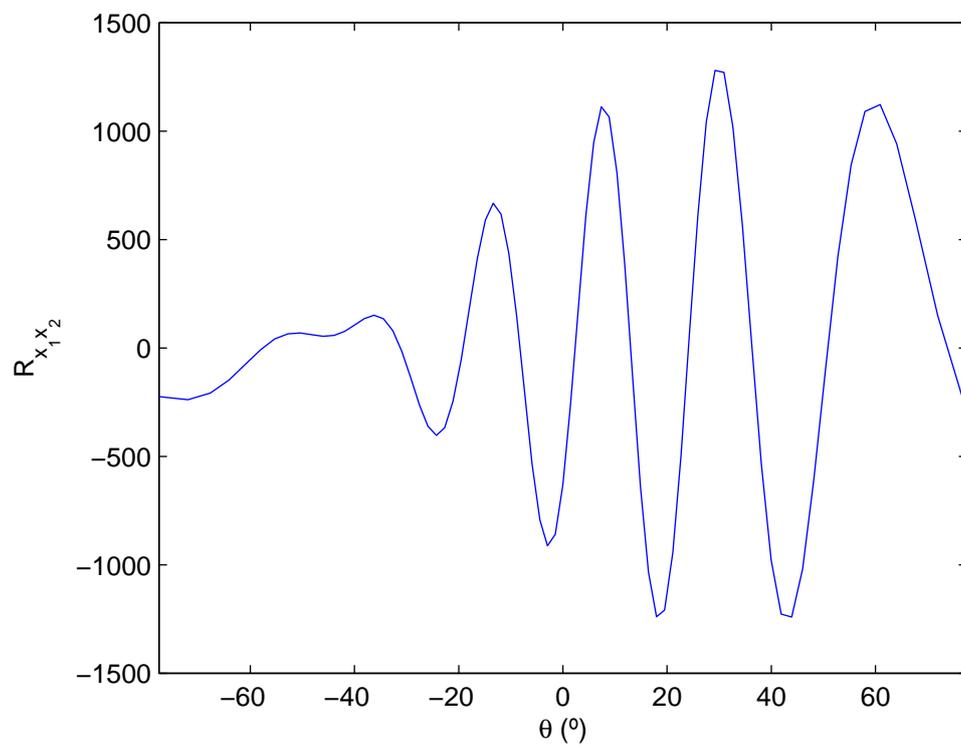


Figura 9 – Gráfico de correlação cruzada para sinal colorido.

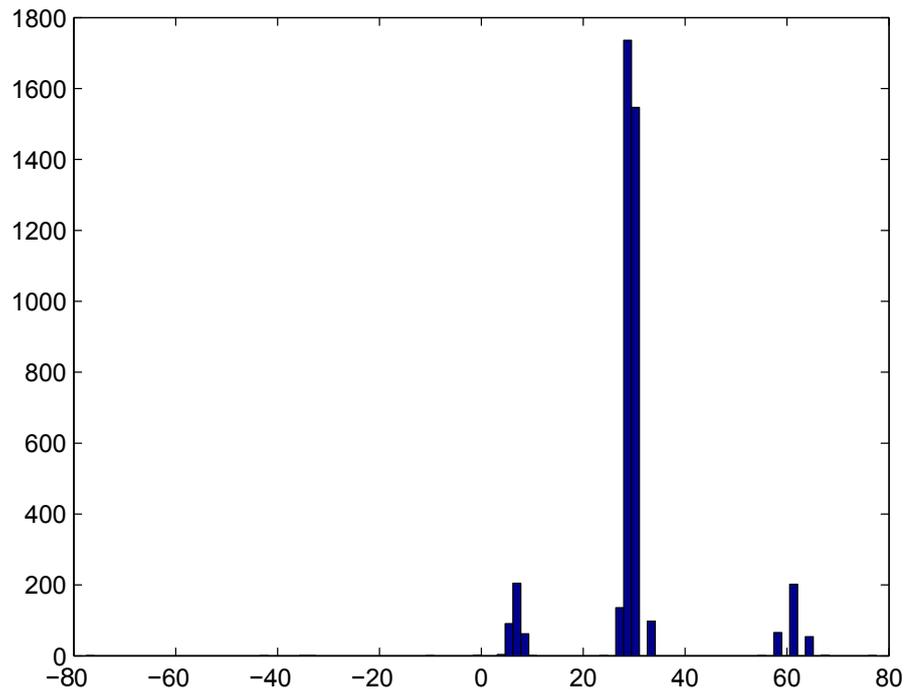


Figura 10 – Histograma de 4214 realizações para sinal colorido acrescido de ruído.

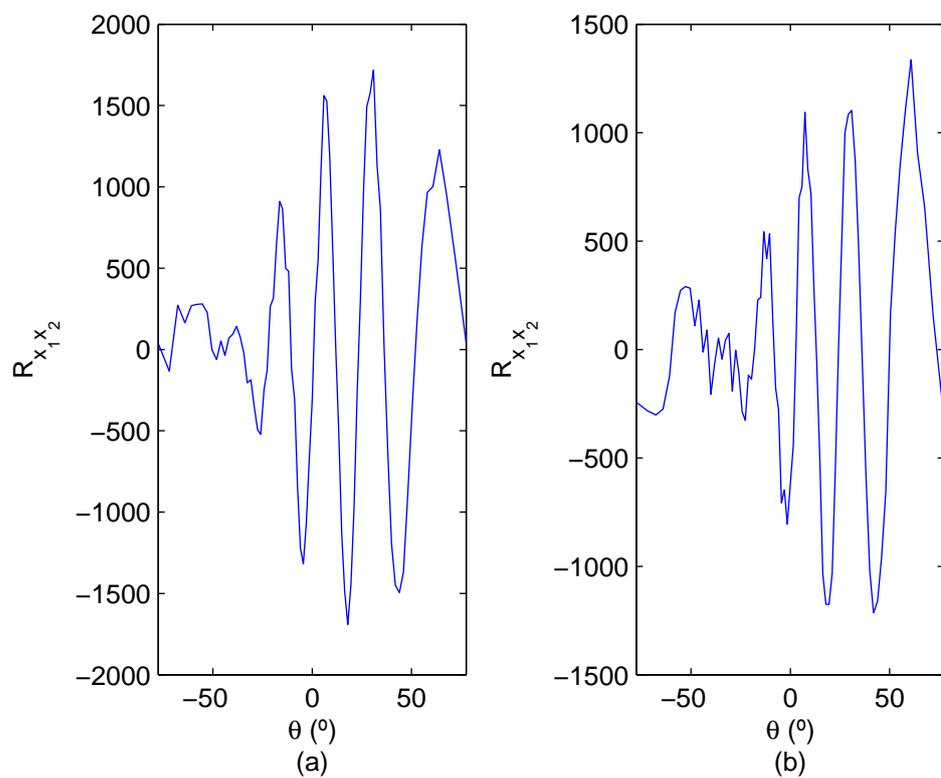


Figura 11 – (a) Gráfico de correlação cruzada com resultado próximo de  $30^\circ$ . (b) Gráfico de correlação cruzada com resultado próximo de  $60^\circ$ .

## 2.4 GCC-Phat

Nos anos 70, uma série de aprimoramentos para a técnica da correlação cruzada, englobados pela sigla GCC (Generalized Cross Correlation), foram propostos [4]. As técnicas GCC são caracterizadas pela implementação da correlação no domínio da frequência. Para tanto, se utilizam do conceito da transformada de Fourier.

A transformada de Fourier é uma transformada integral que, para o processamento digital de sinais, visa apresentar os sinais em função das suas amplitudes em cada faixa de frequência. Como trabalhamos em tempo discreto, utilizamos o conceito da DFT (Discrete Fourier Transform) ou transformada de Fourier em tempo discreto. A DFT de uma sequência  $p[n]$  de  $N$  valores discretos é expressa como

$$\mathcal{F}(p[n]) = P(f) = \lim_{N \rightarrow \infty} \sum_{n=-N}^N p[n] e^{-i2\pi f n}. \quad (2.16)$$

Sabe-se, pela dualidade tempo-frequência, que a correlação no domínio do tempo se torna uma multiplicação entre o primeiro sinal e o conjugado do segundo no domínio da frequência [6]. Sendo assim, temos que a transformada da expressão 2.11 é

$$\Phi(f) = \mathcal{F}(R_{x_1 x_1}(k)) = |X_1^2(f)| e^{-i2\pi f \tau}, \quad (2.17)$$

na qual a exponencial aparece devido à propriedade da transformada relacionada ao atraso  $\tau$  que um dos sinais sofre.

Como característica principal das propostas de GCC, é multiplicada uma função-peso  $\vartheta(f)$  à expressão 2.17. Por fim, após feitas as multiplicações, voltamos ao domínio do tempo para então achar o pico correspondente ao  $k'$  esperado, chegando, portanto, à expressão

$$k' = \operatorname{argmax} R_{x_1 x_1}^{GCC}[n] = \operatorname{argmax} \mathcal{F}^{-1}(\Phi(f)\vartheta(f)). \quad (2.18)$$

Existem diversas propostas para a escolha da função-peso, por exemplo, poderia-se escolher  $\vartheta(f)$  igual a 1, o que resultaria na técnica padrão da correlação cruzada explorada na seção 2.2. Entre as possibilidades, a proposta do PHAT (Phase Transform) é a que mais se adequa às necessidades do trabalho, já que foi feita *ad hoc* para explorar as dificuldades enfrentadas pela técnica padrão ao ser exposta a sinais de voz em ambientes reverberantes [8]. A proposta do GCC-PHAT é de acrescentar a função

$$\vartheta(f) = \frac{1}{|X_1^2(f)|}, \quad (2.19)$$

o que se pode notar é que substituindo 2.19 em 2.18, o que sobra é apenas a componente de fase, justamente aquela que interessa para a estimação de direção de tempo de chegada. Logo, em condições ideais, o que se têm ao realizar a transformada inversa é

$$\mathcal{F}^{-1}(\Phi(f)\vartheta(f)) = \int_{-\infty}^{\infty} e^{i2\pi f(n-\tau)}, \quad (2.20)$$

que resulta em um delta de dirac na posição de  $n = \tau$ .

Além disso, considerando o objetivo de se trabalhar com sinais de voz e o fato de que, para o GCC-PHAT, é necessário trabalhar com o sinal no domínio da frequência, fez-se uma mudança no algoritmo da TDOA, para que fossem consideradas apenas as componentes da transformada na faixa de frequência que vai de  $500Hz$  a  $3,5kHz$ . Esse corte foi implementado tanto no código do GCC-PHAT, quanto no da técnica padrão.

## 2.5 Comparações Teóricas

Com a introdução de uma técnica alternativa ao método padrão de correlação cruzada, fica necessário estabelecer uma comparação teórica entre as duas técnicas de forma que exista uma base de comparação para os resultados práticos. Sendo assim, foi feita uma comparação entre algumas das questões apresentadas na seção 2.3. As deduções teóricas de cada parâmetro se mantém para o GCC-PHAT como era para a técnica padrão, sendo assim, as comparações serão feitas através de resultados de testes.

O sinal utilizado nas figuras 12, 13 e 14 é o mesmo que foi reproduzido para obter os resultados teóricos discutidos posteriormente. É uma gravação de 1 minuto e 5 segundos aproximadamente, na qual é lido um trecho de um livro. Já o utilizado na figura 15 é um ruído branco gerado em MatLab. Para que a janela fosse considerada nos cálculos, foi imposto um limiar mínimo de potência da janela baseado na potência média do sinal todo e do ruído.

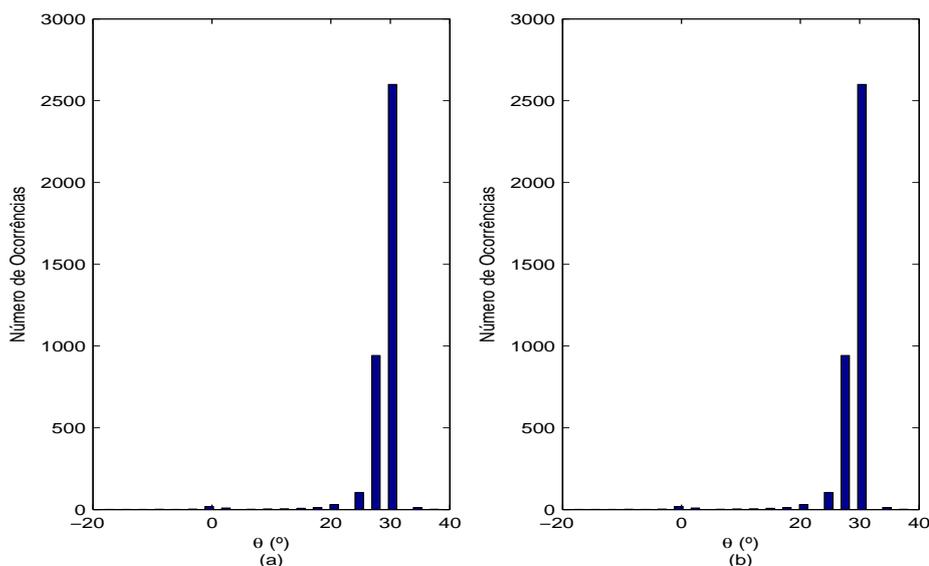


Figura 12 – Histogramas de teste com sinal de voz sem ruído, ângulo de chegada de  $30^\circ$  e distância entre os microfones de 15cm. Janela utilizada de  $2^8$  amostras. (a) Correlação Cruzada. (b) GCC-PHAT

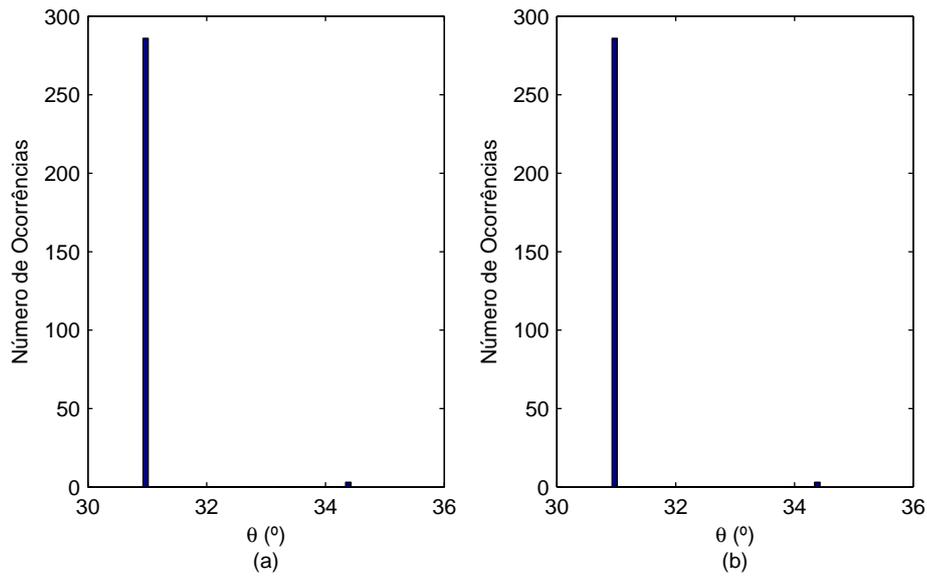


Figura 13 – Histogramas de teste com sinal de voz sem ruído, ângulo de chegada de  $30^\circ$  e distância entre os microfones de 15cm. Janela utilizada de  $2^{12}$  amostras. (a) Correlação Cruzada. (b) GCC-PHAT

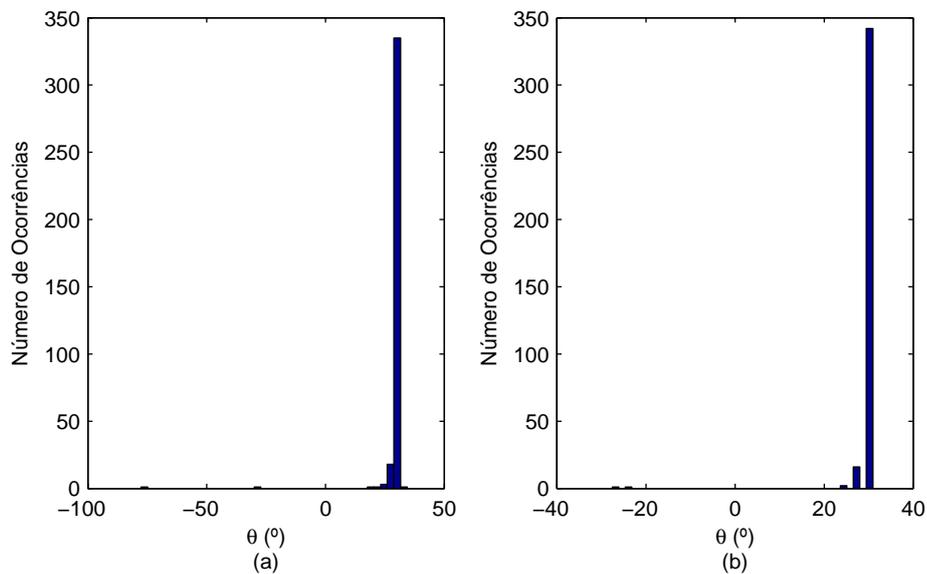


Figura 14 – Histogramas de teste com sinal de voz com ruído, ângulo de chegada de  $30^\circ$  e distância entre os microfones de 15cm. Janela utilizada de  $2^{12}$  amostras. SNR de 4.8dB. (a) Correlação Cruzada. (b) GCC-PHAT

Nas figuras 12 e 13 é possível notar a diferença de resultados em função do tamanho da janela utilizada, enquanto com  $2^8$  amostras, as técnicas exibem certas quantidades de resultados diferentes do esperado, com  $2^{12}$  apenas uma estimativa resultou diferente das outras, e mesmo essa está há apenas  $4^\circ$  do ângulo imposto. A diferença no número de resultados é devido à diferença no número de amostras utilizadas em cada janela.

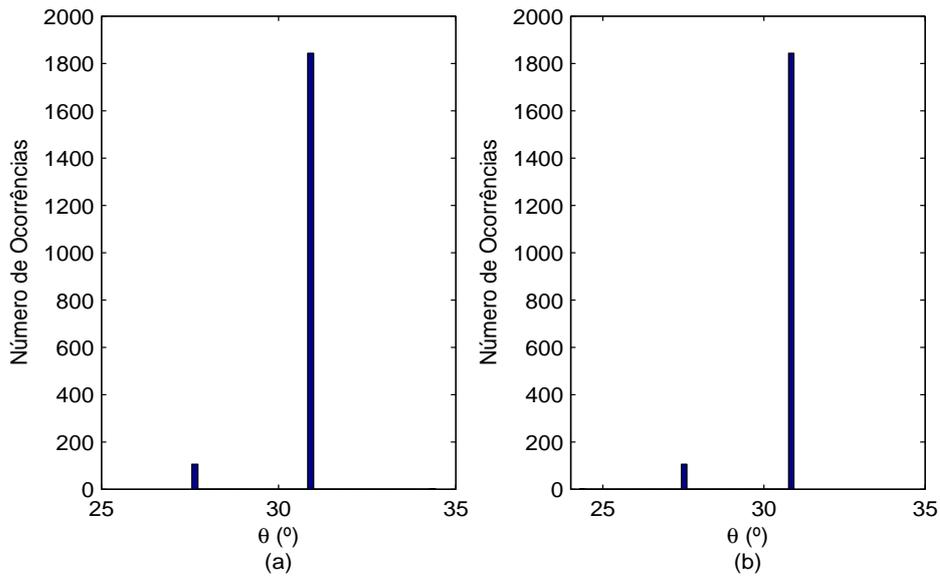


Figura 15 – Histogramas de teste com sinal branco sem ruído, ângulo de chegada de  $30^\circ$  e distância entre os microfones de 15cm. Janela utilizada de  $2^8$  amostras. (a) Correlação Cruzada. (b) GCC-PHAT

Na figura 14, percebe-se agora a influência do ruído na estimação. O uso da janela de  $2^{12}$  amostras, que na figura 13 garantiu quase a totalidade de acertos, acabou afetada quando acrescentado o ruído.

Por fim, a figura 15 apresenta os resultados de um teste usando um sinal do tipo branco, sem ruído e com janela de  $2^8$  amostras. O que se pode perceber é que, para ruído branco, as técnicas apresentam uma precisão e acurácia semelhantes a quando se utiliza uma janela maior com sinal de voz. Isso se deve às características discutidas na subseção 2.3.3.

Comparando as 4 imagens observadas aqui, é possível perceber que, em ambientes idealizados o GCC-PHAT não aparenta ter ganho algum em relação à técnica padrão sob a ótica de presença ou não de ruído, tamanho da janela e tipo de sinal.

## 2.6 Testes Práticos

Com a nova técnica definida, iniciou-se uma bateria de testes práticos para poder, enfim, comparar a correlação cruzada padrão com o GCC-Phat. Os testes pretenderam analisar, a partir das limitações expostas na seção 2.3, quão bem cada técnica se sai em relação a variação de diferentes parâmetros. Os parâmetros analisados são o ângulo de chegada e distância entre os microfones. Além disso, analisou-se também, sendo este impossível de se testar em ambiente computacional, a influência da distância ao arranjo.

Os testes foram feitos utilizando um arranjo de dois microfones a condensador,



Figura 16 – Divisor de Áudio



Figura 17 – Arranjo de Microfones e Aplicativo de Celular

conectados através de um divisor de áudio, como o da figura 16, no qual em uma das entradas o canal esquerdo foi isolado, e na outra entrada, o canal direito. Utilizou-se um aplicativo de celular para medir o ângulo desejado, como exibido na figura 17, e uma trena esticada para medir a distância entre os microfones e a distância ao arranjo.

A sala utilizada para as medições tem aproximadamente  $8,4m^2$ . Os microfones estavam sobre uma mesa, dispostos na horizontal, e a caixa de som usada para emitir o sinal foi disposta na mesma elevação dos microfones. O arranjo estava à uma distância de aproximadamente  $0,5m$  da parede de trás,  $1,8m$  da parede lateral mais distante e  $1,0m$  da parede lateral mais próxima. Um esboço da configuração do local está apresentado na figura 18.

Por fim, todos os testes foram realizados com janelas de  $2^{14}$  amostras. A SNR média das janelas utilizadas foi calculada após a filtragem das frequências indesejadas, o que resultou em uma maior relação sinal-ruído.

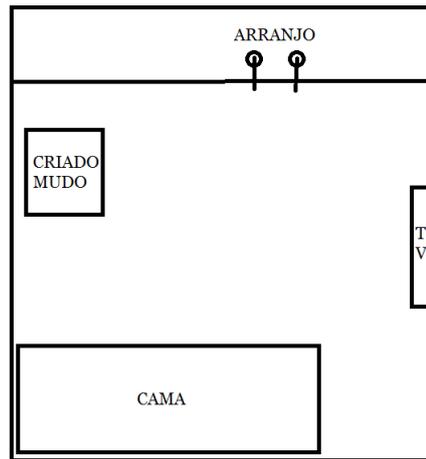


Figura 18 – Esboço da sala utilizada nos testes

### 2.6.1 Ângulo de Chegada

Analisando as figuras 19, 20 e 21 percebe-se que não há relação aparente entre o ângulo de chegada e a qualidade dos resultados obtidos. Observa-se também a questão de como a reverberação atinge os microfones em relação ao ângulo de chegada. Essa questão é dependente da configuração da sala utilizada. Nos resultados da figura 21, por exemplo, percebe-se nitidamente a reverberação atingindo o arranjo vinda da parede anterior aos microfones, distribuída através dos ângulos negativos.

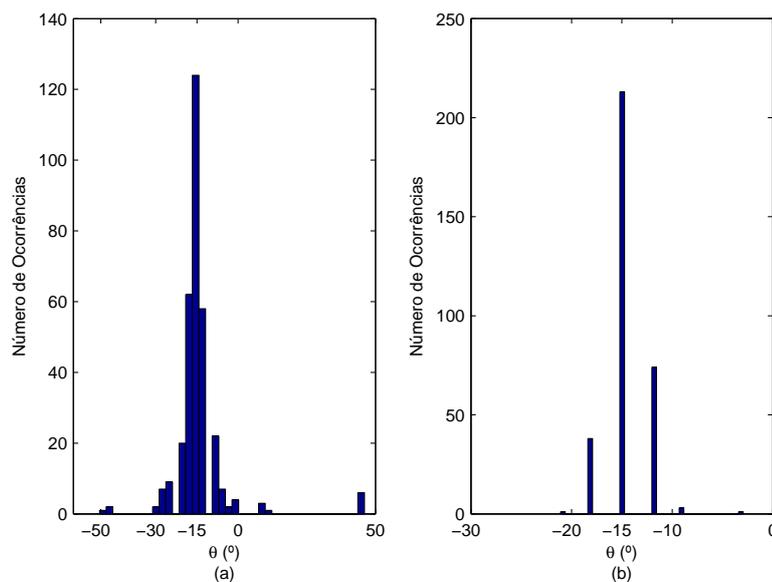


Figura 19 – Histogramas de teste com sinal de voz, ângulo de chegada de  $-15^\circ$ , distância ao arranjo de 1,5m e distância entre os microfones de 15cm. 330 janelas utilizadas para estimação. SNR média das janelas de 70,12dB. (a) Correlação Cruzada. (b) GCC-PHAT

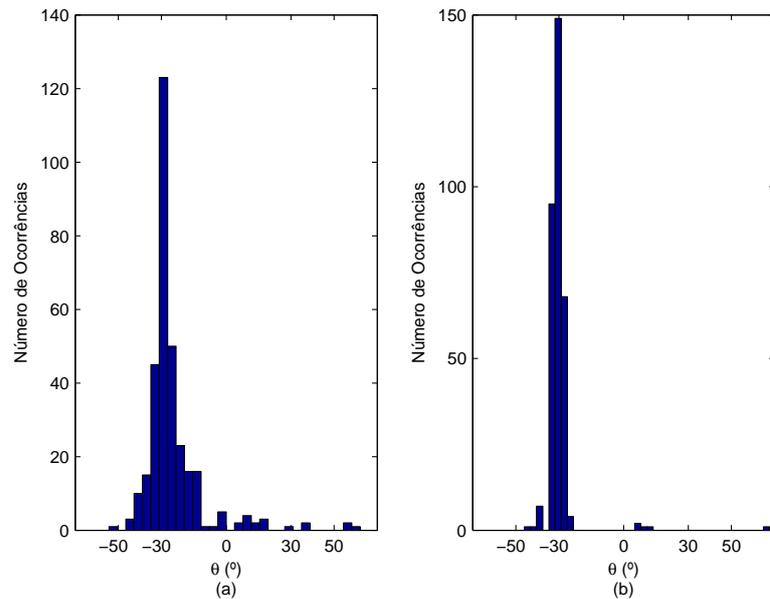


Figura 20 – Histogramas de teste com sinal de voz, ângulo de chegada de  $-30^\circ$ , distância ao arranjo de 1,5m e distância entre os microfones de 15cm. 330 janelas utilizadas para estimação. SNR média das janelas de 70,27dB. (a) Correlação Cruzada. (b) GCC-PHAT

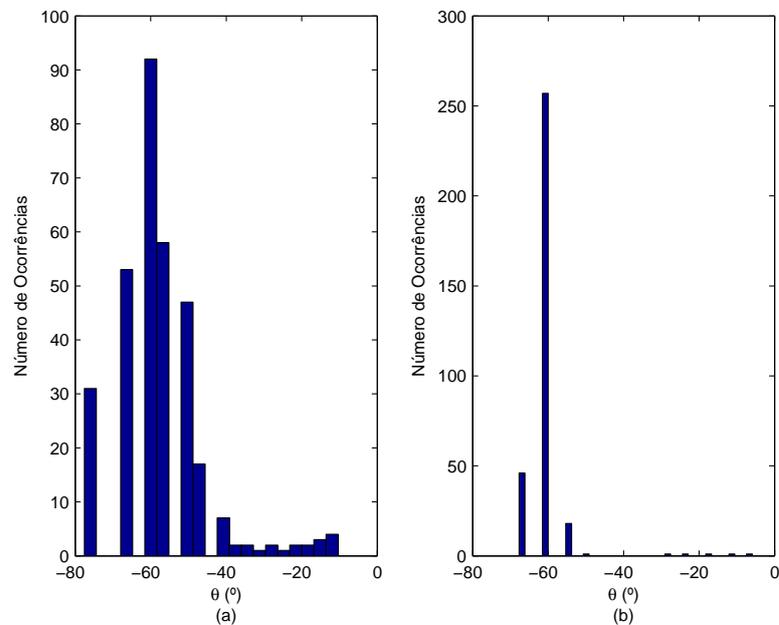


Figura 21 – Histogramas de teste com sinal de voz, ângulo de chegada de  $-60^\circ$ , distância ao arranjo de 1,5m e distância entre os microfones de 15cm. 327 janelas utilizadas para estimação. SNR média das janelas de 69,30dB. (a) Correlação Cruzada. (b) GCC-PHAT

## 2.6.2 Distância Entre os Microfones

Analisando as figuras 20, 23 e 23 observa-se uma maior dispersão dos resultados conforme a distância aumentava. Isso é provavelmente causado devido à alteração no padrão de reverberação captado pelos microfones, quanto estes sofrem um deslocamento na sua posição. Porém, essa degradação nos resultados se apresenta em menor número nas as estimações feitas com o GCC-PHAT, que concentra muito mais resultados em torno do ângulo correta.

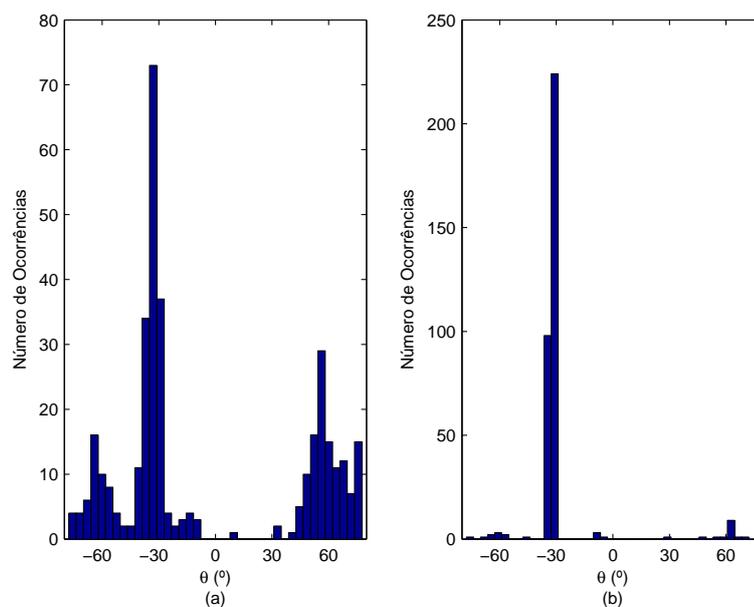


Figura 22 – Histogramas de teste com sinal de voz, ângulo de chegada de  $-30^\circ$ , distância ao arranjo de 1,5m e distância entre os microfones de 30cm. 351 janelas utilizadas para estimação. SNR média das janelas de 68,54dB. (a) Correlação Cruzada. (b) GCC-PHAT

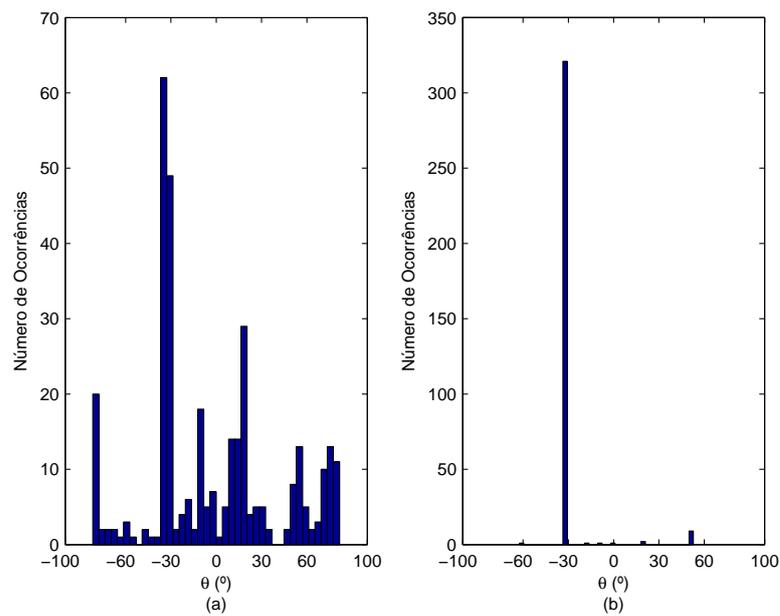


Figura 23 – Histogramas de teste com sinal de voz, ângulo de chegada de  $-30^\circ$ , distância ao arranjo de 1,5m e distância entre os microfones de 60cm. 336 janelas utilizadas para estimação. SNR média das janelas de 68,06dB. (a) Correlação Cruzada. (b) GCC-PHAT

### 2.6.3 Distância ao Arranjo

A distância ao arranjo parece colaborar mais para a perda de confiabilidade da técnica. Observa-se que, enquanto os resultados da figura 24 exibem medições bastante precisas em ambas as técnicas, a figura 25 já exhibe resultados espúrios até mesmo no GCC-PHAT. Sabendo que nas três medições variou-se apenas a distância ao arranjo, estima-se que a SNR decrescente entre cada um dos testes seja a causa da deterioração dos resultados.

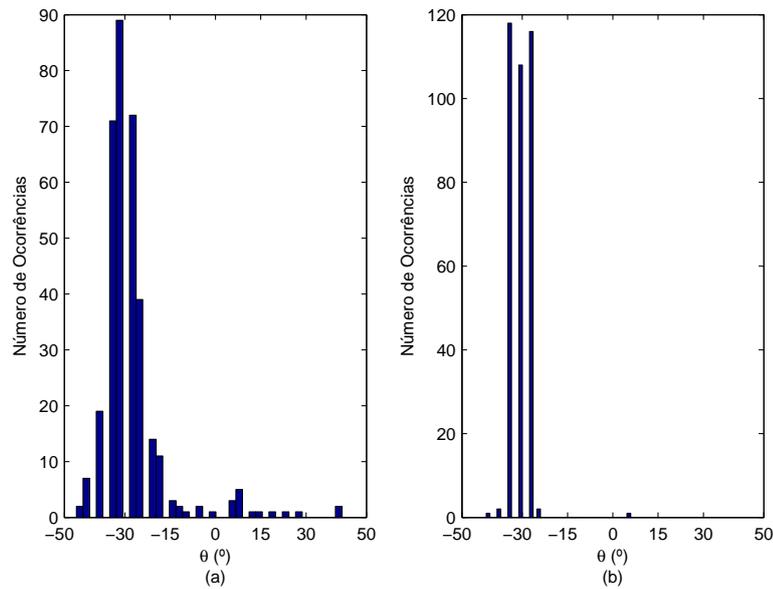


Figura 24 – Histogramas de teste com sinal de voz, ângulo de chegada de  $-30^\circ$ , distância ao arranjo de 1.0m e distância entre os microfones de 15cm. 348 janelas utilizadas para estimação. SNR média das janelas de 71,97dB. (a) Correlação Cruzada. (b) GCC-PHAT

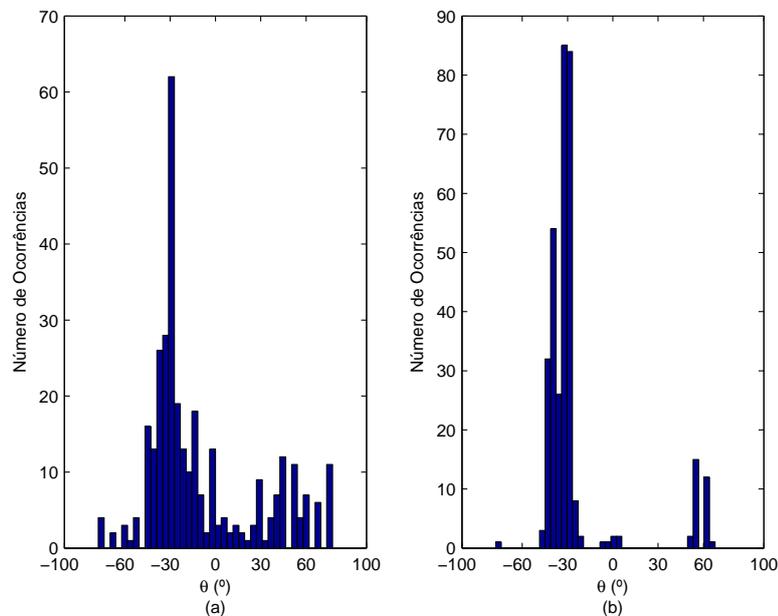


Figura 25 – Histogramas de teste com sinal de voz, ângulo de chegada de  $-30^\circ$ , distância ao arranjo de 2.0m e distância entre os microfones de 15cm. 331 janelas utilizadas para estimação. SNR média das janelas de 69,61dB. (a) Correlação Cruzada. (b) GCC-PHAT

### 2.6.4 Considerações quanto a sinal branco

Foi feito também uma comparação entre as técnicas quando expostas a um sinal de entrada branco no ambiente real. O resultado se mostrou diferente do que era esperado após as simulações, já que nestas, ambas as técnicas tiveram eficácia e precisão semelhantes tanto quando expostas à sinais coloridos (voz) quanto à sinais brancos. O que se vê na figura 26, quando comparada à figura 20 é que as técnicas tiveram resultados muito melhores com sinais brancos.

Sabe-se que diferentes frequências podem ser mais ou menos reverberantes em certas configurações de sala. A ideia que se têm é que, como sinais de voz possuem distribuição de frequência irregular, é possível que exista uma concentração maior nas frequências mais reverberantes, sendo assim, os resultados desses testes são mais influenciados pela reverberação do que os com sinais brancos.

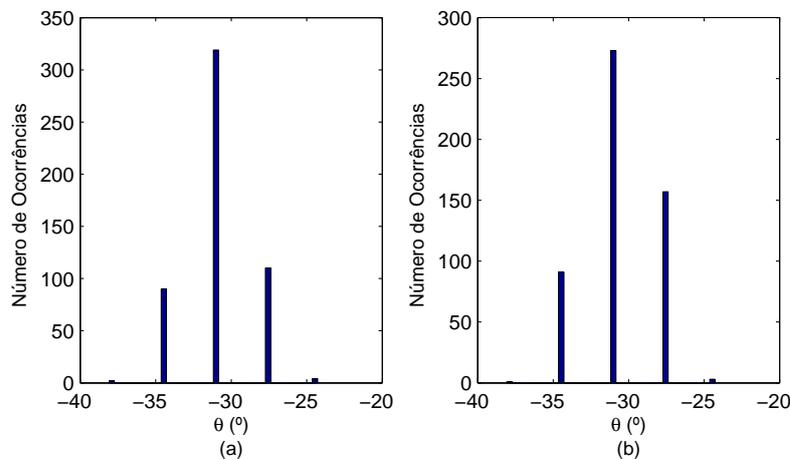


Figura 26 – Histogramas de teste com sinal branco, ângulo de chegada de  $-30^\circ$ , distância ao arranjo de 1,5m e distância entre os microfones de 15cm. 525 janelas utilizadas para estimação. SNR média das janelas de 71,18dB. (a) Correlação Cruzada. (b) GCC-PHAT

## 2.7 Conclusões

Iniciou-se este trabalho com o estudo da TDOA através da correlação cruzada. Para que a técnica pudesse ser implementada, fez-se uma análise quanto a relação entre o ângulo de chegada, a diferença de tempo de chegada e as outras variáveis de ~~meio~~ a estipular o modelo teórico utilizado para os cálculos da estimação. Posto isso, implementou-se a técnica padrão da correlação cruzada, e, através de simulações teóricas, foram feitas as análises das limitações da técnica quando exposta a diferentes situações.

Com a técnica implementada, e sua funcionalidade testada em ambiente computacional, um teste prático foi feito antes que se partisse para a segunda fase do projeto. Nessa segunda fase, estudou-se a transformada de Fourier de sinais, a dualidade tempo-frequência da correlação e a utilização de funções peso para atingir diferentes objetivos na técnica de correlação. Criada para combater os efeitos da reverberação, escolheu-se a técnica do GCC-PHAT como objetivo de implementação da segunda fase.

Implementada a técnica, novos testes teóricos foram feitos, de modo a estimar as capacidades e limitações do GCC-PHAT quando comparado à técnica padrão de correlação cruzada. Por fim, foi feita uma sequência de testes práticos para comparar as técnicas no ambiente real e comparar também com os resultados teóricos obtidos nas simulações.

Através da análise teórica e estudos práticos, percebeu-se que, em ambientes tendendo ao ideal, a técnica da correlação cruzada se mostra bastante confiável, até mesmo na presença de ruído e sinais coloridos. Porém, quanto transposta a ambientes reais, a precisão e acurácia da técnica tende a diminuir. Efeitos como a reverberação causam acentuado impacto negativo nas estimativas da direção de chegada. É possível observar desde uma distribuição maior em torno do ponto desejado até mesmo resultados absolutamente espúrios ao verificar os resultados dos testes.

Criada justamente para se contrapor justamente à esses efeitos da reverberação, a técnica do GCC-PHAT exibiu, em praticamente todos os casos testados, uma melhora pronunciada na capacidade da técnica em entregar resultados condizentes com o esperado. Essa melhora, em certos casos, chegou a até 300%, como no caso da figura 23. Sendo assim, é possível justificar a popularidade da técnica GCC-PHAT para a estimação da direção de chegada através do TDOA.

# Referências

- [1] J. Benesty, J. Chen, and Y. Huang. *Microphone Array Signal Processing*. Springer Topics in Signal Processing. Springer, 2008.
- [2] M. A. Khalid, M. I. K. Babar, M. H. Zafar, and M. F. Zuhairi. Gunshot detection and localization using sensor networks. In *Smart Instrumentation, Measurement and Applications (ICSIMA), 2013 IEEE International Conference on*, pages 1–6, Nov 2013.
- [3] B. C. Kirkwood. Acoustic source localization using time-delay estimation. Master’s thesis, Technical University of Denmark, 2003.
- [4] C. H. Knap and G. C. Carter. The generalized correlation method for estimation of time delay. In *IEEE Transactions on Acoustics, Speech, and Signal Processing*, pages 320–327, Aug 1976.
- [5] B. P. Lathi. *Modern Digital & Analog Communications Systems 2e*. HRW series in electrical and computer engineering. Oxford University Press, Incorporated, 1995.
- [6] A.V. Oppenheim, A.S. Willsky, and I.T. Young. *Signals and systems*. Prentice-Hall signal processing series. Prentice-Hall, 1983.
- [7] R. J. Urick. *Principles of Underwater Sound*. Península Publications, 1983.
- [8] C. Zhang, D. Florência, and Z. Zhang. Why does phat works well in low noise, reverberant environments. In *International Conference on Acoustics, Speech and Signal Processing*, pages 2565–2568, May 2008.