



LOCALIZAÇÃO DE FONTES SONORAS EM RECINTOS FECHADOS APLICADA À TELEVIGILÂNCIA MÉDICA

Rogério Réus Santin

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Elétrica – PROEE, da Universidade Federal de Sergipe, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Elétrica.

Orientador: Jugurta Rosa Montalvão Filho

São Cristóvão-SE, Brasil

Dezembro de 2012

LOCALIZAÇÃO DE FONTES SONORAS EM RECINTOS FECHADOS APLICADA
À TELEVIGILÂNCIA MÉDICA

Rogério Réus Santin

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA – PROEE DA UNIVERSIDADE FEDERAL DE SERGIPE COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA ELÉTRICA.

Examinada por:

Prof. Charles Casimiro Cavalcante, D.Sc.

Prof. Eduardo Oliveira Freire, D.Sc.

Prof. Jugurta Rosa Montalvão Filho, D.Sc.

SÃO CRISTÓVÃO-SE, BRASIL
DEZEMBRO DE 2012

Santin, Rogério Réus

Localização de Fontes Sonoras em Recintos Fechados
Aplicada à Televigilância Médica/Rogério Réus Santin. – São
Cristóvão: UFS/PROEE, 2012.

XIV, 52 p.: il.; 29, 7cm.

Orientador: Jugurta Rosa Montalvão Filho

Dissertação (mestrado) – UFS/PROEE, 2012.

Referências Bibliográficas: p. 50 – 52.

1. Direção de chegada. 2. Correlação Cruzada
Generalizada. 3. Transformação de Fase. 4. DUET. I.
Montalvão Filho, Jugurta Rosa. II. Universidade Federal de
Sergipe, Programa de Pós-Graduação em Engenharia Elétrica.
III. Título.

*À minha família, pedra angular de
toda construção.*

Agradecimentos

Para que esta dissertação chegasse a termo, enquanto resultado de um longo caminho percorrido, várias pessoas doaram seu tempo, conhecimento, paciência e companheirismo. Sem o apoio dos colegas de turma, dos professores da UFS e da Coppe e, principalmente, do meu orientador, pouco teria sido construído. A todos a minha mais profunda gratidão.

Resumo da Dissertação apresentada ao PROEE/UFS como parte dos requisitos necessários para a obtenção do grau de Mestre (Me.)

LOCALIZAÇÃO DE FONTES SONORAS EM RECINTOS FECHADOS APLICADA À TELEVIGILÂNCIA MÉDICA

Rogério Réus Santin

Dezembro/2012

Orientador: Jugurta Rosa Montalvão Filho

Programa: Engenharia Elétrica

Devido ao fenômeno global do envelhecimento da população, o tratamento domiciliar - *home care* - torna-se mais comum a cada dia, especialmente para os casos de tratamentos longos, por tratar-se de um método terapêutico mais humano e econômico. O problema decorrente desta prática é como assegurar uma rápida intervenção caso necessário. Um sistema de vigilância médica pode ser a solução para manter os cuidadores informados e aumentar a segurança dos pacientes. Neste trabalho é investigada a possibilidade de se detectar, selecionar e determinar a direção de chegada de sons recebidos por sensores instalados em um ambiente interno, com o objetivo de diagnosticar uma situação de emergência, tal como uma queda do paciente. Usando um arranjo de microfones, após a detecção de um evento acústico alvo, os sons são gravados para análise. Ao passarem pelo critério de seleção, os sons são processados por um algoritmo de estimação de ângulo de chegada - *Direction of Arrival* (DOA). Cinco estimadores de DOA foram implementados e comparados empiricamente, sendo três existentes, a saber, a Correlação Cruzada Generalizada - *Generalized Cross Correlation* - (GCC), a Correlação Cruzada Generalizada com Transformação de Fase - *Generalized Cross Correlation - Phase Transform* - (GCC-PHAT), a Técnica de Estimação e Separação de Sinais Degenerados - *Degenerate Unmixing and Estimation Technique* - (DUET); e dois novos algoritmos propostos neste trabalho, tendo sido denominados Detector de Borda de Ataque e Técnica do Espectrograma. Ao final, através da perspectiva da televigilância médica, avaliou-se o desempenho dos mesmos.

Abstract of Dissertation presented to PROEE/UFS as a partial fulfillment of the requirements for the degree of Master

SOUND SOURCES LOCALIZATION IN INDOOR ENVIRONMENTS APPLIED TO REMOTE MEDICAL SURVEILLANCE

Rogério Réus Santin

December/2012

Advisor: Jugurta Rosa Montalvão Filho

Department: Electrical Engineering

Due to the worldwide people ageing phenomenon, the health care at home (home care), a more humanized and cost saving method, for elderly people and long term care, is nowadays more and more usual. The problem is to assure prompt help when needed. A surveillance system can be the solution to improve caregiver awareness and patient safety. This work investigates the possibility of indoor sounds direction of arrival detection, selection and determination, attempting to find out signals of a patient fall or collapse. The sounds after an acoustic event are recorded using a microphone array. If they pass the selection criteria, they are processed by a *Direction of Arrival* (DOA) detection algorithm. Five DOA algorithms were implemented and evaluated empirically. Three of them are already known, nominally, *Generalized Cross Correlation* (GCC), *Generalized Cross Correlation - Phase Transform* (GCC-PHAT), *Degenerate Unmixing and Estimation Technique* (DUET), and two are proposed at this work, *Attack Border Detector* and *Spectrogram Technique*. Performances were evaluated by taking the home care perspective into account.

Sumário

Lista de Figuras	x
Lista de Tabelas	xii
Lista de Abreviaturas	xiii
1 Introdução	1
2 Revisão Bibliográfica	3
2.1 Ondas sonoras	3
2.2 Propriedades das ondas sonoras	5
2.3 Um pouco mais sobre a reverberação	6
2.4 Métodos de estimação da direção de chegada	7
2.4.1 Métodos de Apontamento Direcionado	10
2.4.2 Métodos de Decomposição por AutoValores	10
2.4.3 Métodos de Estimativa de Atraso	10
2.5 Trabalhos correlatos	11
2.6 Uma ferramenta: o espectrograma	12
2.7 Síntese	13
3 Sons Utilizados nos Testes e Avaliações	14
3.1 Síntese	17
4 Algoritmos Estudados	18
4.1 Correlação Cruzada Generalizada	19
4.2 Correlação Cruzada Generalizada com Transformação de Fase	20
4.3 Técnica de Estimação e Separação de Sinais Degenerados	23
4.4 Método da Borda de Ataque	25
4.5 Técnica do Espectrograma	28
4.6 Síntese	30

5	Resultados e Discussões	31
5.1	Escolha de parâmetros	31
5.2	Comparativo do GCC, GCC-PHAT e DUET perante reverberações	39
5.3	O Algoritmo Detector de Borda de Ataque	43
5.4	A Técnica do Espectrograma	45
5.5	Testes com outros sinais	46
5.6	Síntese	48
6	Conclusões	49
	Referências Bibliográficas	50

Lista de Figuras

2.1	Algumas características das ondas sonoras.	3
2.2	Reverberações no interior de um ambiente.	6
2.3	Diferentes tempos de chegada dos sinais sonoros aos sensores em função do ângulo de incidência.	8
2.4	Ambiguidade entre dois sinais com ângulos de incidência espelhados horizontalmente.	9
2.5	Espectrograma da palavra <i>Socorro</i>	13
3.1	Caixas acústicas de referência.	15
3.2	Microfones de referência captando os sons ambientes.	16
3.3	Equipamentos utilizados nos experimentos.	17
4.1	Agrupamento dos pontos tempo-frequência em duas regiões.	24
4.2	Ilustração dos caminhos do sinal direto e da 1ª reflexão no piso.	25
4.3	Ondas sonoras características captadas pelos sensores.	26
4.4	Quadros pequenos de sinais defasados e o efeito do seu janelamento, representado por envoltórias.	27
4.5	Espectrogramas de sinais com reverberações e defasados temporalmente.	28
4.6	Fluxograma do algoritmo da Técnica do Espectrograma.	30
5.1	Deteção de atraso com a utilização de diferentes janelamentos.	32
5.2	Assertividade do algoritmo GCC para diversos tamanhos de quadros.	33
5.3	Algoritmo GCC para atrasos de 0, 11 e 22 amostras.	34
5.4	Algoritmo GCC-PHAT para atrasos de 0, 11 e 22 amostras.	35
5.5	Algoritmo GCC para diversas relações sinal-ruído.	36
5.6	Algoritmo GCC-PHAT para diversas relações sinal-ruído.	37
5.7	Algoritmo GCC para quadros de 512 amostras e diversas relações sinal-ruído.	38
5.8	Algoritmo GCC-PHAT para quadros de 512 amostras e diversas relações sinal-ruído.	39
5.9	Histograma do atraso detectado para o 1º sinal.	40
5.10	Histograma do atraso detectado para o 2º sinal.	40

5.11	Histograma do atraso detectado para o 1º sinal com emulação.	41
5.12	Histograma do atraso detectado para o 2º sinal com emulação.	41
5.13	Histograma do atraso detectado para o 1º sinal gravado.	42
5.14	Histograma do atraso detectado para o 2º sinal gravado.	43

Lista de Tabelas

5.1	Número de acertos para 40 simulações com diversos tamanhos de quadros.	32
5.2	Erro relativo em número de amostras com o Detector de Borda para o 1º sinal gravado.	44
5.3	Erro relativo em número de amostras com o Detector de Borda para o 2º sinal gravado.	44
5.4	Erro relativo em número de amostras com o Detector de Borda para o 3º sinal gravado.	44
5.5	Erro relativo e desvio padrão com a Técnica do Espectrograma para o 1º sinal gravado.	45
5.6	Erro relativo e desvio padrão com a Técnica do Espectrograma para o 2º sinal gravado.	46
5.7	Erro relativo e desvio padrão com a Técnica do Espectrograma para o 3º sinal gravado.	46
5.8	Erro médio e desvio padrão para o 4º sinal gravado.	46
5.9	Erro médio e desvio padrão para o 5º sinal gravado.	47
5.10	Erro médio e desvio padrão para o 6º sinal gravado.	47

Lista de Abreviaturas

AED	Decomposição Adaptativa por Autovalores - Adaptive Eigen-value Decomposition -, p. 12
CC	Correlação Cruzada - Cross Correlation -, p. 11
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico -, p. 2
DFT	Transformada Discreta de Fourier - Discrete Fourier Transform -, p. 18
DOA	Direção de Chegada - Direction of Arrival -, p. vi
DSB	Apontamento por Atraso e Soma - Delay and Sum Beamformer -, p. 10
DUET	Técnica de Estimação e Separação de Sinais Degenerados - Degenerate Unmixing and Estimation Technique -, p. vi
EM	Esperança - Maximização - Expectation - Maximization -, p. 23
ESPRIT	Técnicas de Estimação de Parâmetros de Sinais via Invariância Rotacional - Estimation of Signal Parameters via Rotational Invariance Techniques -, p. 12
EVD	Decomposição por Autovalores - Eigen-Value Decomposition -, p. 10
GCC	Correlação Cruzada Generalizada - Generalized Cross Correlation -, p. vi
GCC-PHAT	Correlação Cruzada Generalizada com Transformação de Fase - Generalized Cross Correlation - PHase Transform -, p. vi
GXPSD	Densidade Espectral de Potência Cruzada Generalizada - Generalized Cross Power Spectral Density -, p. 20

IBGE	Instituto Brasileiro de Geografia e Estatística, p. 1
LSE	Erro Quadrático Mínimo - Least Squared Error -, p. 11
MCT	Ministério da Ciência e Tecnologia -, p. 2
ML	Máxima Verossimilhança - Maximum Likelihood -, p. 11
MSE	Erro Quadrático Médio - Mean Squared Error -, p. 11
MUSIC	Classificação de Múltiplos Sinais - Multiple Signal Classification -, p. 10
SBF	Apontamento Direcionado - Steered Beamformer -, p. 10
SNR	Relação Sinal-Ruído - Signal to Noise Ratio -, p. 11
SRP	Potência da Resposta Direcionada - Steered Response Power -, p. 12
SUS	Sistema Único de Saúde, p. 1
TDE	Estimação de Atraso - Time Delay Estimation -, p. 10
TIDES	Seleção de Atraso - Time Delay Selection -, p. 12
ULA	Arranjo Linear Uniforme - Uniform Linear Array -, p. 7
XPSD	Densidade Espectral de Potência Cruzada - Cross Power Spectral Density -, p. 18

Capítulo 1

Introdução

O rápido envelhecimento populacional que ocorre há décadas nos países desenvolvidos tem, progressivamente, se tornado um fenômeno global neste século, afetando também os países em desenvolvimento. Os idosos hoje constituem 19% da população nos países desenvolvidos, e serão 33% (37% na Europa) em 2050 [Doukas e Maglogiannis 2011]. No Brasil, a população de idosos representa um contingente de quase 15 milhões de pessoas com mais de 59 anos de idade (8,6% da população) e, segundo o Instituto Brasileiro de Geografia e Estatística (IBGE), nos próximos 20 anos a população idosa do país poderá ultrapassar os 30 milhões de pessoas e representar quase 13% da população.

Neste contexto de grande crescimento do número de idosos, faz-se necessária a implantação de políticas públicas de saúde que provejam serviços de qualidade para todos, com racionalidade econômica. No Brasil, o governo federal criou um programa chamado *Melhor em Casa*, o qual prevê a assistência médica domiciliar para pacientes do Sistema Único de Saúde (SUS). A iniciativa, que já é encontrada em alguns pontos do País, permite que pessoas idosas, com doenças crônicas ou em recuperação de cirurgias sejam atendidas em casa, reduzindo assim a necessidade e o tempo de internação. A meta é implantar 1000 equipes até 2014, além de 400 equipes multidisciplinares de apoio [ASCOM/MS 2011].

Para atingir uma grande parcela populacional, proporcionando a atenção e a segurança necessárias, a alternativa do cuidado domiciliar - *home care* - com monitoramento remoto demonstra ser uma solução muito mais acolhedora e econômica que a internação hospitalar [Vacher et al. 2006], sendo que os progressos em tecnologias de vigilância e assistência médica remota prometem ser uma boa aposta para ajudar a atingir as metas de universalização do atendimento.

Visando dar uma maior segurança para as pessoas envolvidas neste processo, torna-se fundamental o monitoramento diuturno do ambiente em que vive o paciente para a detecção de emergências, especialmente no caso de pessoas sofrendo de doenças crônicas ou de idosos vivendo sozinhos. O principal objetivo deve ser a detecção de situações de emergência tais como quedas, as quais podem provocar lesões severas e a necessidade de

uma rápida intervenção. Os requisitos mandatórios para um sistema de segurança pessoal consistem na utilização de tecnologias de monitoramento pouco invasivas, que preservem a privacidade de seus usuários e que permitam a mobilidade. Sistemas que tenham estas características têm uma maior aceitação por parte das pessoas a serem monitoradas [Stout 2010]. Ao mesmo tempo, estes sistemas devem possuir uma taxa elevada de acertos na interpretação dos eventos, para não caírem em descrédito junto aos usuários.

Com o objetivo de atender a estas demandas, é avaliada neste estudo uma solução para os problemas de invasibilidade, mobilidade, privacidade e assertividade na detecção de eventos anômalos, através da estimativa da direção de chegada de sons em ambientes internos. Esta etapa, enquanto parte de um conjunto de esforços paralelos no âmbito do projeto AtenSom [Ministério da Ciência e Tecnologia (MCT) / Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) N^o14/2010], deverá servir de base para um sistema de vigilância para detecção de emergências, tais como quedas de idosos ou desmaios de pacientes.

A escolha de um sistema de monitoramento acústico, ao mesmo tempo em que evita a exposição das pessoas, como no caso do videomonitoramento, implica ainda, via de regra, baixa carga computacional e complexidade dos algoritmos de detecção. Por outro lado, em ambientes fechados e com dimensões reduzidas, como tipicamente vem a ser o local onde o paciente permanece, têm-se efeitos indesejados que dificultam a localização da fonte emissora. Estes efeitos originam-se do fato de que o comprimento das ondas sonoras tem a mesma ordem de grandeza dos objetos e das estruturas do ambiente. Em um cômodo fechado, o som chega aos microfones não somente pelo caminho direto, mas também através de inúmeras reflexões nas paredes, piso e teto, sendo este efeito especialmente danoso no caso da fonte sonora estar próxima a uma superfície refletora, como quando uma pessoa caída no chão passa a emitir sons. Este fenômeno físico é conhecido como reverberação, e sua ocorrência pode degradar severamente o desempenho dos algoritmos, induzindo a falsos resultados. Assim, durante as pesquisas fez-se necessário o estudo e a busca de soluções específicas, tendo sido propostos neste trabalho dois novos algoritmos para fazer frente aos desafios desse cenário.

Para melhor compreensão das dificuldades inerentes à localização de fontes sonoras em ambientes fechados, especialmente os de pequenas dimensões, os quais constituem-se nos ambientes a serem monitorados para televigilância médica, encontra-se no capítulo 2 uma descrição de alguns fenômenos sonoros, com ênfase na reverberação. No mesmo capítulo tem-se uma breve explanação de algumas publicações com linhas de pesquisa semelhantes, e suas conclusões pertinentes. Depois, no capítulo 3, discorre-se sobre os sons reais digitalizados utilizados nos experimentos. No capítulo 4 são demonstradas as equações e explicado o princípio de funcionamento de cada um dos algoritmos utilizados. Já no capítulo 5 são apresentados os resultados obtidos nas diversas situações experimentais ora investigadas. Por último têm-se as principais conclusões advindas deste trabalho.

Capítulo 2

Revisão Bibliográfica

2.1 Ondas sonoras

A onda sonora é um tipo de onda mecânica produzida pela vibração de um objeto em um meio. Ela é a propagação de uma frente de compressão mecânica ou onda mecânica; é uma onda longitudinal, que se move de forma circuncêntrica, apenas em meios materiais - que têm massa e elasticidade, como os sólidos, líquidos ou gasosos - tendo em vista que a onda se propaga através de compressões e rarefações do meio. Tais compressões e rarefações possuem um comportamento cíclico e senoidal, conforme pode-se observar na Figura 2.1, a qual contém uma ilustração do deslocamento de uma onda sonora, com um gráfico representando as pressões envolvidas [BARBA 1998].

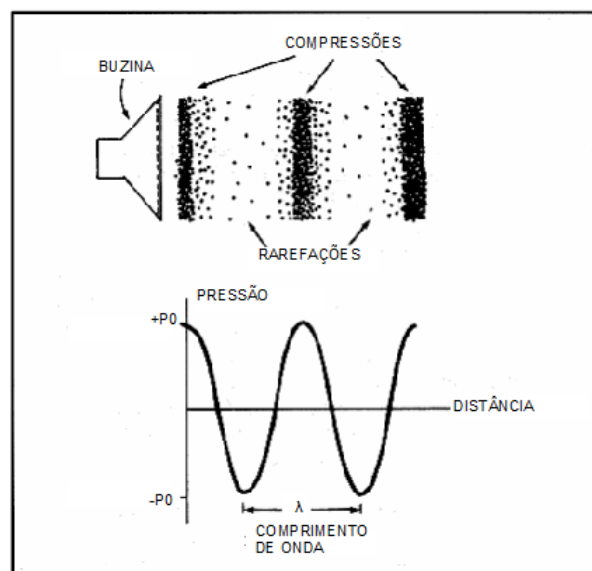


Figura 2.1: Algumas características das ondas sonoras.

Suas principais características são:

Amplitude - É a pressão instantânea ou deslocamento de partículas em um meio. Sua unidade é o pascal [Pa]. A amplitude não tem relação com a velocidade da onda ou sua frequência;

Fase - Tendo em vista que os movimentos de compressão e rarefação são cíclicos, medindo-se a sequência de valores de pressão aos quais um determinado local está sendo submetido, pode-se determinar a fase da onda em um determinado instante. A fase é uma medida arbitrária e depende do referencial escolhido;

Período - Consiste no tempo que decorre entre um determinado local ser submetido à pressão correspondente a um ponto qualquer em um ciclo e o mesmo local voltar a ser submetido a outro ponto de mesma fase no próximo ciclo. Para sons audíveis, sua ordem de grandeza é de milissegundos;

Frequência - É o número de ciclos que ocorrem por segundo, sendo o inverso do período, e sua unidade é o hertz [Hz];

Comprimento de Onda - É o período espacial da onda, isto é, a distância entre dois pontos consecutivos de mesma fase;

Velocidade - É o deslocamento de uma frente de onda por uma determinada distância, dividido pelo tempo decorrente para tal. Sua unidade é metros por segundo [m/s]. Por depender exclusivamente do meio, ondas de frequências diferentes propagam-se com a mesma velocidade no mesmo meio.

Os sons audíveis pelo ouvido humano são, na sua maior parte, combinações de sinais, os quais devem possuir frequências entre 20 Hz e 20.000 Hz. Acima e abaixo desta faixa estão ultrassom e infrassom, respectivamente. A unidade usual da intensidade, ou energia sonora, é o decibel [dB], sendo 0 dB o limiar da audição humana, correspondendo a uma intensidade de $1 \cdot 10^{-12} W/m^2$ [Henderson 2012].

A velocidade das ondas sonoras depende das propriedades elásticas e inerciais do meio através do qual a onda está se propagando, sendo inversamente proporcional a elas. De uma forma geral, a velocidade do som é maior nos sólidos, menor nos líquidos e menor ainda nos gases, devido à elasticidade, a qual tem maior influência que a inércia. No ar, que é o meio pelo qual recebemos normalmente os sons, as características que sobressaem-se são a temperatura, de modo diretamente proporcional e, em menor grau, a umidade, de modo inversamente proporcional. Neste trabalho, por simplificação, será considerada uma velocidade fixa de propagação do som no ar de 340 m/s.

Os sons são usados de diversas maneiras, muito especialmente para comunicação através da fala ou, por exemplo, música. A percepção do som também pode ser usada para adquirir informações sobre o ambiente, como características espaciais (forma, topografia) e presença de outros animais ou objetos. Vários animais, incluindo os seres hu-

manos, captam os sons através de seus dois ouvidos (a chamada audição estereofônica), o que permite a eles estimar a distância e a posição da fonte sonora. Essa capacidade possibilita que muitas espécies utilizem o som para detecção do perigo, orientação, caça e comunicação. Morcegos, baleias e golfinhos, por exemplo, usam a ecolocalização para voar ou nadar por entre obstáculos e caçar suas presas, sendo que navios e submarinos utilizam um aparelho baseado neste princípio, chamado sonar.

2.2 Propriedades das ondas sonoras

As ondas sonoras apresentam também as seguintes propriedades:

Interferência - Consiste no encontro de dois ou mais sons de fontes diferentes, que interagem construtiva (quando coincidem duas compressões ou duas rarefações) ou destrutivamente (quando coincidem uma compressão e uma rarefação) em um determinado ponto, a cada instante;

Reflexão - Quando ondas sonoras provenientes de uma fonte pontual encontram um obstáculo plano e rígido, produz-se reflexão das ondas sobre o obstáculo, com o surgimento de uma série de ondas refletidas que se propagam em sentido inverso ao das ondas incidentes e se comportam como se emanassem de uma fonte virtual no lado oposto do obstáculo, a uma distância simétrica em relação à fonte original;

Refração - Ocorre quando uma onda sonora passa de um meio para o outro, mudando sua velocidade de propagação e comprimento de onda, mas mantendo constante a sua frequência;

Difração - Fenômeno em que uma onda sonora pode transpor obstáculos através de desvios nas extremidades dos corpos que encontra, evidenciando que as ondas sonoras não se propagam apenas em linha reta;

Ressonância - Quando uma das frequências das ondas que chegam até um sistema coincide com a frequência natural de vibração deste sistema, ocorre um fenômeno chamado ressonância, que consiste na amplificação desta frequência.

A reflexão pode ocasionar o eco (que é a reflexão da onda sonora por obstáculo suficientemente distante) e a reverberação (superposição linear de inúmeras reflexões com o sinal original), dependendo do tempo que as ondas refletidas levam para atingir o ouvinte. Como os ambientes internos de uma casa são relativamente pequenos, as reflexões das ondas sonoras nas paredes, piso e teto originam a reverberação, que afeta diretamente os resultados obtidos, conforme demonstrado no decorrer desse trabalho, motivo pelo qual ela será mais detalhada a seguir.

2.3 Um pouco mais sobre a reverberação

Reverberação é a persistência do som em um determinado espaço após a produção de um som original. Tem-se reverberação quando um som é produzido em um espaço fechado ou semifechado, sendo que as sucessivas reflexões deste som somam-se e decaem suavemente devido à absorção pelas paredes e pelo ar. Diferentemente do eco, que é uma reflexão sonora que aparece cerca de 100 ms após o som inicial, a reverberação é uma combinação de milhares de pequenas reflexões, que se sucedem muito rapidamente para serem percebidas como sons separados (o intervalo entre elas é de $10\mu s$ à 1 ms), tendo seu volume reduzido paulatinamente, até não mais serem audíveis [Skaloumbakas 2010].

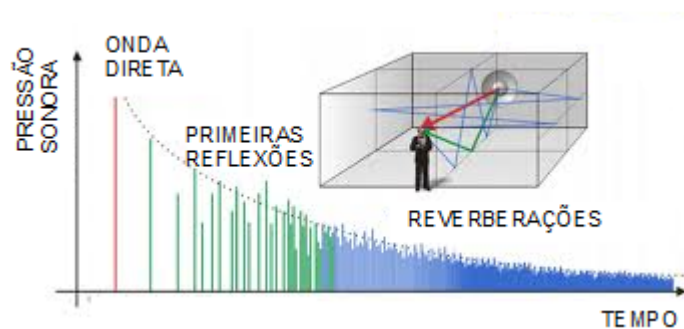


Figura 2.2: Reverberações no interior de um ambiente.

Alguns fatos relevantes sobre a reverberação:

- Ocorre frequentemente em espaços pequenos, com uma ou mais dimensões, como altura, comprimento ou profundidade, possuindo 17 metros ou menos. Como a velocidade do som é de cerca de 340 m/s, as trajetórias das ondas sonoras refletidas devem ser menores que 34 m para que o intervalo entre a captação do som original e das ondas refletidas seja menor que 100 ms; assim sendo o cérebro interpretará todas as ondas sonoras como um único som prolongado;
- A reverberação aumenta o nível de ruído e o volume aparente dos sons dentro de um ambiente;
- O período de tempo que a reverberação leva para tornar-se inaudível é proporcional às dimensões do ambiente e inversamente proporcional à absorção do som pelas paredes. Materiais duros, como o concreto e o vidro, tendem a ser ótimos refletos, enquanto que materiais mais moles, como a fibra de vidro ou tecidos, tendem a absorver muito mais o som;

- Materiais lisos tendem a refletir as ondas sonoras para uma única direção, enquanto materiais mais ásperos tendem a difundir o som, refletindo-o para várias direções, "enchendo" e tornando mais vivo o ambiente. Devido a isso, materiais rugosos são preferidos em salas de concertos e auditórios;
- Inconscientemente, o fenômeno da reverberação é utilizado para orientação em um determinado espaço, e a razão entre o som direto e o reverberado nos dá uma percepção de profundidade e distância. Em grandes espaços, por exemplo, a intensidade do som direto decai muito mais rapidamente que os sons da reverberação;
- O fenômeno da reverberação possibilita a criação de efeitos como o reforço sonoro em locais destinados a apresentações públicas, sendo que o tempo de reverberação desejado para um determinado ambiente depende do tipo de música a ser executado (ou som a ser produzido) naquele espaço. Salas utilizadas para fala (teleconferência, p. ex.), tipicamente necessitam de um tempo menor de reverberação para que a fala possa ser facilmente entendida. Se o som refletido de uma sílaba ainda é ouvido quando a sílaba seguinte é falada, pode-se tornar uma palavra ininteligível. Por outro lado, se o tempo de reverberação é muito curto, o balanço tonal e o volume sonoro podem ser prejudicados;
- O tempo de decaimento da reverberação é objeto de estudos cada vez mais aprofundados, especialmente nos casos de salas de espetáculos, onde sua duração deve ser bem definida, a fim de se obter um ótimo desempenho musical [Montalvão Filho 1995];
- Efeitos de reverberação também são utilizados frequentemente em estúdios para dar "profundidade" aos sons. A reverberação muda a percepção da estrutura harmônica de uma nota, mas não altera suas frequências (*pitch*).

2.4 Métodos de estimação da direção de chegada

O princípio fundamental da estimação da direção de chegada - *Direction of Arrival* - (DOA) usando arranjos lineares de microfones - *Uniform Linear Arrays* - (ULA) é a utilização da informação de fase presente nos sinais captados pelos sensores espacialmente separados. Devido à distância entre os microfones, os sinais chegam até eles em tempos diferentes, dependendo da sua direção de chegada, conforme esquema apresentado na Figura 2.3 [Varma 2002].

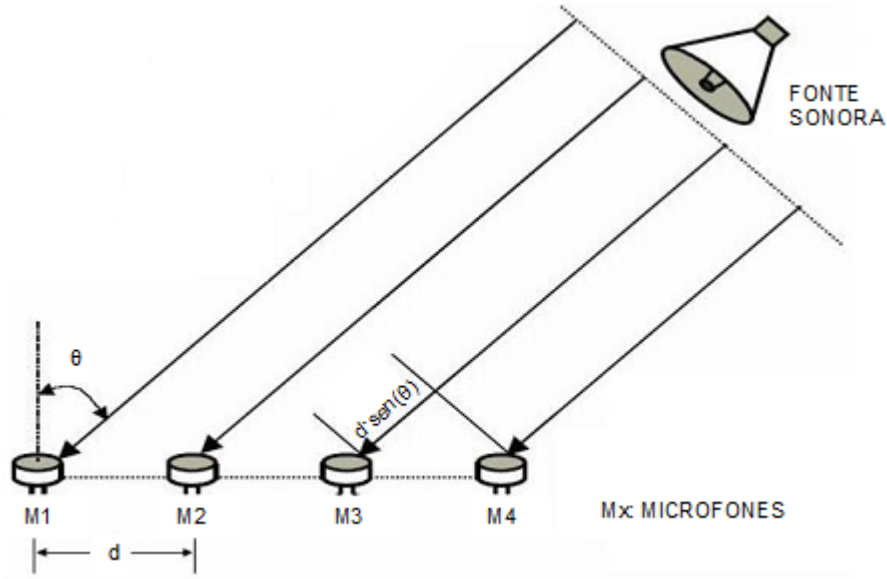


Figura 2.3: Diferentes tempos de chegada dos sinais sonoros aos sensores em função do ângulo de incidência.

Através do atraso - *time delay* - entre dois sinais captados por microfones consecutivos pode-se deduzir o ângulo de chegada do sinal, de acordo com:

$$\theta = \arcsin \frac{\tau \cdot c}{d}; \quad (2.1)$$

em que θ é o ângulo de incidência, τ o atraso relativo em segundos, c a velocidade do som no ar em metros por segundo e d a distância entre os microfones em metros. Assim, conhecendo-se a distância entre microfones, pode-se estimar o ângulo de chegada do som através do atraso relativo entre as réplicas (idealmente idênticas, a não ser pelo atraso) de sinais sonoros capturados pelos microfones.

Percebe-se pela Equação (2.1) que os valores de τ que podem ser interpretados univocamente estão entre $-d/c$ e $+d/c$, tendo-se em vista que a função seno varia entre -1 e +1, repetindo-se ciclicamente. Estes limites, impostos pela distância entre os sensores e pela velocidade do som no meio, são importantes para os algoritmos e também para as defasagens artificiais criadas entre sinais para simulações.

Deve-se observar neste tipo de arranjo que a maior frequência que pode ser utilizada sem ambiguidades é aquela dada pela relação c/d . Esta relação deriva do fato de que se deve ter, no máximo, um comprimento de onda inteiro entre os dois sensores para interpretar cada fase univocamente. Por exemplo, para $d = 0,34$ m, considerando $c = 340$ m/s, temos uma frequência máxima $f_{MAX} = 1$ kHz, a qual possui um comprimento de onda de 0,34 m em um meio com esta velocidade. Como os sinais de fala e de ruídos são composições de frequências, e não uma única frequência pura, concentrando-se sua potência sonora nas frequências mais baixas (abaixo de 1 kHz), este aspecto pode ser desconsiderado neste trabalho.

Outro ponto a se notar é que a resolução espacial, ou seja, a quantidade de ângulos passíveis de detecção, é proporcional ao atraso máximo multiplicado pela frequência de amostragem ($Res = \tau_{MAX} \cdot f_s$), sendo assim proporcional à distância entre os sensores também.

Voltando à Figura 2.3, vê-se que a sua configuração permite a distinção de ângulos de incidência entre -90° e $+90^\circ$ em relação à perpendicular à reta que une os sensores. Porém, podem existir sinais chegando com ângulos de incidência além desta faixa, do lado oposto aos sensores. Estes últimos criam uma ambiguidade, conforme pode-se observar na Figura 2.4.

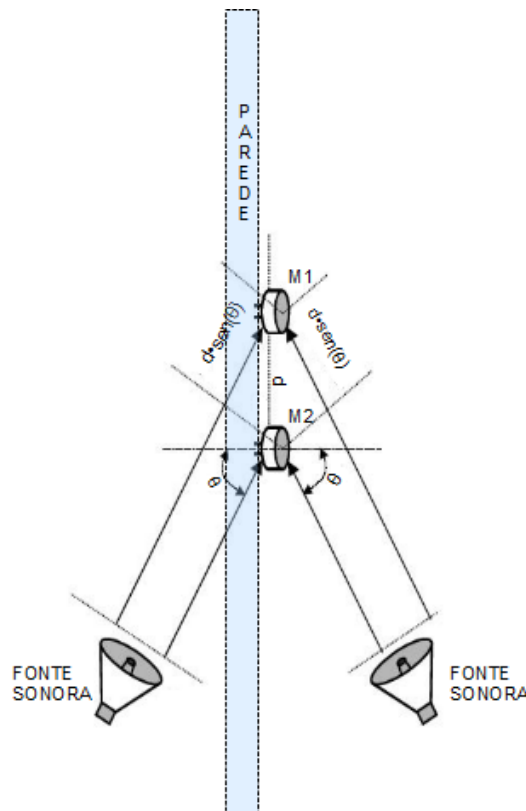


Figura 2.4: Ambiguidade entre dois sinais com ângulos de incidência espelhados horizontalmente.

Como na aplicação típica deste trabalho têm-se os sensores próximos a uma parede, seja por necessidade de instalação do equipamento, para manter a funcionalidade do recinto ou por questões estéticas, esta ambiguidade resta resolvida pela grande atenuação que os sinais sonoros sofrem ao passar através de um obstáculo como uma parede. Porém, esta configuração pode ter um custo elevado: o grande incremento das reverberações pela proximidade dos sensores a uma superfície refletora. Como solução para este grave problema, sugere-se o emprego de materiais fonoabsorventes atrás do arranjo de microfones.

Têm-se a seguir os três tipos principais de estimadores de direção de chegada dos sons, para os métodos que empregam ULA.

2.4.1 Métodos de Apontamento Direcionado

Os Métodos de Apontamento Direcionado - *Steered Beamformer* - (SBF) combinam sinais de diversos sensores separados espacialmente, de tal forma que a saída do algoritmo enfatiza os sinais vindos de uma determinada direção de busca. O resultado é que, se houver sinal naquela direção, a saída do algoritmo será alta, e se não houver sinal algum, a saída será baixa. Assim, se o arranjo "olhar" para todas as direções, aquela que possuir o sinal mais forte será a estimativa da direção de chegada dos sinais. O mais comum destes métodos é o Apontamento por Atraso e Soma - *Delay and Sum Beamformer* - (DSB), no qual os sinais provenientes de uma dada direção são amplificados através da compensação de seus atrasos (combinação aditiva), ao mesmo tempo em que se espera que os sinais das demais direções sejam combinados de forma subtrativa, devido aos atrasos não compensados entre eles. A principal vantagem desses métodos é que, com apenas um ciclo de cálculos, estimam todas as direções de sinais que chegam aos sensores, ou seja, são inerentemente capazes de determinar a direção de chegada de múltiplas fontes. Porém, isto ocorre a um custo computacional elevado, dependendo da resolução e das direções de busca pretendidas.

2.4.2 Métodos de Decomposição por AutoValores

Os Métodos de Decomposição por AutoValores - *EigenValue Decomposition* - (EVD) decompõem primeiramente a matriz de correlação cruzada dos sinais provenientes dos sensores do arranjo nos seus subespaços de sinal e de ruído. A seguir, faz-se uma busca por todos os possíveis valores de DOA, de forma a determinar as direções mais prováveis de chegada dos sinais. Tais como os métodos anteriores, estes também são inerentemente capazes de determinar múltiplas fontes, isto a um custo computacional ainda mais elevado, porém com uma resolução melhor, já que a métrica utilizada resulta em picos muito mais definidos para os atrasos detectados. O algoritmo mais conhecido deste tipo é o Classificador de Múltiplos Sinais - *Multiple Signal Classification* - (MUSIC).

2.4.3 Métodos de Estimativa de Atraso

Os Métodos de Estimativa de Atraso - *Time Delay Estimation* - (TDE) estimam inicialmente o atraso entre os sinais de cada par de sensores do arranjo, normalmente através da Correlação Cruzada Generalizada - *Generalized Cross Correlation* - (GCC), e depois

utilizam essas informações de atraso entre pares de sensores para calcular, usualmente via Erro Quadrático Mínimo - *Least Squared Error* - (LSE), a melhor estimativa de atraso global.

A simplicidade do algoritmo da Correlação Cruzada - *Cross Correlation* - (CC) e o fato de que se pode obter uma solução de forma fechada com a utilização do mesmo, ao invés de procurar em todas as direções, fê-lo uma escolha muito popular para a estimativa da DOA utilizando arranjos de microfones, pertencendo os algoritmos estudados e propostos a esta classe de estimadores.

2.5 Trabalhos correlatos

Em [Knapp e Carter 1976] observa-se a base teórica para o desenvolvimento de diversos algoritmos e métodos para a estimação da direção de chegada dos sons. Desta publicação tiveram origem o método da Correlação Cruzada Generalizada e sua evolução, a Correlação Cruzada Generalizada com Transformação de Fase - *Generalized Cross Correlation - PHASE Transform* - (GCC-PHAT). Estes métodos têm sido, desde então, os algoritmos preferidos para a estimação da DOA. Ambos são baseados em um estimador de Máxima Verossimilhança - *Maximum Likelihood* - (ML) utilizado para determinar o atraso relativo entre sinais com ruído não correlacionado recebidos por dois sensores separados espacialmente. Os sinais oriundos dos dois canais passam por um pré-filtro, e então é calculada a correlação entre eles. O número de períodos de defasagem entre os sinais é utilizado como argumento temporal, e aquele que faz com que o correlator atinja um máximo é a estimativa do atraso. Pode-se escolher os pré-filtros de forma que haja um reforço nas frequências que possuem uma boa relação sinal-ruído - *Signal to Noise Ratio* - (SNR), e uma atenuação onde o ruído predomina. No caso do algoritmo com transformação de fase, os pré-filtros são escolhidos de forma a ter-se um branqueamento do sinal, nivelando todas as amplitudes, como ver-se-á no decorrer deste trabalho.

Em [Azaria e Hertz 1984] são derivadas expressões para cálculo do Erro Quadrático Médio - *Mean Squared Error* - (MSE) na estimação do tempo de atraso para qualquer sinal estacionário, através do método da correlação cruzada, demonstrando-se serem idênticas às expressões para o cálculo dos erros na estimação de sinais gaussianos através do método da Máxima Verossimilhança.

Em [Champagne et al. 1996] são estudados, via simulações, os efeitos nocivos das reverberações no desempenho do algoritmo GCC. Especial atenção deve ser dada ao fato de que, se a fonte ou um dos receptores encontrarem-se muito próximos a uma superfície refletora (chão, paredes ou teto), o efeito da reverberação torna-se extremamente forte e ocorre uma grande degradação na performance do algoritmo.

Em [Rickard e Dietrich 2000] é apresentada a Técnica de Estimação e Separação de Sinais Degenerados - *Degenerate Unmixing and Estimation Technique* - (DUET). Ela

permite a estimativa da direção de chegada de vários sinais com somente dois receptores, respeitada a condição de ortogonalidade entre eles para cada ponto tempo-frequência em cada quadro das misturas de sinais - W-Disjoint Orthogonality. Com a obtenção de estimativas da atenuação e da defasagem entre os sinais, procede-se posteriormente ao agrupamento - *clustering* - e separação dos mesmos. Esta nova técnica de estimação de DOA é comparada através de simulações às Técnicas de Estimação de Parâmetros de Sinais via Invariância Rotacional - *Estimation of Signal Parameters via Rotational Invariance Techniques* - (ESPRIT), MUSIC e outras, demonstrando a superioridade de seus resultados para relações sinal-ruído de até 15 dB.

Em [Varma 2002] tem-se uma comparação entre as técnicas de Potência da Resposta Direcionada - *Steered Response Power* - (SRP) e GCC-PHAT. Para compensar o baixo desempenho do GCC-PHAT na presença de reverberações, criou-se um método alternativo para melhorar seus resultados, o qual foi chamado Seleção de Atraso - *Time Delay Selection* - (TIDES).

Em [Brutti et al. 2008] tem-se uma comparação entre os métodos de Decomposição Adaptativa por Autovalores - *Adaptive EigenValue Decomposition* - (AED) e GCC-PHAT. Os resultados demonstram não haver diferenças significativas entre esses métodos para elevadas relações sinal-ruído, sendo que, para relações abaixo de 15 dB, o método GCC-PHAT tem uma melhor performance, mesmo com o maior processamento do método AED.

Já em [Talantzis et al. 2005] tem-se a comparação, via simulações, entre GCC-PHAT e um método de Estimativa de Atraso por Teoria da Informação - *Information Theory Delay Estimation* -, levando-se em conta a presença de reverberações nos sinais recebidos. Variaram-se os tempos de reverberação, espaçamento dos microfones e tamanho dos blocos de dados. Observa-se que o algoritmo proposto, possivelmente por utilizar um processamento muito maior, demonstra possuir um desempenho superior ao GCC-PHAT para todas as situações.

2.6 Uma ferramenta: o espectrograma

Como o espectrograma servirá de base para a criação de um novo método de estimação de direções de chegada, segue uma breve explanação do mesmo: um espectrograma é uma representação espectral variável no tempo, formando uma imagem (Figura 2.5) que mostra o comportamento da densidade espectral de um sinal.

O formato mais comum de representação gráfica é a bidimensional, com o eixo horizontal representando o tempo e o eixo vertical representando as frequências; uma terceira

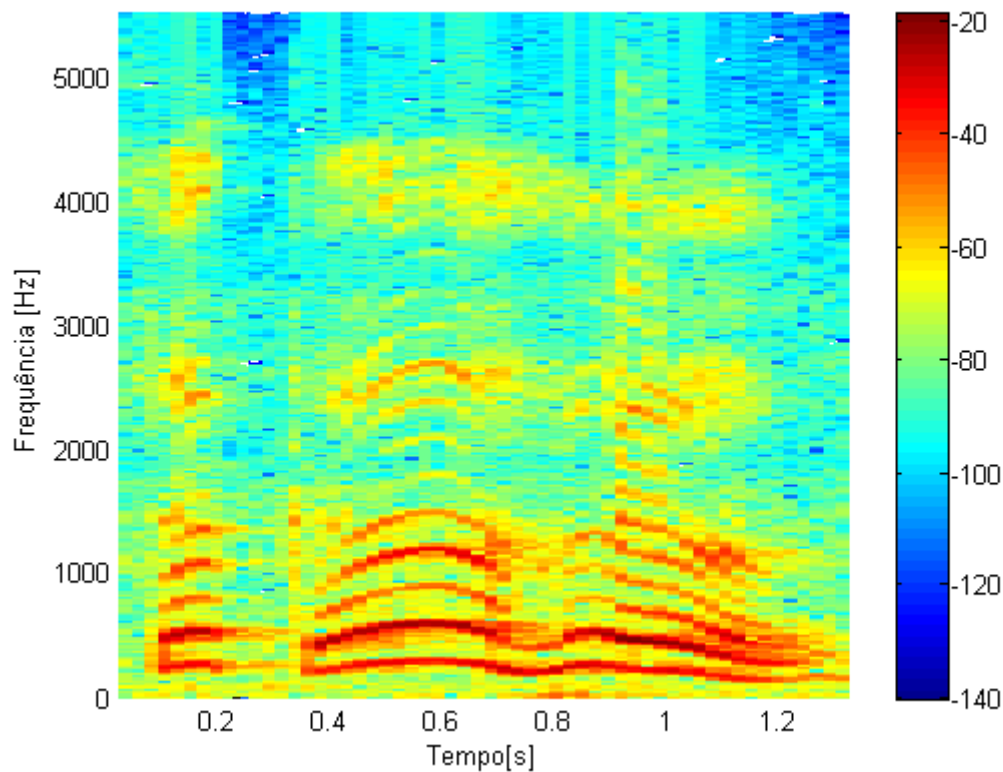


Figura 2.5: Espectrograma da palavra *Socorro*.

dimensão indicando a amplitude de uma frequência específica para um determinado quadro é representada pela intensidade ou cor de cada ponto da imagem. Há muitas variações no formato: a amplitude pode ser representada pela altura de uma superfície tridimensional ao invés de cores ou intensidades, os eixos de frequência e amplitude podem ser lineares ou logarítmicos, dependendo da finalidade do espectrograma. Para áudio, normalmente o eixo da amplitude é logarítmico, em decibéis (dB), e o eixo das frequências pode ser linear para enfatizar as relações harmônicas, ou logarítmico, para enfatizar as relações musicais (tonais).

2.7 Síntese

Neste capítulo fez-se uma revisão básica dos conceitos de ondas sonoras e suas propriedades, dando-se ênfase à reverberação, pelo importante papel que ela desempenha nos ambientes internos. Foram citados os principais métodos de estimação de direção de chegada de sons captados por arranjos lineares de sensores. Também foram elencados os trabalhos correlatos, com as conclusões mais impactantes para este trabalho. Por fim, foi apresentado o espectrograma, uma ferramenta muito útil, que servirá de base para o desenvolvimento de um dos algoritmos propostos.

O próximo capítulo apresenta os diversos sons utilizados neste trabalho.

Capítulo 3

Sons Utilizados nos Testes e Avaliações

Para avaliar-se o desempenho dos algoritmos, foram utilizados inicialmente alguns sons monofônicos de fala da língua inglesa, obtidos em repositórios públicos como [Open_Speech_Repository 2012]. Eles foram digitalizados com taxas de amostragem de 8.000 e de 22.050 amostras por segundo, e codificados com 16 bits por amostra. Foram utilizados trechos de fala mais intensa, com duração de pouco menos de três segundos para cada sinal. Os sinais foram duplicados para formarem sinais estereofônicos, sendo que as cópias dos sinais sofreram defasagens variáveis, de modo que os atrasos relativos entre os dois canais dos sinais estereofônicos fossem diferentes para cada sinal. Na etapa seguinte foram acrescentados ruídos não correlacionados em quantidades controladas (-20 e -40 dB) aos sinais estereofônicos. Além disso, um dos sinais de voz feminina foi processado por um *software* utilizado em estúdios de gravações sonoras para geração de reverberações. O resultado foi a geração de distorções de fase mínima diferenciadas por canal, emulando reverberações, que constituem-se em um dos grandes problemas de detecção e análise de sons em ambientes internos.

Posteriormente, para uma investigação mais rigorosa dos efeitos da reverberação, utilizaram-se equipamentos de gravação profissionais, com microfones omnidirecionais de referência, do fabricante Behringer, modelo ECM 8000 - <http://www.behringer.com/EN/Products/ECM8000.aspx> -, os quais possuem resposta aproximadamente plana em todo o espectro de áudio; e gravador de sons do fabricante M-Audio, modelo Fast Track Pro - <http://www.maudiodobrasil.com.br/> - para a captação, gravação e digitalização de diversas amostras de sons. Como emissores de sons, foram utilizadas duas caixas acústicas de referência do fabricante Yamaha, modelo MONITOR YAMAHA BI AMPL HS 50 M - http://uk.yamaha.com/en/products/proaudio/speakers/hs_series/hs50m/), as quais possuem resposta aproximadamente plana em toda a faixa de áudio, e têm potência RMS máxima de 70 W, cada. Adicionalmente ao fato das gravações terem sido feitas em salas pequenas, com grande efeito reverberativo e sem nenhum

tratamento acústico, teve-se também o incremento da taxa de amostragem utilizada nas gravações para 44.100 amostras por segundo, possibilitando uma maior resolução espacial para as direções de chegada dos sons. Tem-se nas Figuras 3.1, 3.2 e 3.3 imagens dos equipamentos utilizados nesta etapa.



Figura 3.1: Caixas acústicas de referência.



Figura 3.2: Microfones de referência captando os sons ambientes.



Figura 3.3: Equipamentos utilizados nos experimentos.

3.1 Síntese

Neste capítulo foram explicados os tipos de sons utilizados nos experimentos e sua origem, e também demonstrados os equipamentos utilizados para a gravação e digitalização de sinais.

O próximo capítulo traz a explicação e o equacionamento dos algoritmos existentes estudados e daqueles propostos neste trabalho.

Capítulo 4

Algoritmos Estudados

A Correlação Cruzada Generalizada (GCC) é considerada o algoritmo mais utilizado para estimação de atrasos entre sinais [Talantzis et al. 2005]. Além disso, ela serve de base para vários outros algoritmos, conforme pode-se observar neste e em outros trabalhos. Por conta disto, ela foi o primeiro algoritmo a ser estudado. Como uma escolha natural, deu-se sequência aos estudos com a sua evolução, o algoritmo GCC-PHAT; já o algoritmo DUET foi escolhido pela possibilidade de separação do sinal de interesse das fontes interferentes. Para fazer frente ao desafio representado pelas reverberações, foram criados o algoritmo Detector de Borda de Ataque, com sua extrema simplicidade e bons resultados; e a Técnica do Espectrograma, a qual apresenta uma expectativa de grande robustez, devido ao seu maior processamento.

A estimativa da diferença de tempo - *Time Delay Estimation* - (TDE) entre os sinais que chegam a um par de microfones pode ser feita através da função de correlação cruzada entre os dois sinais. A diferença de tempo que maximiza a função de correlação cruzada é o atraso - *time delay* - entre os dois sinais captados e, através deste atraso, pode-se deduzir o ângulo de chegada do sinal, conforme visto através da Equação (2.1).

Assim, para calcular a TDE, considere-se dois microfones, representados por i e j , sendo $x_i(n)$ e $x_j(n)$ os respectivos sinais por eles captados e digitalizados, onde n é o índice da amostra no tempo; aplica-se então uma função de janelamento (para evitar que ocorram descontinuidades ao início e final de cada quadro) e a Transformada Discreta de Fourier - *Discrete Fourier Transform* - (DFT) a ambos. Esse processamento resulta em $X_i(k)$ e $X_j(k)$, sinais discretos contendo a amplitude e a fase de cada frequência, sendo k o contador de frequências discretas, conforme a relação:

$$\omega_k = \frac{2 \cdot \pi \cdot k}{N}; \quad (4.1)$$

em que ω_k é a frequência discreta, em rad/s, associada ao índice k .

A Densidade Espectral da Potência Cruzada - *Cross Power Spectral Density* - (XPSD)

entre estes dois sinais é dada pela seguinte expressão:

$$\Phi_{X_i X_j} = X_i(k) \cdot X_j^*(k); \quad (4.2)$$

na qual * denota o complexo conjugado.

A correlação cruzada entre estes dois sinais é dada pela transformada inversa de Fourier da XPSD. Para sinais amostrados e frequências discretas utiliza-se a seguinte aproximação:

$$\hat{R}_{X_i X_j}(l) = \frac{1}{M} \sum_{k=0}^{M-1} \hat{\Phi}_{X_i X_j}(k) \cdot e^{\frac{j \cdot 2 \cdot \pi \cdot k \cdot l}{M}}; \quad (4.3)$$

na qual M é o número de termos da XPSD; $\hat{\Phi}_{X_i X_j}$ é a estimativa da XPSD entre os sinais X_i e X_j ; e l é o índice relativo ao atraso. Como a XPSD pode ser apenas estimada para observações finitas, pode-se apenas estimar o atraso com o valor desta correlação, sendo que a mesma pode ser calculada para atrasos positivos e negativos. O índice l que maximiza o valor da correlação é igual ao número de amostras correspondente à defasagem entre os dois sinais. A expressão que determina esta defasagem é:

$$\tau_{i,j} = \frac{1}{f_s} \cdot \operatorname{argmax}[\hat{R}_{X_i X_j}(l)]. \quad (4.4)$$

4.1 Correlação Cruzada Generalizada

Conforme visto na Equação (4.3), a correlação cruzada pode ser calculada através da expressão

$$\hat{R}_{X_i X_j}(l) = \frac{1}{M} \sum_{k=0}^{M-1} \hat{\Phi}_{X_i X_j}(k) \cdot e^{\frac{j \cdot 2 \cdot \pi \cdot k \cdot l}{M}};$$

na qual Φ é a XPSD [Equação (4.2)], sendo definida pela equação:

$$\Phi_{X_i X_j} = X_i(k) \cdot X_j^*(k).$$

De forma mais geral, conforme [Knapp e Carter 1976], tem-se a generalização da correlação cruzada - *Generalized Cross Correlation* - (GCC) por aplicar um pré-filtro aos sinais originais, antes do cálculo da correlação, cujas expressões são dadas por:

$$y_i = h_i(n) * x_i(n); \quad (4.5)$$

$$y_j = h_j(n - l) * x_j(n - l); \quad (4.6)$$

nas quais * denota a função convolução.

Como consequência, no domínio frequência tem-se a Densidade Espectral da Potência Cruzada Generalizada - *Generalized Cross Power Spectral Density* - (GXPSD), dada por:

$$\Phi_{Y_i Y_j}^g(k) = [H_i(k) \cdot X_i(k)] \cdot [H_j(k) \cdot X_j(k)]^*. \quad (4.7)$$

Combinando-se os dois pré-filtros,

$$\Psi_{i,j}(k) = H_i(k) \cdot H_j^*(k); \quad (4.8)$$

obtem-se então a expressão

$$\Phi_{Y_i Y_j}^g(k) = \Psi_{i,j}(k) \cdot X_i(k) \cdot X_j^*(k); \quad (4.9)$$

através da qual pode-se obter finalmente a GCC, calculando-se a inversa da DFT da GXPSD:

$$R_{Y_i Y_j}^g(l) = \frac{1}{M} \sum_{k=0}^{M-1} \Phi_{Y_i Y_j}^g(k) \cdot e^{\frac{j \cdot 2 \cdot \pi \cdot k \cdot l}{M}}. \quad (4.10)$$

Neste ponto o algoritmo calcula o valor de l que maximiza o valor da correlação, obtendo a estimativa da defasagem entre os sinais, de forma análoga à Equação (4.4).

4.2 Correlação Cruzada Generalizada com Transformação de Fase

A Transformação de Fase - *PHase Transform* - (PHAT) é a função mais utilizada como pré-filtro para a GCC, sendo que a mesma atua como um nivelador de amplitudes, tentando obter a composição do ruído branco. Sua escolha visa garantir uma maior robustez do algoritmo, já que a mesma busca igualar as amplitudes de todas as frequências do sinal, obtendo um espectro o mais parecido possível com o do ruído branco (amplitude constante para todas as frequências). Através da propriedade que o ruído branco possui de não correlacionar-se com nenhum outro sinal, exceto consigo mesmo, espera-se obter um resultado próximo a zero para qualquer correlação entre dois sinais distintos, ou mesmo entre um sinal e sua réplica defasada; e um pico muito agudo (idealmente, a função impulso) para a correlação do sinal sincronizado com sua réplica.

Por uma questão de familiaridade com o domínio das frequências contínuas e facilidade de visualização, tem-se a seguir, para dois sinais analógicos idênticos defasados no tempo, a sua demonstração. Inicia-se com a equação do pré-filtro no domínio frequência:

$$\Psi^g(\omega) = \frac{1}{|\hat{\Phi}_{X_i X_j}(\omega)|}; \quad (4.11)$$

em que ω é a frequência analógica. Pode-se então calcular a estimativa da GCC-PHAT através de

$$R_{X_i X_j}^{PHAT}(\tau) = \frac{1}{2 \cdot \pi} \int_{-\pi}^{\pi} \frac{1}{|\hat{\Phi}_{X_i X_j}(\omega)|} \cdot \hat{\Phi}_{X_i X_j}(\omega) \cdot e^{j \cdot \omega \cdot \tau} d\omega. \quad (4.12)$$

Utilizando a expressão

$$\hat{\Phi}_{X_i X_j}(\omega) = \hat{\Phi}_{X_i X_i}(\omega) \cdot e^{-j \cdot \omega \cdot D}; \quad (4.13)$$

na qual D é o atraso relativo entre os sinais, obtém-se

$$R_{X_i X_j}^{PHAT}(\tau) = \frac{1}{2 \cdot \pi} \int_{-\pi}^{\pi} \frac{1}{|\hat{\Phi}_{X_i X_i}(\omega) \cdot e^{-j \cdot \omega \cdot D}|} \cdot \hat{\Phi}_{X_i X_i}(\omega) \cdot e^{-j \cdot \omega \cdot D} \cdot e^{j \cdot \omega \cdot \tau} d\omega; \quad (4.14)$$

resultando na expressão

$$R_{X_i X_j}^{PHAT}(\tau) = \frac{1}{2 \cdot \pi} \int_{-\pi}^{\pi} e^{-j \cdot \omega \cdot D} \cdot e^{j \cdot \omega \cdot \tau} d\omega. \quad (4.15)$$

A estimativa se reduz à função impulso para um atraso igual a D , e zero para os demais valores, conforme pode-se observar na equação (4.16).

$$R_{X_i X_j}^{PHAT}(\tau) = \delta(\tau - D). \quad (4.16)$$

Ressalta-se que, para chegar a esta expressão, foi utilizada a premissa de que os dois sinais eram idênticos, apenas defasados temporalmente. Na prática têm-se contaminações por ruído e atenuações (e até mesmo defasagens) variáveis com a frequência, atuando de forma assimétrica nos dois canais. Mesmo assim, resultados aproximados podem ser obtidos, com picos agudos correspondendo aos atrasos relativos entre os sinais, e valores bem mais baixos para os demais valores de defasagem.

Trabalhar no domínio discreto traz vantagens como a simplicidade e a rapidez dos programas computacionais, e desvantagens como os problemas de quantização (o atraso relativo provavelmente não será um múltiplo do período de amostragem). Pela melhor relação custo-benefício, opta-se por trabalhar no domínio discreto. Assim, para implantar o algoritmo GCC-PHAT para sinais amostrados utiliza-se como pré-filtro, ou função peso, a expressão

$$\Psi^{PHAT}(k) = \frac{1}{|\hat{\Phi}_{X_i X_j}(k)|}. \quad (4.17)$$

Valendo-se da relação

$$\hat{\Phi}_{X_i X_j}(k) = \hat{\Phi}_{X_i X_i}(k) \cdot e^{\frac{-j \cdot 2 \cdot \pi \cdot k \cdot D}{M}}; \quad (4.18)$$

em que D , no caso discreto, é o número de amostras de defasagem entre os sinais, pode-se obter então

$$\Psi^{PHAT}(k) = \frac{1}{|\hat{\Phi}_{X_i X_j}(k)|} = \frac{1}{|\hat{\Phi}_{X_i X_i}(k) \cdot e^{\frac{-j \cdot 2 \cdot \pi \cdot k \cdot D}{M}}|}. \quad (4.19)$$

Assim, calcula-se a correlação através de:

$$R_{Y_i Y_j}^{PHAT}(l) = \frac{1}{M} \cdot \sum_{k=0}^{M-1} \Psi^{PHAT}(k) \cdot \hat{\Phi}_{X_i X_j}(k) \cdot e^{\frac{j \cdot 2 \cdot \pi \cdot k \cdot l}{M}} = \quad (4.20)$$

$$= \frac{1}{M} \cdot \sum_{k=0}^{M-1} \frac{\hat{\Phi}_{X_i X_j}(k)}{|\hat{\Phi}_{X_i X_j}(k)|} \cdot e^{\frac{j \cdot 2 \cdot \pi \cdot k \cdot l}{M}} = \quad (4.21)$$

$$= \frac{1}{M} \cdot \sum_{k=0}^{M-1} \frac{\hat{\Phi}_{X_i X_i}(k) \cdot e^{\frac{-j \cdot 2 \cdot \pi \cdot k \cdot D}{M}}}{|\hat{\Phi}_{X_i X_i}(k) \cdot e^{\frac{-j \cdot 2 \cdot \pi \cdot k \cdot D}{M}}|} \cdot e^{\frac{j \cdot 2 \cdot \pi \cdot k \cdot l}{M}} = \quad (4.22)$$

$$= \frac{1}{M} \cdot \sum_{k=0}^{M-1} e^{\frac{-j \cdot 2 \cdot \pi \cdot k \cdot D}{M}} \cdot e^{\frac{j \cdot 2 \cdot \pi \cdot k \cdot l}{M}} = \quad (4.23)$$

$$= \frac{1}{M} \cdot \sum_{k=0}^{M-1} e^{\frac{j \cdot 2 \cdot \pi \cdot k \cdot (l-D)}{M}}. \quad (4.24)$$

Com o auxílio da Equação (4.24) pode-se perceber que, para sinais discretizados, os resultados tendem a ser próximos a 0 para valores de índices de atraso l diferentes do atraso verdadeiro D , devido ao somatório de funções senoidais ao longo do tempo resultar em 1, e então ser dividido pelo número de termos M ; e tendem a ser próximos a 1 quando o índice de atraso l coincidir com o atraso verdadeiro D .

4.3 Técnica de Estimação e Separação de Sinais Degenerados

A Técnica de Estimação e Separação de Sinais Degenerados - *Degenerate Unmixing and Estimation Technique* - (DUET) foi criada originalmente para a separação de diversas fontes sonoras, sendo utilizada quando o número de misturas disponíveis (número de canais captados por microfones distintos) é menor que o número de fontes sonoras, já que, neste caso, o problema não pode ser resolvido através da estimativa da inversa da matriz de misturas, advindo daí a denominação de sinais degenerados [Jourjine et al. 2000]. Utiliza como premissa a existência de transformadas separadas (*W-disjoint orthogonality* ou separação tempo-frequência) para cada componente de cada sinal.

Tal técnica baseia-se no algoritmo da Correlação Cruzada para obtenção da atenuação e defasagem - *delay* - relativos dos sinais correspondentes às misturas, isto para cada frequência em cada quadro (*time-frequency point*). A seguir, ao invés de achar o atraso que melhor atende a todas as frequências de todos os quadros, a técnica utiliza um algoritmo de agrupamento - *clustering* - similar ao algoritmo Esperança-Maximização - *Expectation-Maximization* - (EM), utilizado para o cálculo de Máxima Verossimilhança - *Maximum Likelihood* - (ML). Este algoritmo agrupa os pontos tempo-frequência utilizando como função peso para o cálculo do centro do grupo - *cluster* - o valor calculado através da correlação cruzada para cada ponto contido no grupo. O número de grupos, correspondente ao número de sinais detectados, deve ser escolhido previamente, de forma que, se a estimativa do número de fontes sonoras estiver errada, o resultado restará prejudicado. Na Figura 4.1 tem-se um exemplo de agrupamento dos pontos tempo-frequência no entorno de duas regiões, resultando na separação dos sons captados em dois agrupamentos, cada qual com sua estimativa de ângulo de chegada e atenuação.

AGRUPAMENTO DE SINAIS DE ACORDO COM A ATENUAÇÃO E A DEFASAGEM RELATIVAS

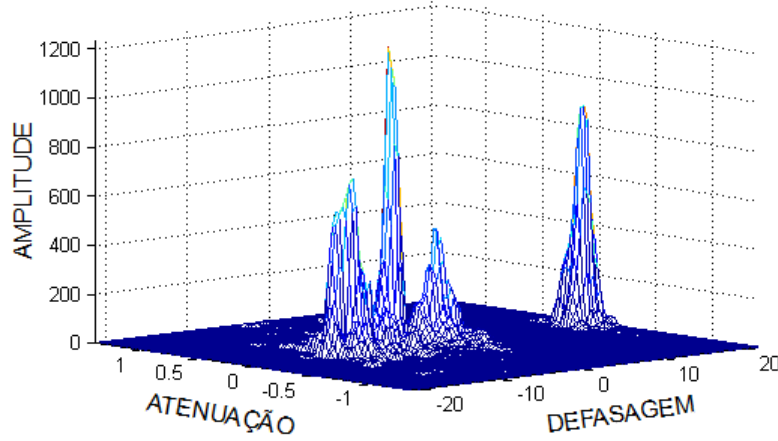


Figura 4.1: Agrupamento dos pontos tempo-frequência em duas regiões.

Para implementar o algoritmo DUET neste trabalho, foram utilizados os quadros de tempo versus frequências calculados pelo algoritmo GCC, e então seus valores foram separados em dois grupos (cada um correspondendo a um canal), com a utilização iterativa das equações obtidas em [Rickard e Dietrich 2000].

$$\hat{\delta}_i = \frac{\sum_{(\omega, \tau) \in \omega_i} \|X_1^W(\omega, \tau)\|^2 \cdot \text{imag}[\log(\frac{X_1^W(\omega, \tau)}{X_2^W(\omega, \tau)})]/\omega}{\sum_{(\omega, \tau) \in \omega_i} \|X_1^W(\omega, \tau)\|^2}; \quad (4.25)$$

$$\hat{a}_i = \frac{\sum_{(\omega, \tau) \in \omega_i} \|X_1^W(\omega, \tau)\|^2 \cdot \|\frac{X_2^W(\omega, \tau)}{X_1^W(\omega, \tau)}\|}{\sum_{(\omega, \tau) \in \omega_i} \|X_1^W(\omega, \tau)\|^2}; \quad (4.26)$$

nas quais $\hat{\delta}_i$ é o atraso estimado do agrupamento i , $\hat{a}_i$ é a atenuação estimada do agrupamento i , ω_i é o conjunto de pontos do agrupamento i , X_i^W é o sinal da mistura i janelado e transformado pela DFT, e imag é a parte imaginária.

Para o caso de um único sinal estereofônico, $\hat{\delta}_i$, com $i=1$, representará o atraso relativo médio entre os dois canais, em número de amostras. Com esse atraso, estima-se o ângulo de chegada do sinal aos sensores.

Devido aos erros elevados obtidos com a utilização dos métodos tradicionais (GCC, GCC-PHAT e DUET) para sinais gravados em ambientes com reverberação (conforme poderá ser visto no capítulo correspondente à análise dos resultados), foram criados dois novos métodos para melhorar a precisão na detecção de ângulos para os casos de ambientes reverberativos, sendo eles apresentados a seguir.

4.4 Método da Borda de Ataque

Considerando-se que o ar é um meio isotrópico (i.e., a velocidade de propagação de onda é a mesma em todas as direções), é fácil concluir que o caminho de propagação mais curto é também o mais rápido. Por isso, tem-se sempre, nos instantes iniciais (ou ataque) de um som, um intervalo de tempo no qual o único sinal sonoro a chegar ao microfone o fez pelo caminho direto entre a fonte sonora e esse. Por um curto período de tempo, pode-se utilizar estas amostras recebidas sem "contaminação" pela reverberação para o cálculo do atraso entre dois sinais captados por microfones em posições distintas. Sendo assim, o desafio é detectar onde inicia-se um novo trecho de fala, ou mesmo um som característico indicando a necessidade de ajuda, após um intervalo de relativo silêncio. O método resultante desta tentativa foi denominado Detector de Borda de Ataque. Tem-se uma ilustração das trajetórias direta e da 1ª reflexão na Figura 4.2.

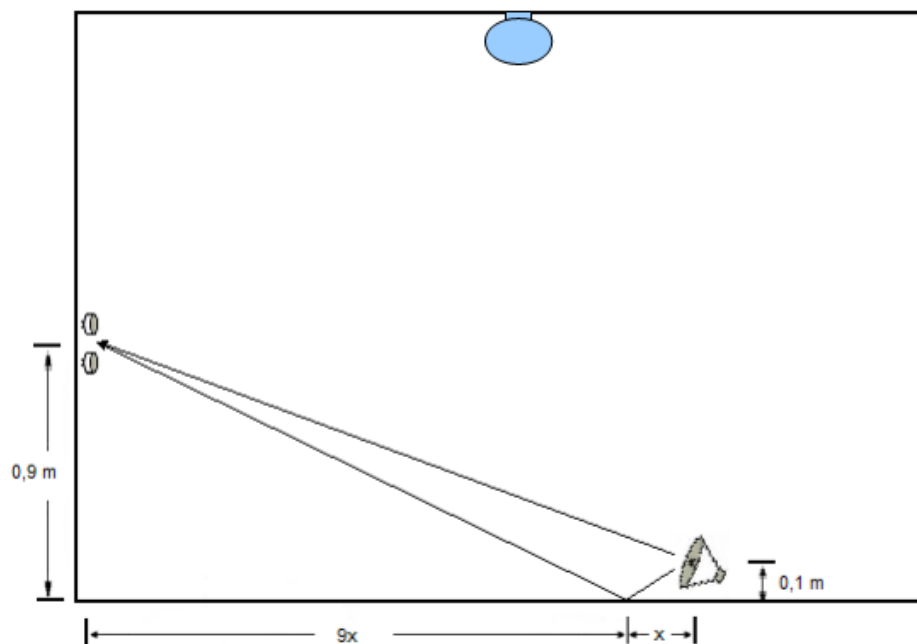


Figura 4.2: Ilustração dos caminhos do sinal direto e da 1ª reflexão no piso.

Para um conjunto de microfones à 0,9 m do chão, e uma fonte sonora a uma distância suficientemente longa¹ do conjunto de microfones e à 0,1 m do chão, simulando uma pessoa caída, tem-se que a diferença entre o caminho direto e o caminho contendo uma

¹Embora não haja consenso sobre a distância mínima necessária para considerarem-se paralelas as ondas sonoras que chegam aos sensores, pode-se demonstrar que o efeito da aproximação da fonte sonora aos mesmos pode ser desprezado, por ser muito inferior a uma amostra, considerando taxas de amostragem de até 44.100 Hz, e distâncias tão curtas quanto 1 m.

reflexão do som no solo é de, no máximo, 0,2 m, correspondendo a um tempo de 0,68 ms, o que, para uma taxa de amostragem relativamente elevada de 44,1 kHz por canal, indica que nesta situação ocorrem menos de 30 amostragens livres de reverberação, as quais constituem o quadro a ser analisado pelo algoritmo. Observa-se na Figura 4.3 como são os sinais captados pelos sensores.

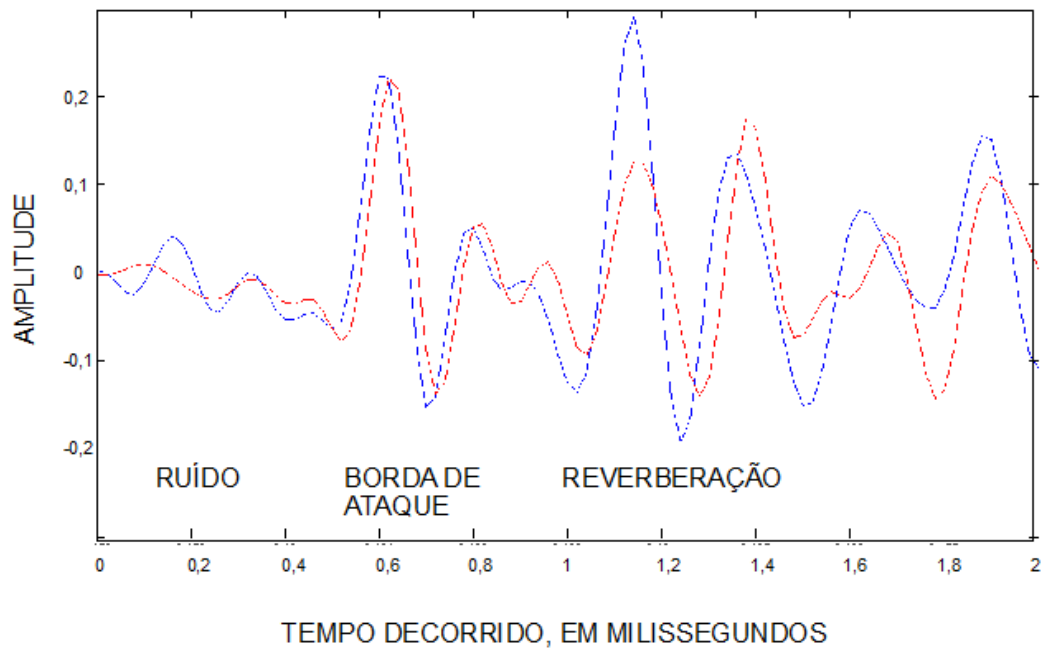


Figura 4.3: Ondas sonoras características captadas pelos sensores.

Por economicidade, parametrizou-se este algoritmo para atuar apenas acima de um determinado limiar de nível sonoro (L), e para utilizar um número reduzido de amostras para processamento (W), comparando os quadros de cada canal no domínio tempo. A escolha do domínio tempo para efetuar o cálculo da correlação entre os sinais deve-se ao fato de que no domínio frequência ter-se-iam problemas decorrentes do janelamento de quadros com poucas amostras. Se os quadros são pequenos, ocorre a tendência à maximização do valor da correlação para atrasos nulos (alinhamento de quadros pelo janelamento), conforme pode-se observar na Figura 4.4.

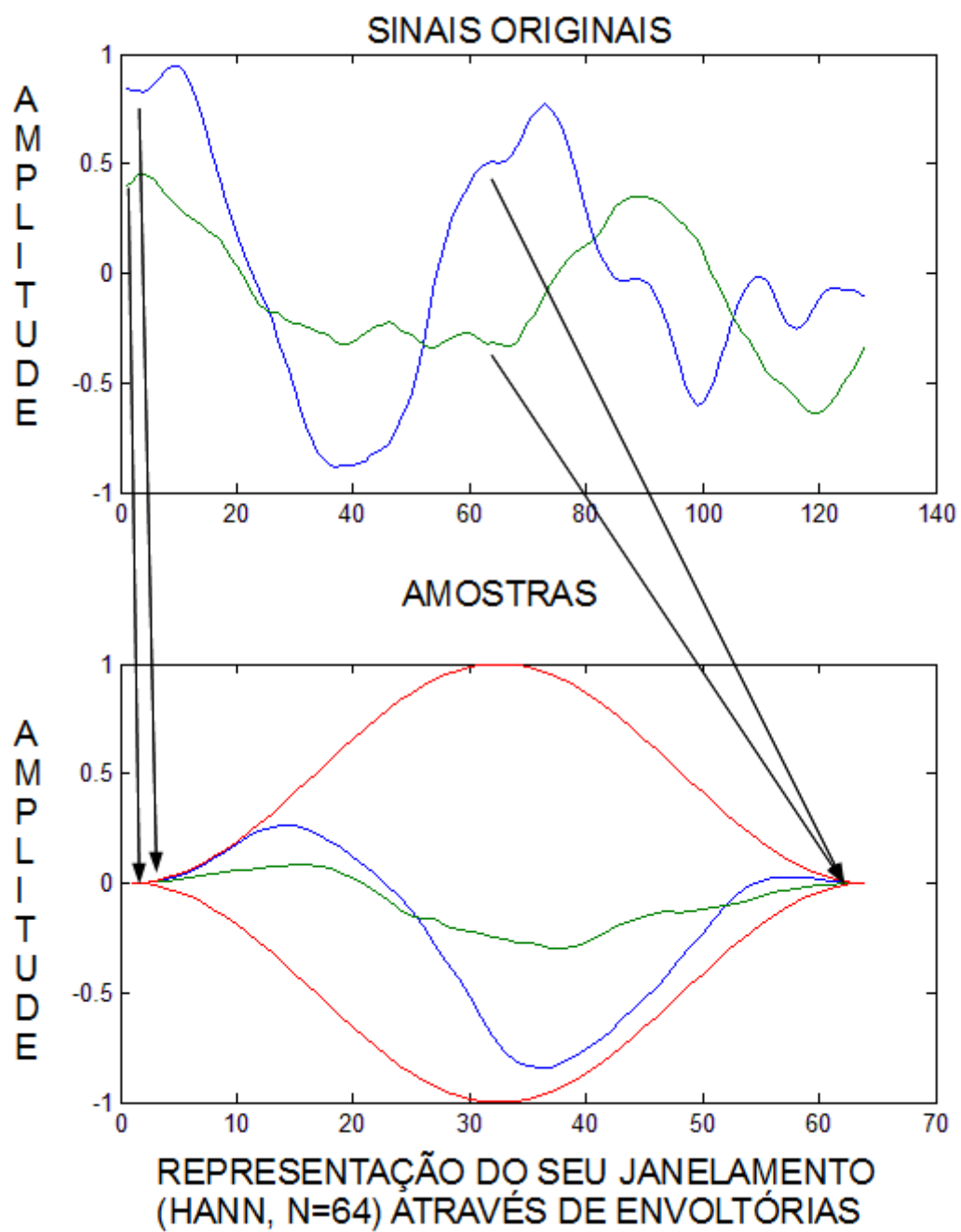


Figura 4.4: Quadros pequenos de sinais defasados e o efeito do seu janelamento, representado por envoltórias.

4.5 Técnica do Espectrograma

Propõe-se agora, para fins de comparação, outra tentativa de detecção da direção de chegada dos sons em ambientes com forte reverberação. Este método baseia-se na ideia de que é possível observar padrões visuais semelhantes em espectrogramas de sinais captados por canais distintos, sendo possível também, em alguns casos, a estimação visual do atraso existente entre os mesmos, conforme pode-se observar na Figura 4.5.

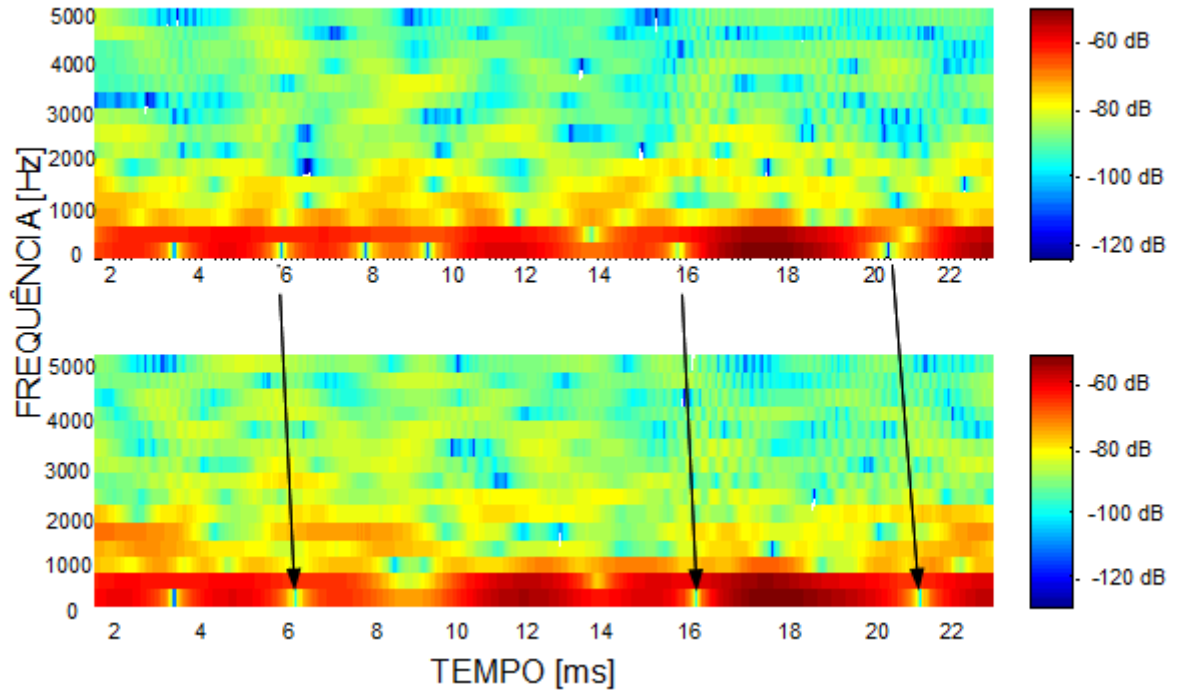


Figura 4.5: Espectrogramas de sinais com reverberações e defasados temporalmente.

A técnica consiste na comparação de um grande número de quadros de dois sinais no domínio frequência, buscando, via erro quadrático mínimo, o ponto ótimo de defasagem entre eles. A equação através da qual isto é feito é:

$$S_{X_i X_j}(l) = \frac{1}{N} \sum_{k=-N/2}^{N/2-1} \text{abs} \left[\log \frac{X_i(k)}{X_j(k-l)} \right]^2; \quad (4.27)$$

em que $S_{X_i X_j}(l)$ é o somatório de todos os N termos da comparação entre os dois quadros, deslocados entre si por l amostras de atraso (l podendo ser positivo ou negativo). Por fim, obtém-se a estimativa do atraso através da equação:

$$\tau_{i,j} = \frac{1}{f_s} \cdot \text{argmin}[S_{X_i X_j}(l)]. \quad (4.28)$$

Assim, criou-se um algoritmo que, após a escolha do limiar de detecção dos sinais a serem analisados, da resolução espectral (o tamanho dos quadros é o dobro da quantidade de frequências W escolhida) e da quantidade de quadros N a serem comparados, realiza-se uma sequência de operações, que incluem: o cálculo, via algoritmo GCC, dos quadros componentes das matrizes tempo versus frequência; a seleção dos quadros com uma defasagem l escolhida para essa iteração; o cálculo do logaritmo dos valores dos quadros; a subtração dos valores calculados de um dos sinais pelos valores calculados do outro sinal; o cálculo do quadrado dos valores obtidos da etapa anterior; a soma de todos os valores obtidos da etapa anterior. Por fim, se já foram utilizados todos os valores de l para originarem os somatórios correspondentes, calcula-se o ângulo de chegada dos sinais sonoros aos sensores relativo ao menor valor obtido através dos somatórios. Por ter-se obtido este valor via minimização da soma do quadrado do logaritmo da divisão entre quadros, espera-se obter uma estimativa confiável para o atraso entre os sinais para esta detecção. Para melhor compreensão, tem-se na Figura 4.6 um fluxograma do algoritmo.

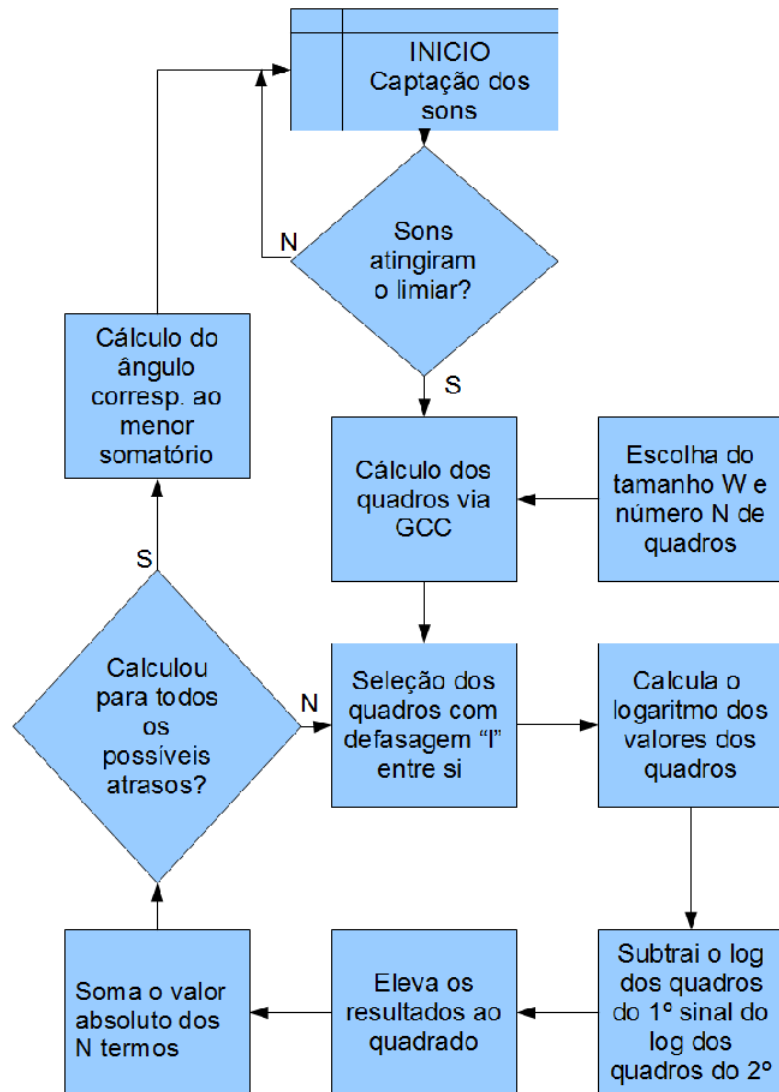


Figura 4.6: Fluxograma do algoritmo da Técnica do Espectrograma.

4.6 Síntese

Neste capítulo foram explicados e equacionados os algoritmos existentes, quais sejam, o GCC, o GCC-PHAT e o DUET; e também apresentados e explicados os dois novos algoritmos propostos, o Detector de Borda de Ataque, e a Técnica do Espectrograma.

No próximo capítulo ver-se-ão a escolha de parâmetros e os resultados das simulações com esses algoritmos para os diversos tipos de sinais.

Capítulo 5

Resultados e Discussões

5.1 Escolha de parâmetros

Para utilizar os algoritmos, inicialmente deve-se escolher os parâmetros mais apropriados para cada finalidade. Por pretender monitorar continuamente um ambiente, deve-se priorizar um curto ciclo de execução, para não haver perda de dados por falta de capacidade de processamento, isto sem reduzir significativamente a qualidade dos resultados obtidos. A primeira escolha a ser feita é qual o tipo de janelamento a ser aplicado aos quadros, para que os dados sofram a mínima alteração possível [Ramirez 1985]. Foram utilizadas as janelas de Hann e de Blackman, por ambas possuírem um comportamento espectral adequado, distorcendo pouco o espectro de frequências dos dados, além do fato de serem de fácil implementação, com baixo custo computacional. Para a janela de Hann o vetor de pesos é dado por:

$$h(n) = 0,5 - 0,5 \cdot \cos\left(\frac{2 \cdot \pi \cdot n}{N-1}\right), \quad 0 \leq n \leq N-1; \quad (5.1)$$

em que N é o tamanho do quadro em número de amostras. E para a janela de Blackman o vetor de pesos é dado por:

$$h(n) = a_0 - a_1 \cdot \cos\left(\frac{2 \cdot \pi \cdot n}{N-1}\right) + a_2 \cdot \cos\left(\frac{4 \cdot \pi \cdot n}{N-1}\right), \quad 0 \leq n \leq N-1; \quad (5.2)$$

em que $a_0 = (1 - \alpha)/2$, $a_1 = 1/2$ e $a_2 = \alpha/2$. Adotando-se o valor do parâmetro α como 0,16, visando obter-se baixa distorção espectral, resulta na expressão

$$h(n) = 0,42 - 0,5 \cdot \cos\left(\frac{2 \cdot \pi \cdot n}{N-1}\right) + 0,08 \cdot \cos\left(\frac{4 \cdot \pi \cdot n}{N-1}\right), \quad 0 \leq n \leq N-1. \quad (5.3)$$

Tem-se a seguir o gráfico 5.1, o qual compara os dois tipos distintos de janelamento para um sinal de fala da língua inglesa, obtido em um repositório público, com 11 amostras de defasagem entre réplicas do mesmo, correspondente a uma direção de chegada de 30° para uma frequência de amostragem de 22.050 Hz e uma distância de 34 cm entre os sensores.

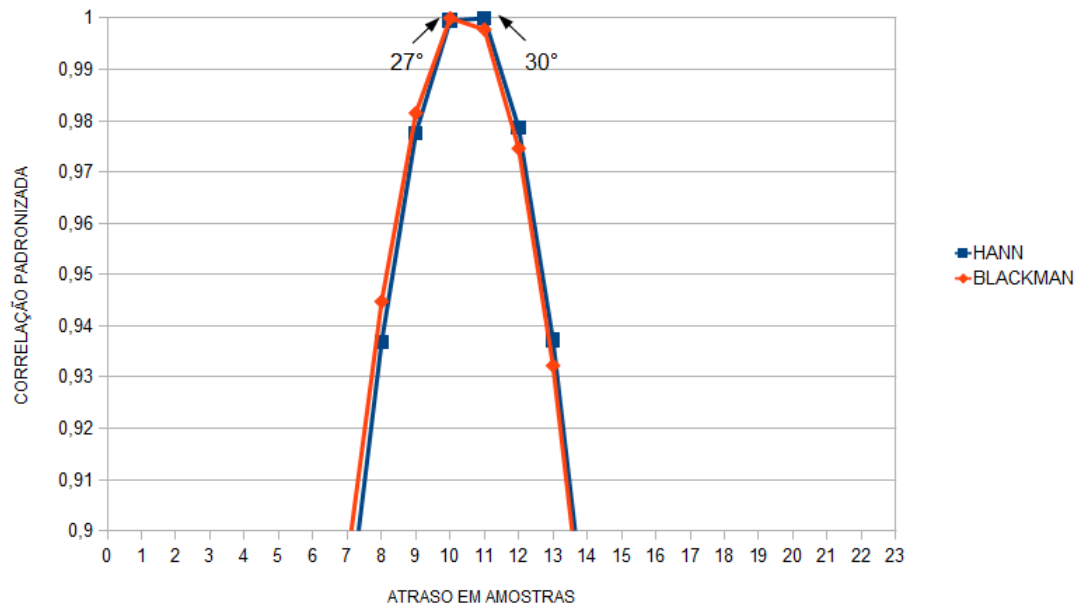


Figura 5.1: Detecção de atraso com a utilização de diferentes janelamentos.

Conforme pode-se observar, o janelamento de Hann aponta para o resultado correto, que é um atraso correspondente a 11 amostras, ao passo que o janelamento de Blackman resulta em um atraso de 10 amostras apenas.

Em seguida foram feitas simulações variando o tamanho do quadro, para ambas as janelas, para 40 trechos de falas da língua inglesa digitalizadas com uma taxa de amostragem de 22.050 Hz, obtidas em um repositório público. Os resultados encontram-se na tabela 5.1.

Tipo de Janela \ Tamanho do Quadro	128	256	512
Hann	19	39	38
Blackman	12	38	36

Tabela 5.1: Número de acertos para 40 simulações com diversos tamanhos de quadros.

Observa-se pela tabela que a janela de Hann tem um desempenho superior ao da janela de Blackman para todos os tamanhos de quadros, e que quadros com 256 amostras tiveram uma taxa de acertos maior. Assim, optou-se pela utilização do janelamento de Hann nos algoritmos deste trabalho.

A próxima escolha a ser feita diz respeito ao tamanho do quadro para a estimação do atraso. Em princípio, o tamanho ideal de quadro é muito dependente do sinal utilizado, variando bastante com a faixa e a quantidade de frequências presentes no sinal. Na tentativa de obter-se uma relação custo-benefício adequada, simulou-se a detecção de atraso para o mesmo sinal com 11 amostras de defasagem, e obteve-se, para diversos tamanhos de quadros, o gráfico 5.2.

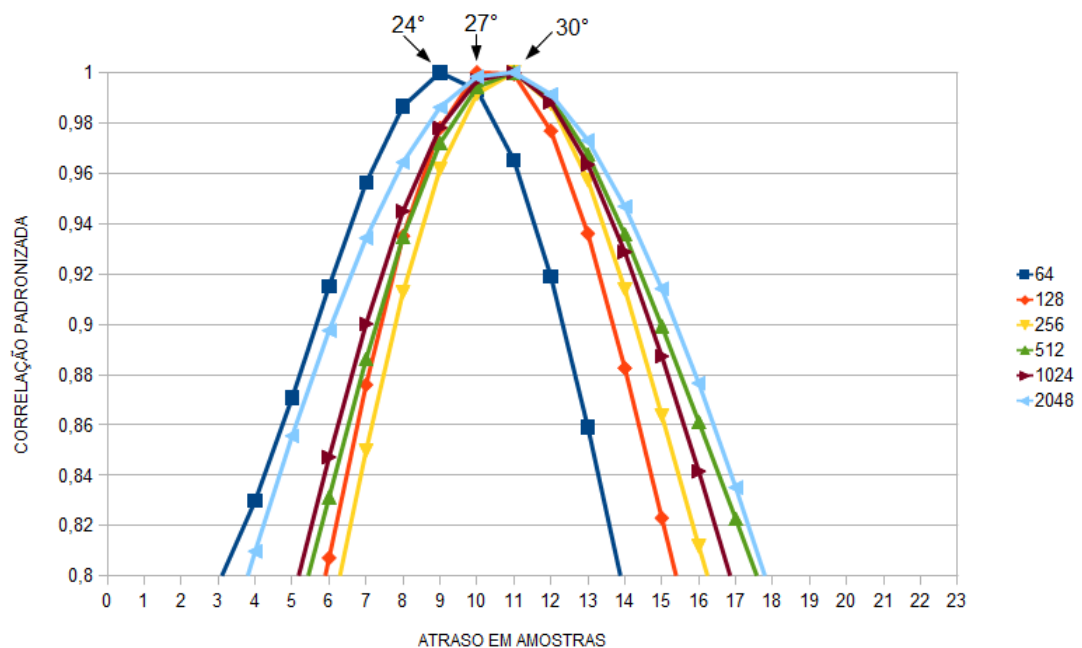


Figura 5.2: Assertividade do algoritmo GCC para diversos tamanhos de quadros.

Pode-se observar que os dois primeiros tamanhos de quadros, de 64 e 128 amostras, nos conduzem a resultados errôneos, tendo em vista que o atraso correto é de 11 amostras, correspondendo a 30° , para esta simulação. Por este motivo, e também pelos resultados obtidos anteriormente com os diferentes janelamentos, optou-se pelo tamanho de quadro imediatamente superior, de 256 amostras por quadro.

Tem-se, na sequência, o gráfico 5.3, com o resultado do uso do algoritmo GCC com quadros de 256 amostras e janelamento do tipo Hann, para três sinais de fala da língua inglesa obtidos no mesmo repositório público, com atrasos distintos entre cada sinal e sua réplica: o primeiro sem defasagem (0°); o segundo com um atraso de 0,5 ms, correspondendo a 11 amostras (ou 30°); e o terceiro com um atraso de 1 ms, correspondendo a 22 amostras (ou 86°) para a mesma frequência de amostragem de 22.050 Hz.

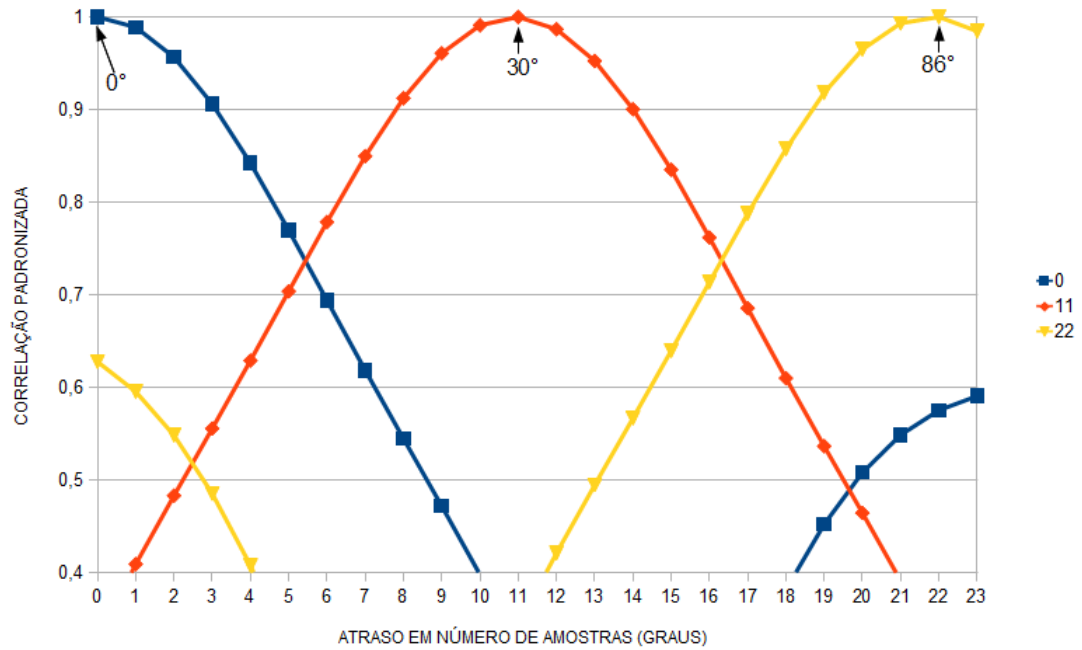


Figura 5.3: Algoritmo GCC para atrasos de 0, 11 e 22 amostras.

Observa-se que os atrasos foram determinados corretamente nos três casos. Além disso, vê-se que o algoritmo apresenta como resultado curvas com decaimento suave no entorno dos valores detectados.

Tem-se agora o gráfico 5.4, o qual demonstra os resultados obtidos com o algoritmo GCC-PHAT, para os mesmos sinais.

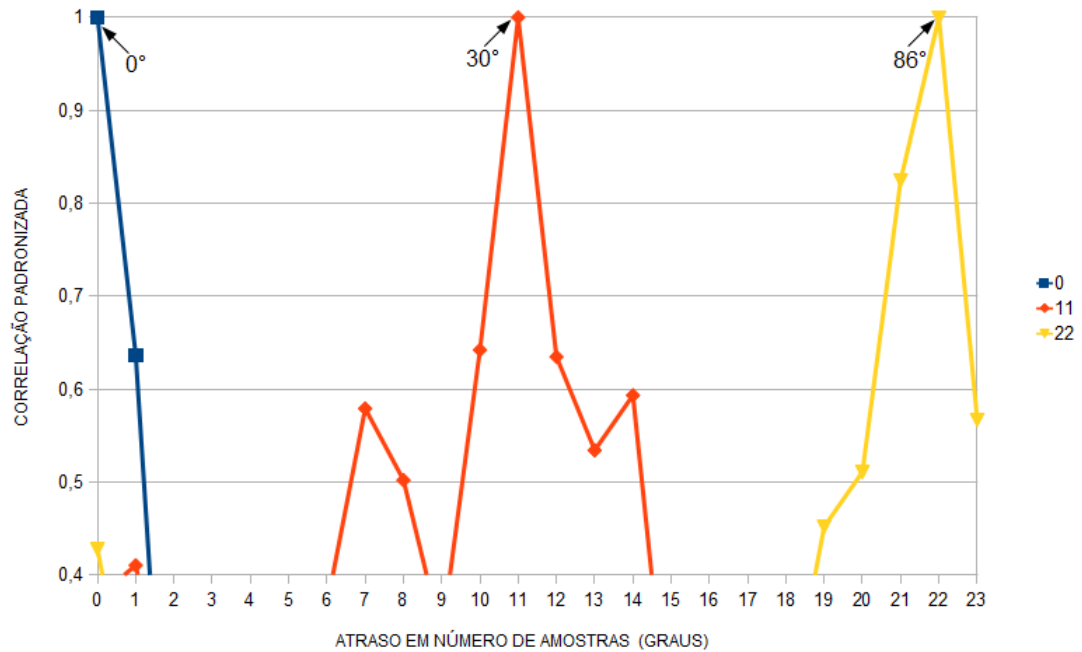


Figura 5.4: Algoritmo GCC-PHAT para atrasos de 0, 11 e 22 amostras.

Nota-se que, além dos atrasos serem determinados corretamente, este algoritmo apresenta uma seletividade muito maior, caracterizada pelos picos agudos no número de amostras correspondente aos atrasos.

Continuando com as simulações, têm-se os gráficos 5.5 e 5.6, os quais apresentam o resultados dos algoritmos GCC e GCC-PHAT, respectivamente, quanto à adição de ruído do tipo branco aos sinais.

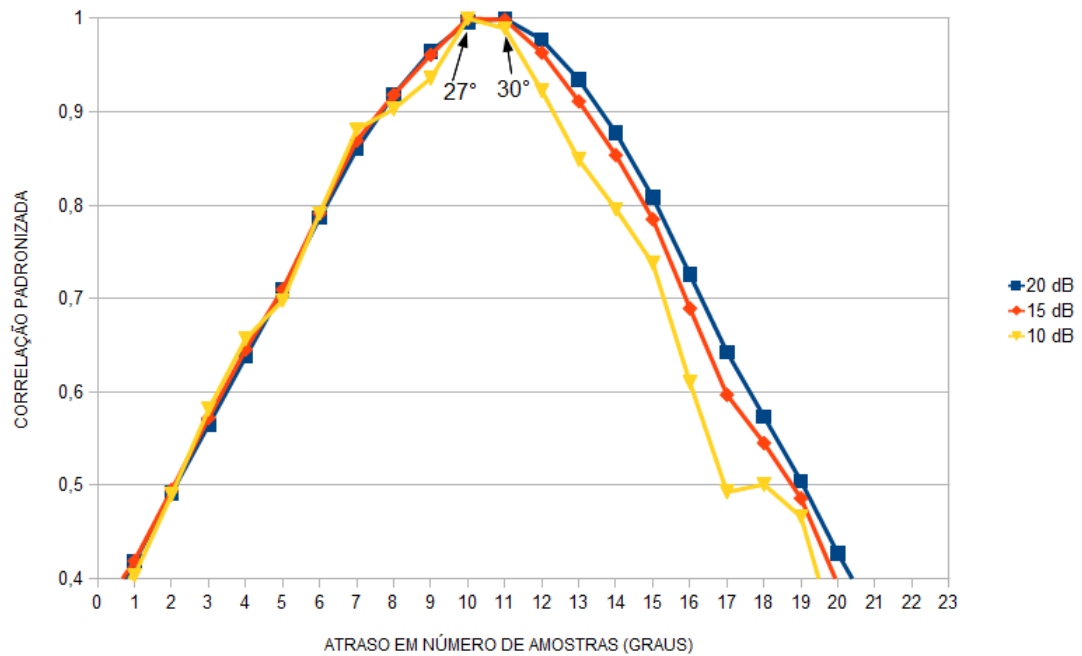


Figura 5.5: Algoritmo GCC para diversas relações sinal-ruído.

Para o algoritmo GCC tem-se que, para 10 dB de relação sinal-ruído, há a indicação errônea do valor do atraso, de 10 amostras (ou 27°) ao invés de 11 amostras (30°).

Tem-se agora o comportamento do algoritmo GCC-PHAT em relação ao ruído.

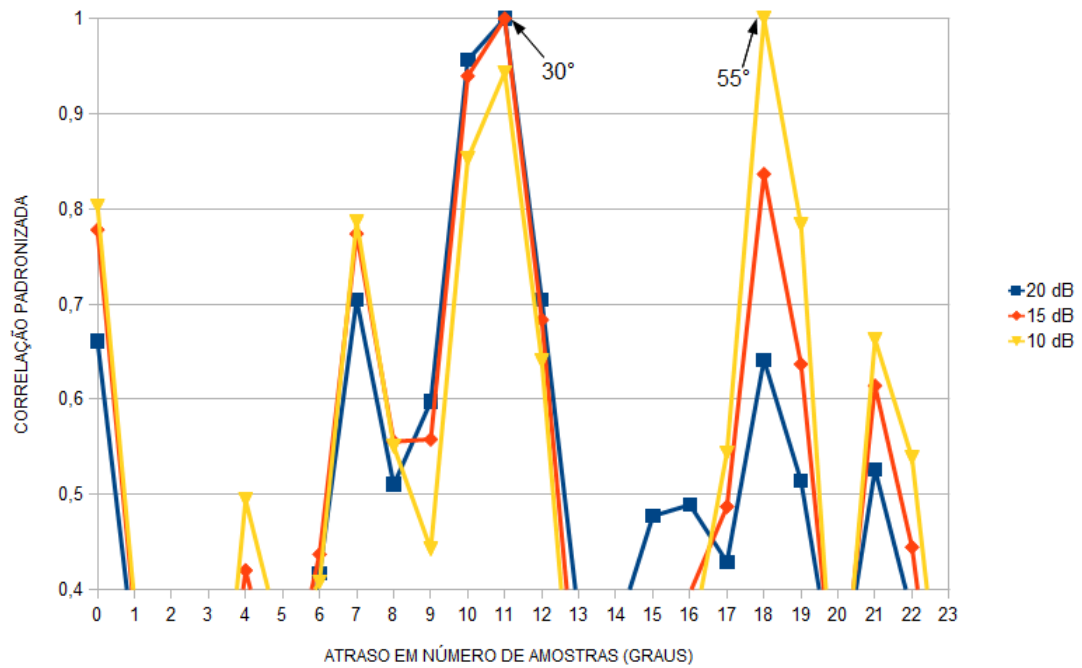


Figura 5.6: Algoritmo GCC-PHAT para diversas relações sinal-ruído.

Observa-se que a seletividade do algoritmo GCC-PHAT faz com que o mesmo possa incorrer em erros grandes, ao confundir os picos de ruído com o sinal. Com uma relação sinal-ruído de 10 dB o algoritmo GCC-PHAT indica erroneamente o atraso em 18 amostras (ou 55°), ao invés de 11 amostras (30°).

Como solução para este problema, optou-se por aumentar o tamanho do quadro, objetivando diluir o efeito do ruído branco no sinal, conforme observa-se nos gráficos 5.7 e 5.8.

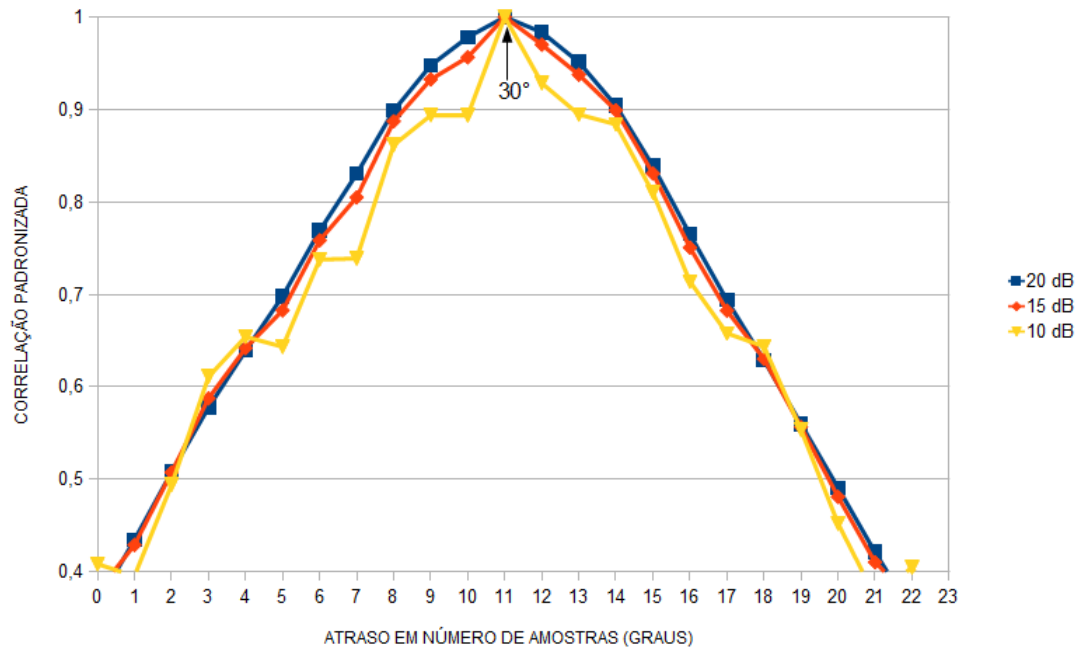


Figura 5.7: Algoritmo GCC para quadros de 512 amostras e diversas relações sinal-ruído.

E para o algoritmo GCC-PHAT:

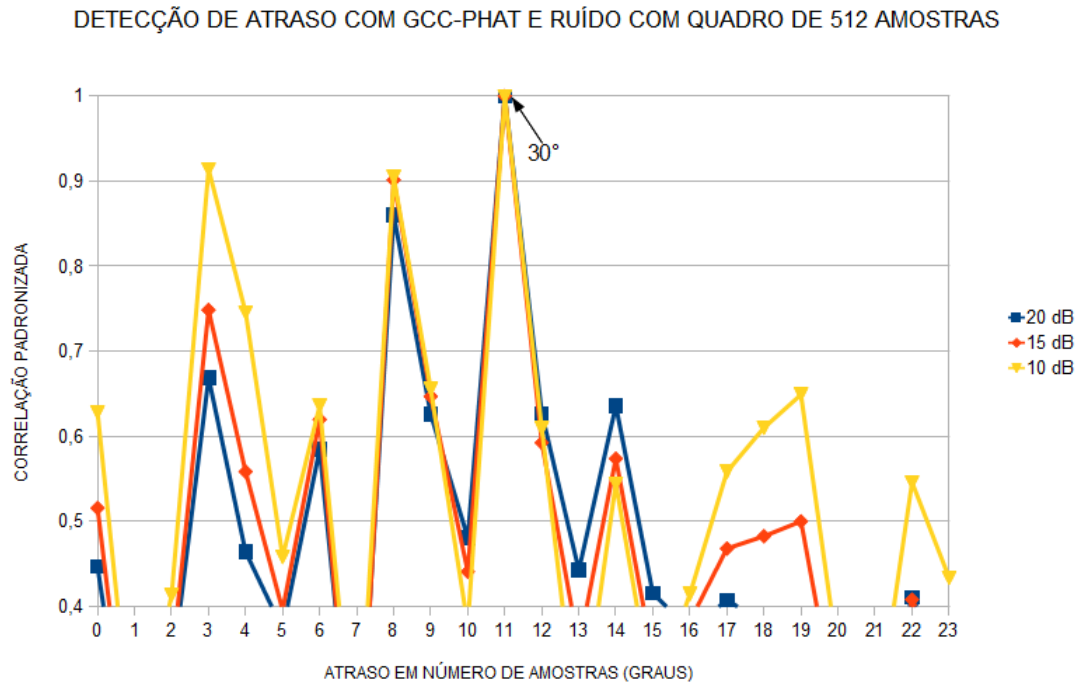


Figura 5.8: Algoritmo GCC-PHAT para quadros de 512 amostras e diversas relações sinal-ruído.

Observa-se que, com quadros de 512 amostras, ao invés das 256 amostras anteriores, ambos os algoritmos conseguiram determinar corretamente o atraso, mesmo com a pior relação sinal-ruído, de apenas 10 dB.

Finda a escolha dos parâmetros para os algoritmos GCC e GCC-PHAT, tem-se na sequência a comparação dos algoritmos GCC, GCC-PHAT e DUET perante sinais com reverberações.

5.2 Comparativo do GCC, GCC-PHAT e DUET perante reverberações

Pode-se observar a seguir gráficos contendo a distribuição dos valores de atraso detectados, juntamente com os valores de desvio padrão junto à legenda, para estes três algoritmos: GCC, GCC-PHAT e DUET. Os dois primeiros (gráficos 5.9 e 5.10) são dos sinais originais, e os dois últimos (gráficos 5.11 e 5.12) são dos mesmos sinais com emulação de reverberações. Para cada gráfico foram realizadas 91 simulações.

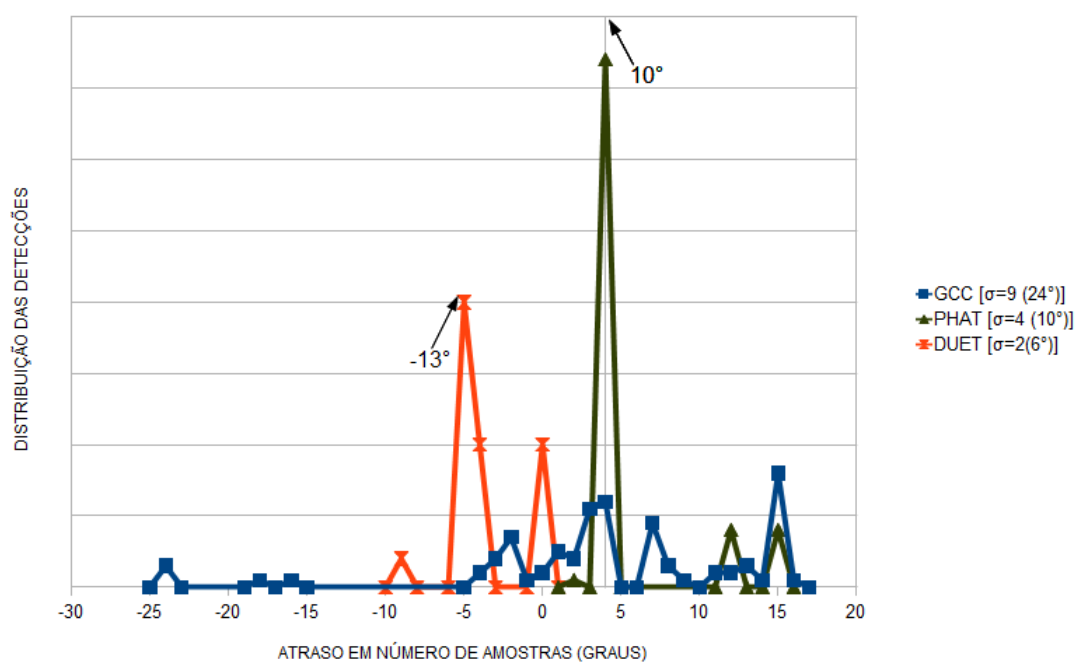


Figura 5.9: Histograma do atraso detectado para o 1º sinal.

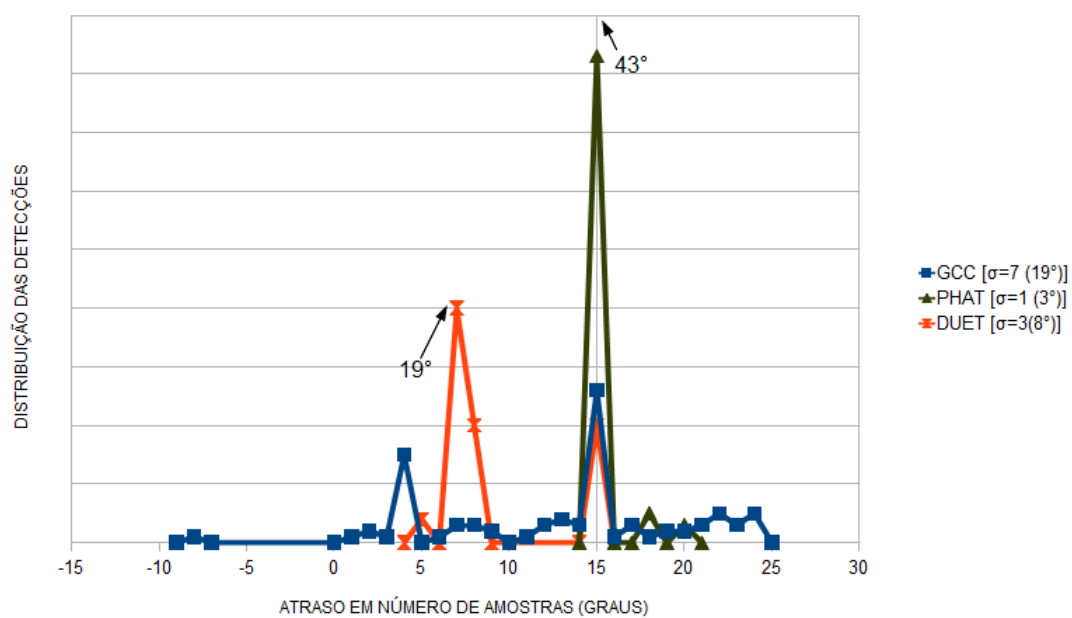


Figura 5.10: Histograma do atraso detectado para o 2º sinal.

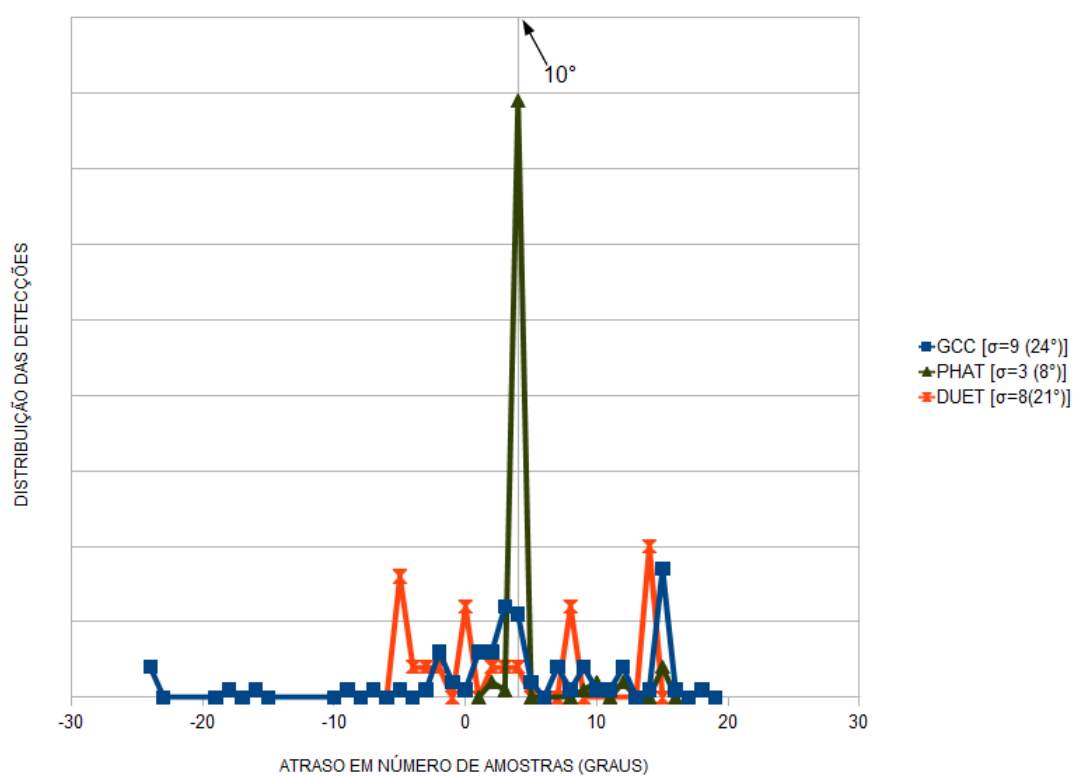


Figura 5.11: Histograma do atraso detectado para o 1º sinal com emulação.

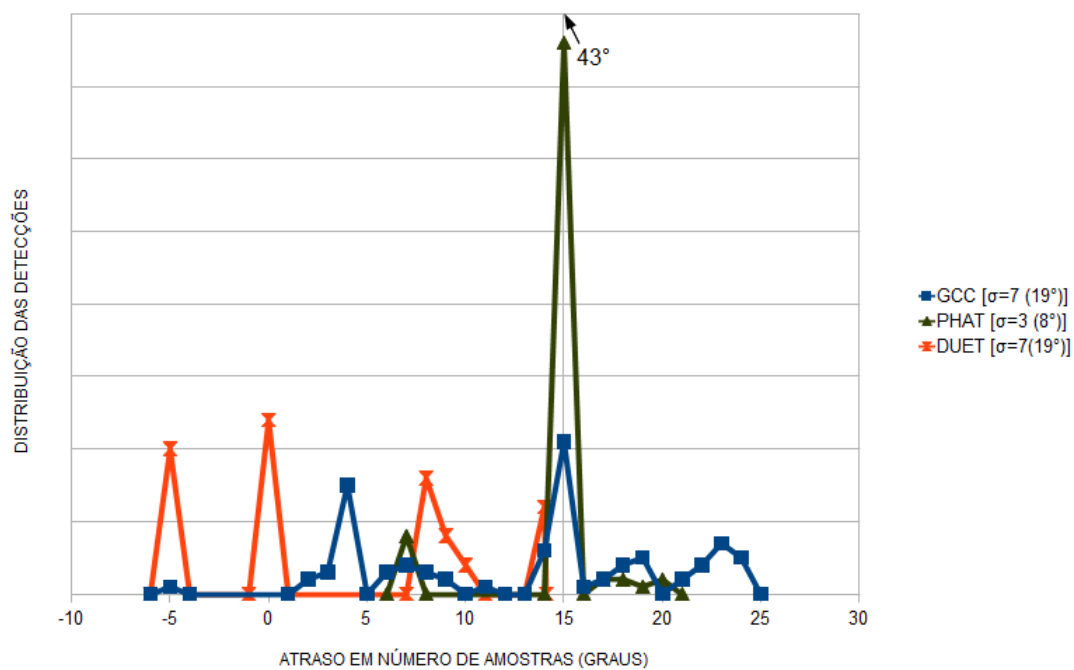


Figura 5.12: Histograma do atraso detectado para o 2º sinal com emulação.

Através dos gráficos percebe-se que os algoritmos GCC e, em menor escala, DUET, possuem um grande espalhamento dos atrasos detectados, sendo que as reverberações ampliam ainda mais esta tendência no caso do DUET. Já o algoritmo GCC-PHAT apresenta uma grande concentração em torno do valor de pico, com pequena variância. Os atrasos corretos encontram-se representados nos gráficos pelo eixo vertical.

O fenômeno da reverberação constitui-se no principal problema originado por ambientes fechados, pois gera diversos picos na função de correlação, devido ao somatório de inúmeras reflexões com o sinal original. Embora estes picos possam ser maiores do que aquele correspondente ao atraso real, com a metodologia utilizada até este ponto (a emulação de reverberações com fase mínima, utilizando *softwares* de estúdios de gravação) não foram detectados grandes efeitos negativos, sendo que os resultados dos três algoritmos apresentaram pouca degradação. Assim, para investigar mais apropriadamente este fenômeno, gravaram-se sinais reais em ambientes com intensa reverberação. Utilizaram-se equipamentos de gravação profissionais, com microfones omnidirecionais de referência (Behringer ECM8000) e gravador de sons (M-Audio Fast Track Pro) para a captação, digitalização e gravação de diversas amostras de sons. Os resultados podem ser vistos nos gráficos 5.13, para o qual foram utilizados 15 trechos; e 5.14, para o qual foram utilizados 21 trechos.

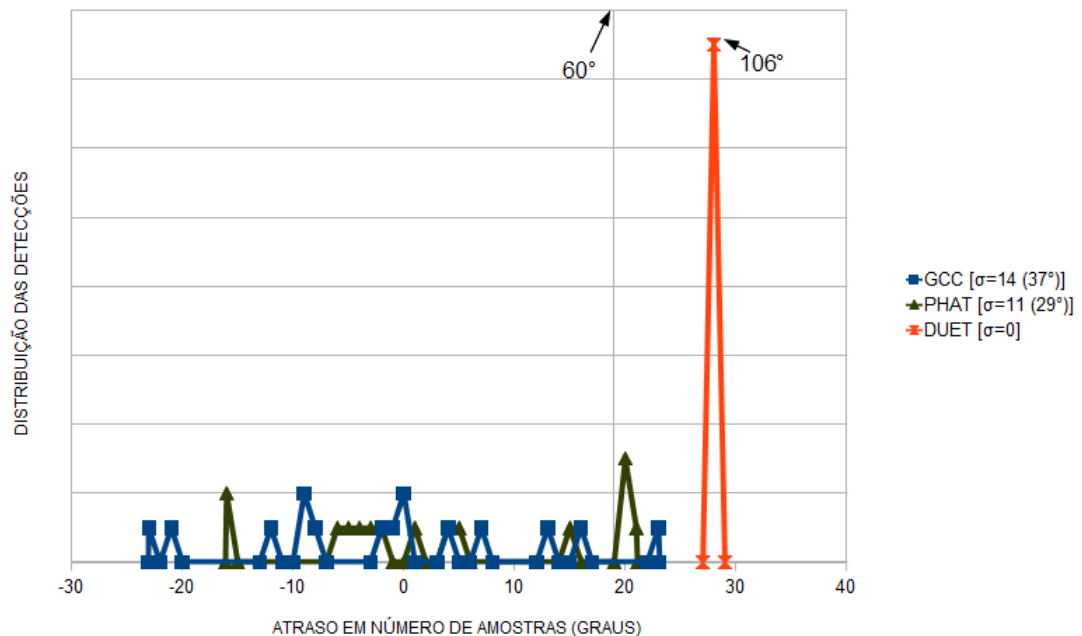


Figura 5.13: Histograma do atraso detectado para o 1º sinal gravado.

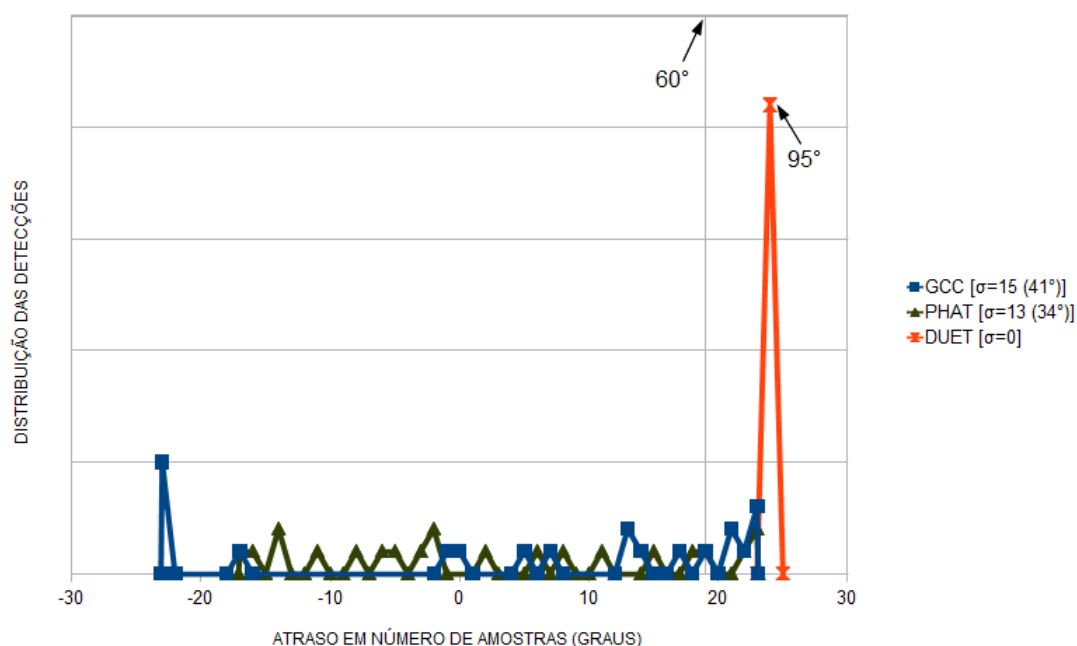


Figura 5.14: Histograma do atraso detectado para o 2º sinal gravado.

Observa-se nos gráficos que a dispersão aumenta muito para estes sinais devido à intensa reverberação, exceção feita ao algoritmo DUET, o qual apresenta, para o caso de uma única fonte sonora, somente um valor médio dos valores de atraso detectados, porém sem exatidão, haja vista a distância para o atraso correto (representado pelo eixo vertical nos gráficos), além do fato de ter estimado atrasos fisicamente irreais. Assim, para tentar obter resultados razoavelmente precisos para as estimativas de atraso, optou-se pela busca de novos métodos de detecção, nominalmente, o algoritmo Detector de Borda de Ataque e a Técnica do Espectrograma.

5.3 O Algoritmo Detector de Borda de Ataque

Este algoritmo permite que sejam selecionados parâmetros de configuração tais como o tamanho do quadro e o limiar de detecção (altura da borda), para evitar a detecção de picos de ruídos. Nas tabelas 5.2, 5.3 e 5.4 podem ser vistos os resultados obtidos em função da variação do parâmetro W (tamanho do quadro) e L (limiar de detecção) para três sinais gravados.

W \ L	.25	.3	.35	.4	.45	.5	.55	.6	.65	.7	.75	.8	.85	.9	.95
64	9	9	8	7	8	8	6	10	12	8	7	4	4	2	-3
72	5	5	5	7	4	6	5	11	5	6	11	3	1	0	-2
80	7	7	6	7	6	9	6	9	10	11	10	0	-1	-1	-3
88	9	8	6	5	6	6	5	10	9	5	2	-1	-1	-1	-3
96	5	5	2	3	6	6	5	9	10	7	-1	-1	-2	-2	-3
104	9	4	2	3	5	5	4	11	10	5	-2	-2	-3	-2	-3
112	6	3	3	4	5	5	7	12	6	0	-2	-2	-3	-1	-3
120	7	5	5	4	4	4	4	10	2	1	-1	-1	-2	-1	-1
128	8	5	4	5	5	4	8	12	2	2	0	-1	-2	-1	-1

Tabela 5.2: Erro relativo em número de amostras com o Detector de Borda para o 1º sinal gravado.

W \ L	.25	.3	.35	.4	.45	.5	.55	.6	.65	.7	.75	.8	.85	.9	.95
64	12	12	9	9	9	8	7	6	5	2	1	-1	-3	-4	-4
72	11	10	8	8	8	7	6	5	4	2	1	-2	-4	-4	-4
80	9	9	7	7	6	5	6	5	3	1	-1	-2	-3	-4	-4
88	7	7	6	5	4	5	7	4	3	0	0	-1	-3	-4	-4
96	8	6	4	4	4	4	6	5	3	1	0	-2	-3	-4	-4
104	7	5	4	4	4	4	7	5	4	1	1	-2	-3	-4	-3
112	7	4	5	5	4	4	6	4	3	1	0	-2	-3	-3	-3
120	6	5	6	7	5	4	7	5	3	1	-1	-2	-3	-3	-3
128	10	7	7	7	7	7	7	5	4	1	-2	-2	-1	-3	-3

Tabela 5.3: Erro relativo em número de amostras com o Detector de Borda para o 2º sinal gravado.

W \ L	.25	.3	.35	.4	.45	.5	.55	.6	.65	.7	.75	.8	.85	.9	.95
64	-9	-5	-4	-3	-2	1	1	2	2	2	2	1	2	2	2
72	-10	-5	-4	-3	-1	2	2	2	2	2	2	2	2	2	2
80	-11	-4	-7	-3	0	2	2	2	2	2	2	2	2	1	1
88	-15	-7	-5	0	1	2	2	2	2	2	2	2	2	1	1
96	-14	-9	-4	0	2	2	2	2	2	2	2	2	2	1	1
104	-11	-8	-3	1	1	2	2	2	2	2	2	2	2	0	1
112	-15	-8	-2	0	1	2	2	2	2	2	2	1	2	1	1
120	-9	-8	-4	-1	0	1	2	2	2	2	1	1	1	0	1
128	-15	-11	-2	-1	1	2	2	2	2	2	1	1	2	0	0

Tabela 5.4: Erro relativo em número de amostras com o Detector de Borda para o 3º sinal gravado.

Através das tabelas pode-se calcular o ponto ótimo para operação deste algoritmo, escolhido via erro quadrático mínimo, obtendo-se $W = 120$ e $L = 0,75$, o qual resulta em erros baixos de detecção (erro correspondente a 1 amostra para os três sinais gravados, conforme pode-se observar pelos números destacados nas tabelas anteriores). Como segunda opção, pode-se utilizar, com pequena desvantagem quanto ao desempenho, porém com grande economia computacional, o ponto $W = 96$ e $L = 0,75$. Considerou-se a possibilidade deste resultado tão favorável ser devido a uma "sintonia" do algoritmo com estes sinais. Para averiguar esta situação, foram realizados testes adicionais, com esse e com o próximo algoritmo.

5.4 A Técnica do Espectrograma

A Técnica do Espectrograma foi elaborada como outra solução alternativa para o problema trazido pelas reverberações. De forma semelhante ao algoritmo Detector de Borda de Ataque, ela também possui diversos parâmetros que podem ser escolhidos para obtenção do ponto ótimo de operação: o número de quadros comparados N , o número de frequências por quadro W e o limiar de detecção de sinal L . Como este algoritmo tem um custo computacional elevado comparativamente aos demais, devido ao grande número de quadros processados por estimativa, optou-se por fixar o valor do limiar de detecção dos sons em 0.95, objetivando diminuir o processamento. Comprovou-se, via simulações, que esta escolha não prejudica o funcionamento do mesmo. A seguir, têm-se as tabelas 5.5, 5.6 e 5.7, as quais contêm os erros relativos de estimação em número de amostras, com os valores de desvio padrão entre parênteses, para diversos valores dos parâmetros número de quadros N e de frequências W selecionados.

$N \setminus W$	32	48	64	80	96	112	128	256
500	-7 (9)	-7 (9)	-2 (11)	-2 (11)	-2 (11)	-3 (11)	1 (12)	1 (0)
600	-20 (7)	-7 (9)	1 (12)	1 (12)	1 (12)	1 (12)	1 (12)	1 (0)
750	-4 (17)	-7 (17)	1 (12)	1 (12)	1 (12)	1 (12)	1 (12)	1 (0)
1000	1 (3)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)

Tabela 5.5: Erro relativo e desvio padrão com a Técnica do Espectrograma para o 1º sinal gravado.

N \ W	32	48	64	80	96	112	128	256
500	-1 (17)	-1 (2)	-1 (2)	-1 (6)	-1 (6)	-3,5 (9)	-3,5 (8)	-1 (2)
750	-1 (17)	-1 (2)	-1 (2)	-1 (15)	-1 (0)	-1 (15)	-1 (15)	-1 (2)
1000	-1 (0)	-1 (0)	-1 (0)	-1 (0)	-1 (0)	-1 (2)	-1 (3)	-1 (2)

Tabela 5.6: Erro relativo e desvio padrão com a Técnica do Espectrograma para o 2º sinal gravado.

N \ W	32	48	64	80	96	112	128	256
500	20 (10)	8 (14)	10,5 (13)	11 (8)	1 (9)	11 (10)	1 (12)	2,5 (9)
750	20 (3)	8 (11)	10,5 (11)	1 (9)	1 (11)	1 (8)	1 (10)	0,5 (9)
1000	20 (2)	20 (8)	8,5 (11)	1 (10)	1 (10)	1 (9)	1 (7)	0,5 (9)

Tabela 5.7: Erro relativo e desvio padrão com a Técnica do Espectrograma para o 3º sinal gravado.

Pode-se observar nas tabelas o valor do erro médio em número de amostras, bem como o desvio padrão, para cada valor de parâmetro experimentado. Para os sinais gravados, o ponto de operação escolhido como o de melhor relação custo-benefício foi $N = 1000$ e $W = 80$, tendo sido os resultados correspondentes destacados nas tabelas anteriores. Definidos os parâmetros de funcionamento, resta agora submeter outros sinais a este algoritmo e ao Detector de Borda de Ataque, para verificar se o bom desempenho dos mesmos repete-se para qualquer sinal, ou se há algum tipo de conformação aos três sinais utilizados inicialmente.

5.5 Testes com outros sinais

Mantendo-se condições semelhantes às das gravações dos sinais anteriores, realizou-se a aquisição de mais um sinal de áudio. Deste novo sinal foram utilizados 562 trechos, e assim pôde-se comparar a eficácia de todos os algoritmos, conforme pode-se ver na tabela 5.8.

Parâm. \ Alg.	GCC	GCC-PHAT	DUET	Detec. Borda	Téc. Espec.
Erro relativo	19 (60°)	1 (3°)	8 (21°)	1 (3°)	1 (3°)
Desvio padrão	$\sigma = 14$ (39°)	$\sigma = 12$ (33°)	$\sigma = 0$	$\sigma = 4$ (10°)	$\sigma = 0$

Tabela 5.8: Erro médio e desvio padrão para o 4º sinal gravado.

Observa-se que o algoritmo GCC-PHAT obteve um resultado razoável com erro relativo médio de apenas 1 amostra, mas com uma dispersão de valores maior que a desejável. Já o algoritmo Detector de Borda obteve erro idêntico, porém com uma dispersão

bem menor, ao passo que a Técnica do Espectrograma apresentou o mesmo erro e sem dispersão alguma. Esses resultados comprovam a eficácia dos algoritmos propostos, os quais conseguem detectar muito bem a direção de chegada de sinais em ambientes com reverberações. Contudo, objetivando testar os limites do funcionamento dos mesmos, foram gravados sinais em condições extremas, quais sejam:

- Recinto pequeno e fechado;
- Paredes com revestimento cerâmico e piso de concreto, sem nenhum material fono-absorvente ou tratamento acústico;
- Ambiente com dimensões quase cúbicas, de 2,85 x 2,95 x 2,75 m (l x h x p);
- Fonte sonora próxima ao piso (refletor quase ideal).

Têm-se a seguir as tabelas 5.9 e 5.10, as quais trazem os erros relativos de detecção, bem como o desvio padrão, para esses sinais.

Parâm. \ Alg.	GCC	GCC-PHAT	DUET	Detec. Borda	Téc. Espec.
Erro relativo	43 (162°)	11 (30°)	15 (43°)	15 (43°)	2 (5°)
Desvio padrão	$\sigma = 13$ (36°)	$\sigma = 9$ (24°)	$\sigma = 0$	$\sigma = 3$ (8°)	$\sigma = 19$ (60°)

Tabela 5.9: Erro médio e desvio padrão para o 5° sinal gravado.

Parâm. \ Alg.	GCC	GCC-PHAT	DUET	Detec. Borda	Téc. Espec.
Erro relativo	36 (129°)	30 (111°)	17 (50°)	8 (21°)	12 (33°)
Desvio padrão	$\sigma = 17$ (50°)	$\sigma = 13$ (36°)	$\sigma = 0$	$\sigma = 15$ (43°)	$\sigma = 5$ (13°)

Tabela 5.10: Erro médio e desvio padrão para o 6° sinal gravado.

Para uma análise mais eficiente desse cenário extremo, deve-se levar em consideração que um aposento padrão possui cerca de três metros na sua menor dimensão, resultando em ângulos possíveis de detecção de -18° à -90° para sons vindos do solo, com o centro dos sensores a uma altura de 0,9 m, sendo razoável considerar que, na maior parte dos casos (média $\pm$ desvio padrão), os sons virão de ângulos entre -24° e -42° . Assim, para o 5° sinal gravado, do qual foram utilizados 17 trechos, apenas a Técnica do Espectrograma obteve um erro relativo médio que pode ser considerado bom, mesmo com um espalhamento grande. Os algoritmos GCC, GCC-PHAT, DUET e Detector de Borda apresentaram um erro de detecção relativamente grande para este sinal. Já para o 6° sinal

gravado, do qual foram utilizados 13 trechos, têm-se erros ainda toleráveis apenas para o algoritmo Detector de Borda, sendo o mesmo capaz de detectar um alarme (estimação da direção de origem do sinal como sendo do solo) para a maioria dos casos, evitando um erro por falta de atuação. Os demais demonstraram ser inadequados para este sinal com intensas reverberações.

5.6 Síntese

Neste capítulo foram escolhidos os parâmetros de funcionamento e testados os algoritmos conhecidos e os propostos, em função de sinais defasados, de sinais com emulação de reverberações e de sinais gravados em ambientes reverberativos, de forma a se poder comparar o desempenho de todos os algoritmos. Os resultados foram analisados, e as conclusões estão dispostas a seguir.

Capítulo 6

Conclusões

Este trabalho analisou o comportamento de algoritmos de estimação da direção de chegada de sons através dos atrasos entre canais distintos de um arranjo linear de microfones. Investigaram-se os algoritmos mais conhecidos (GCC, GCC-PHAT e DUET), e foram propostos dois novos algoritmos, o Detector de Borda de Ataque e a Técnica do Espectrograma, com a finalidade de obter um algoritmo cuja atuação ocorra de forma eficaz e confiável, reduzindo o número de alarmes falsos, enquanto parte de um sistema de televigilância médica, com o objetivo de detectar situações emergenciais em ambientes domésticos.

Através das análises, verificou-se que o algoritmo GCC-PHAT apresentou bons resultados para sinais com baixa reverberação. Para sinais apresentando níveis moderados de reverberação, obtiveram-se bons resultados com os dois algoritmos propostos, sendo o seu desempenho muito superior ao dos algoritmos mais utilizados. Ainda assim, em situações de reverberação muito intensa, a utilização desses algoritmos resultou em erros grandes, embora tenha-se verificado que, para a maior parte das situações, os erros de detecção não os impediriam de atingirem seu objetivo, qual seja, o de detectar sons vindos da direção do solo, para fins de televigilância médica. Uma possível solução para os casos de reverberação extrema, que não chegou a ser implantada neste trabalho, é tratar acusticamente os ambientes, com a instalação de cortinas, carpetes, forrações e outros materiais fonoabsorventes de baixo custo e fácil aplicação, de forma a torná-los minimamente compatíveis com a aplicação. Pode-se ainda utilizar médias móveis para os valores de ângulos detectados, ou mesmo utilizar dois arranjos de microfones em paredes opostas, aumentando a confiabilidade do sistema.

Por terem sido obtidos resultados promissores, recomenda-se a continuidade deste estudo em ambientes reais de televigilância médica, objetivando a aquisição de mais dados para estudo e a possível otimização dos algoritmos através da utilização de um arranjo com maior número de elementos sensores, da implantação de um algoritmo de rastreamento do sinal - *tracking* -, e da desejável separação dos sinais monitorados de fontes sonoras interferentes.

Referências Bibliográficas

- [ASCOM/MS 2011]”Melhor em casa leva atenção à casa do brasileiro”. Portal da Saúde, Ministério da Saúde, 2011. Disponível em: <http://portalsaude.saude.gov.br/portalsaude/noticia/3061/162/melhor-em-casa-leva-atencao-a-casa-do-brasileiro.html>. Acesso em: 10 dez. 2012.
- [Azaria e Hertz 1984]AZARIA, M.; HERTZ, D. ”Time delay estimation by generalized cross correlation methods”. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, ASSP-32, n. 2, p. 280 – 285, apr 1984.
- [BARBA 1998]BARBA, M. C. P. *La Fisica en la Medicina*. 2ª Ed.;Carretera Picacho-Ajusco 227; 14200 México, D.F.: FONDO DE CULTURA ECONÓMICA, S. A. DE C. V., 1998. ISBN 968-16-5701-2.
- [Brutti et al. 2008]BRUTTI, A.; OMOLOGO, M.; SVAIZER, P. ”Comparison between different sound source localization techniques based on a real data collection”. In: *Hands-Free Speech Communication and Microphone Arrays*. [S.l.: s.n.], 2008.
- [Champagne et al. 1996]CHAMPAGNE, B.; BEDARD, S.; STEPHENNE, A. ”Performance of time-delay estimation in the presence of room reverberation”. *Speech and Audio Processing, IEEE Transactions on*, v. 4, n. 2, p. 148 –152, mar 1996. ISSN 1063-6676.
- [Doukas e Maglogiannis 2011]DOUKAS, C. N.; MAGLOGIANNIS, I. ”Emergency fall incidents detection in assisted living environments utilizing motion, sound, and visual perceptual components”. *IEEE Transactions on Information Technology in Biomedicine*, v. 15, n. 2, p. 277–289, 2011.
- [Henderson 2012]HENDERSON, T. ”Sound waves and music”. The Physics Classroom, 2012. Disponível em: <http://www.physicsclassroom.com/Class/sound>. Acesso em: 10 dez. 2012
- [Jourjine et al. 2000]JOURJINE, A.; RICKARD, S.; YILMAZ, O. ”Blind separation of disjoint orthogonal signals: demixing n sources from 2 mixtures”. In: *Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proceedings. 2000*

- IEEE International Conference on*. [S.l.: s.n.], 2000. v. 5, p. 2985 –2988 vol.5. ISSN 1520-6149.
- [Knapp e Carter 1976]KNAPP, C.; CARTER, G. "The generalized correlation method for estimation of time delay". *Acoustics, Speech and Signal Processing, IEEE Transactions on*, v. 24, n. 4, p. 320 – 327, aug 1976. ISSN 0096-3518.
- [Montalvão Filho 1995]MONTALVÃO FILHO, J. R. *Avaliação da acustica de recintos pelo metodo dos elementos finitos*. Dissertação (Mestrado) — Universidade Estadual de Campinas, 1995.
- [Open_Speech_Repository 2012]OPEN_SPEECH_REPOSITORY. "The open speech repository", 2012. Disponível em: http://www.voiptroubleshooter.com/open_speech. Acesso em: 10 dez. 2012.
- [Ramirez 1985]RAMIREZ, R. W. *The FFT Fundamentals and Concepts*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1985. ISBN 0-13-314386-4.
- [Rickard e Dietrich 2000]RICKARD, S.; DIETRICH, F. Doa estimation of many w-disjoint orthogonal sources from two mixtures using duet. In: *Statistical Signal and Array Processing, 2000. Proceedings of the Tenth IEEE Workshop on*. [S.l.: s.n.], 2000. p. 311 –314.
- [Skaloumbakas 2010]SKALOUMBAKAS, C. "Reverberation time $rt(60)$ - definitions, theory, example cases etc.". Audio&Music, Friends&Family, 2010. Disponível em: <http://cpskal.gr/paper43.htm>. Acesso em: 10 dez. 2012.
- [Stout 2010]STOUT, H. "Technologies help adult children monitor aging parents". Home and Garden, The New York Times, 2010. Disponível em: <http://www.nytimes.com/2010/07/29/garden/29parents.html>. Acesso em: 10 dez. 2012.
- [Talantzis et al. 2005]TALANTZIS, F.; CONSTANTINIDES, A. G.; POLYMENAKOS, L. C. The generalized correlation method for estimation of time delay. *Communications and Signal Processing, IEEE Collections on*, v. 12, n. 8, p. 561 – 564, aug 2005. ISSN 1070-9908.
- [Vacher et al. 2006]VACHER, M.; SERIGNAT, J. F.; CHAILLOL, S.; ISTRATE, D.; POPESCU, V. Speech and sound use in a remote monitoring system for health care. In: SOJKA, P.; KOPECEK, I.; PALA, K. (Ed.). *Text, Speech and Dialogue*. [S.l.]: Springer Berlin Heidelberg, 2006, (Lecture Notes in Computer Science, v. 4188). p. 711–718. ISBN 978-3-540-39090-9.

[Varma 2002]VARMA, K. *Time-Delay-Estimate Based Direction-of-Arrival Estimation for Speech in Reverberant Environments*. Dissertação (Mestrado) — Virginia Polytechnic Institute and State University, 2002.