

UNIVERSIDADE REGIONAL DE BLUMENAU
CENTRO DE CIÊNCIAS EXATAS E NATURAIS
CURSO DE CIÊNCIA DA COMPUTAÇÃO – BACHARELADO

**UM ALGORITMO PARA DETECÇÃO E LOCALIZAÇÃO DE
FONTES SONORAS EM UM ESPAÇO 2D**

JEAN KIRCHNER

BLUMENAU
2011

2011/1-20

JEAN KIRCHNER

**UM ALGORITMO PARA DETECÇÃO E LOCALIZAÇÃO DE
FONTES SONORAS EM UM ESPAÇO 2D**

Trabalho de Conclusão de Curso submetido à
Universidade Regional de Blumenau para a
obtenção dos créditos na disciplina Trabalho
de Conclusão de Curso II do curso de Ciência
da Computação — Bacharelado.

Prof. Aurélio Faustino Hoppe, Mestre - Orientador

**BLUMENAU
2011**

2011/1-20

UM ALGORITMO PARA DETECÇÃO E LOCALIZAÇÃO DE FONTES SONORAS EM UM ESPAÇO 2D

Por

JEAN KIRCHNER

Trabalho aprovado para obtenção dos créditos
na disciplina de Trabalho de Conclusão de
Curso II, pela banca examinadora formada
por:

Presidente: _____
Prof. Aurélio Faustino Hoppe, Mestre – Orientador, FURB

Membro: _____
Prof. Dalton Solano Dos Reis, Mestre – FURB

Membro: _____
Prof. Antonio Carlos Tavares, Mestre – FURB

Blumenau, 30 de junho de 2011

A todos que de uma forma ou de outra, tentam
tornar o mundo um lugar melhor para se viver
como uma sociedade

AGRADECIMENTOS

Ao meu orientador Aurélio, por sempre estar disponível para qualquer problema, em praticamente qualquer horário ou dia.

Ao professor Evandro Felin Londero pela grande ajuda com matemática e sua empolgação com o trabalho.

Ao Everton e grupo de Radio e TV Furb por se disponibilizarem a ajudar com os recursos necessários neste trabalho.

À minha família, por entender minha falta em várias reuniões familiares e atenção que não pude oferecer.

Aos meus amigos, pelos empurrões, cobranças e festas que não pude comparecer.

À minha empresa Senior Sistemas por permitir algumas faltas necessárias para o desenvolvimento deste trabalho.

Ao professor Fernando Santos junto com meu orientador Aurélio por ajudar a manter a minha motivação na universidade.

À todos os outros professores e orientadores que de uma forma ou de outra contribuíram para o meu desenvolvimento intelectual e pessoal necessários para o desenvolvimento deste trabalho.

Faça o que ame e ame o que faça, esse é um bom caminho para o sucesso e a felicidade.

Filosofia própria

RESUMO

Este trabalho descreve a implementação de um algoritmo para localização de fontes sonoras em um espaço 2D. Para desenvolver este trabalho utilizou-se três microfones como receptores da fonte sonora. Aplicou-se a técnica de correlação cruzada para medir os atrasos entre a chegada do som nos três microfones. A partir disso, foi possível usar técnicas de triangulação para estimar a localização da fonte sonora. Neste trabalho, são apresentados detalhes das técnicas e modelagens matemáticas utilizadas para medir os atrasos entre os microfones e para efetuar a triangulação. Experimentos iniciais indicam que o algoritmo desenvolvido tem qualidade o suficiente para estimar a posição da fonte sonora, considerando um ambiente controlado sem ruídos, mas é imaturo para ser utilizado em um ambiente com a necessidade de resposta em tempo real. Contudo, são apresentadas algumas sugestões de técnicas que podem melhorar a precisão considerando ruídos e o tempo de resposta.

Palavras-chave: Processamento de som. Processamento de sinais. Localização sonora.

ABSTRACT

This paper describes an implementation of a sound source localization algorithm in 2D space. To develop this work, it was used three microphones as receptors to capture the sound of a sound source, that using the cross correlation technique to measure the time difference of arrival among the three microphones it is possible to use trilateration techniques to estimate the localization. It is presented details of the techniques used to measure the sound delays among microphones and to make the trilateration. The developed software have enough quality to estimate the position of the sound source, considering a controlled environment without noise, but it is immature to be used on a real environment with the need of a real time response. It is presented some suggestions of techniques that can be developed to improve the precision considering noise and response time.

Key-words: Sound processing. Signal processing. Sound localization.

LISTA DE ILUSTRAÇÕES

Figura 1 - Formato do sinal digital de áudio	17
Quadro 1 - Relações entre as propriedades do formato do sinal digital sonoro	18
Figura 2 - Dois sinais de entrada e sua correlação cruzada	20
Figura 3 - Representação do deslocamento dos vetores para o cálculo de correlação cruzada	21
Quadro 2 - Equação do <i>dot product</i>	21
Quadro 3 - Relação física MRU	22
Figura 4 - Representação de ângulos iguais opostos pelo vértice	24
Figura 5 - Representação da relação de ângulos de uma reta transversal cortando duas paralelas	24
Figura 6 - Representação das características de uma reta r , com uma abertura de ângulo a ..	26
Quadro 4 - Sistema linear de exemplo.....	26
Quadro 5 - Resolução do sistema linear representado no Quadro 4.....	27
Figura 7 - Representação de hipérbolas e suas relações	27
Figura 8 - Representação do ângulo a formado entre emissor e receptores	28
Figura 9 - Fórmula para encontrar a	28
Quadro 6 - Características dos trabalhos relacionados.....	30
Figura 10 - Diagrama de caso de uso	32
Figura 11 - Diagrama das classes desenvolvidas.....	33
Figura 12 - Modelo de três microfones exemplificando um cenário de captura, e o desenho de um emissor, representando o som chegando nos três microfones.....	35
Quadro 7 - Operações de correlação cruzada efetuadas	36
Quadro 8 - Código fonte representando implementação da correlação cruzada por <i>dot product</i>	37
Quadro 9 - Cálculos para encontrar atrasos em segundos	38
Figura 13 - representação dos parâmetros de distância	39
Quadro 10 - Parâmetros obtidos após conversão para domínio de espaço.....	39
Quadro 11 - Parâmetros de velocidade e tempo conhecidos	39
Quadro 12 - Alteração na fórmula MRU.....	39
Quadro 13 - Diferenças de distância encontradas	40
Figura 14 - ângulos entre perpendiculares dos lados e o emissor	40
Quadro 14 - Obtensão de ângulos.....	41

Quadro 15 - Ângulos suplementares encontrados	41
Figura 15 - Modelo matemático da primeira tentativa	41
Figura 16 - Objetivo final da técnica empregada	43
Figura 17 - características e relações utilizadas para encontrar os ângulos.....	44
Quadro 16 - Normalização de ângulos	45
Quadro 17 - Ângulos suplementares	45
Quadro 18 - Isolando variáveis na equação da reta	46
Figura 18 - Modelo pontos médios e microfones	46
Quadro 19 - Encontrando equações das retas	47
Quadro 20 - Sistema linear genérico entre duas equações de retas	47
Quadro 21 - Resolução de um sistema de equações genéricas de reta	47
Figura 19 - Representação das interseções encontradas e a fonte sonora	49
Quadro 22 - Teste de correlação cruzada em sinais hipotéticos.....	51
Quadro 23 - Teste de correlação cruzada em sinais hipotéticos.....	52
Figura 20 - Representação das interseções encontradas e a fonte sonora	52
Quadro 24 - Correlação cruzada dos sinais, em relação a posição sonora conhecida.....	53
Figura 21 - Fonte sonora estimada com erro tolerável	54
Figura 22 - Fonte sonora não detectada.....	55
Figura 23 - Fonte sonora estimada com erro não tolerável	56

LISTA DE SIGLAS

DSP – *Digital Signal Processing*

GCC-PHAT – *Generalized Cross Correlation – Phase Transform*

HRTF – *Head Related Transfer Functions*

IDE – *Integrated Development Environment*

ILD – *Interaural Level Difference*

ITD – *Interaural Time Difference*

MRU – *Movimento Retilíneo Uniforme*

UML – *Unified Modeling Language*

WAV – *Waveform Audio File Format*

SUMÁRIO

1 INTRODUÇÃO.....	14
1.1 OBJETIVOS DO TRABALHO	14
1.2 ESTRUTURA DO TRABALHO	14
2 FUNDAMENTAÇÃO TEÓRICA	16
2.1 PROCESSAMENTO DIGITAL DE SINAIS	16
2.1.1 Formato do sinal digital sonoro.....	17
2.1.2 Localização da fonte sonora.....	18
2.1.2.1 ILD.....	19
2.1.2.2 ITD.....	19
2.1.2.2.1 CORRELAÇÃO CRUZADA	20
2.2 RELAÇÕES FÍSICAS.....	22
2.2.1 MRU (Movimento Retilíneo Uniforme)	22
2.2.2 Velocidade do som.....	22
2.3 FUNDAMENTOS MATEMÁTICOS.....	23
2.3.1 Teorema de Tales	23
2.3.1.1 Ângulos opostos pelo vértice.....	23
2.3.1.2 Retas paralelas cortadas por uma transversal	24
2.3.2 Triângulo equilátero	25
2.3.3 Relações trigonométricas no triângulo retângulo.....	25
2.3.4 Reta	25
2.3.4.1 Coeficiente angular	25
2.3.5 Sistemas lineares	26
2.3.6 Hipérbole.....	27
2.4 TRABALHOS CORRELATOS.....	29
3 DESENVOLVIMENTO DO ALGORITMO	31
3.1 REQUISITOS DO TRABALHO PROPOSTO.....	31
3.2 ESPECIFICAÇÃO	31
3.2.1 Diagrama de casos de uso	32
3.2.2 Diagrama de classes	32
3.3 DESENVOLVIMENTO DO ALGORITMO PARA LOCALIZAÇÃO DA FONTE SONORA.....	35

3.3.1 Cálculo do ITD.....	35
3.3.1.1 Correlação cruzada	36
3.3.1.2 Cálculo da diferença no domínio de tempo	37
3.3.2 Triangulação para encontrar a fonte sonora	38
3.3.2.1 Primeiro passo, transformar parâmetros para domínio de espaço	38
3.3.2.2 Ângulos dos pontos médios	40
3.3.2.3 Tentativa 1, relações angulares e trigonometria básica	41
3.3.2.4 Tentativa 2, estimar posição baseado em interseção de retas	42
3.3.2.4.1 Encontrando os ângulos	43
3.3.2.4.2 Encontrando as equações lineares.....	45
3.3.2.4.3 Encontrando as interseções	47
3.3.2.4.4 Estimando a posição da fonte sonora.....	48
3.3.3 Otimizações.....	49
3.3.4 Técnicas e ferramentas utilizadas.....	50
3.3.5 Operacionalidade da implementação	51
3.4 RESULTADOS E DISCUSSÃO	51
3.4.1 Resultados da correlação cruzada	51
3.4.2 Resultados da triangulação.....	53
4 CONCLUSÕES.....	57
4.1 EXTENSÕES	58

1 INTRODUÇÃO

Desde os primeiros meses de vida, várias espécies de animais desenvolvem a habilidade de percepção sonora, que é de suma importância para a sobrevivência de muitas delas. A percepção sonora e consequentemente a localização das fontes ajudam estas espécies na sua sobrevivência e interação ou, colocando de uma forma mais genérica, ajudam estas espécies a se beneficiar e reagir melhor no ambiente no qual estão inclusas. Esta habilidade biológica foi-se desenvolvendo ao longo dos anos, que é decorrente de uma grande evolução natural das espécies, o que por si só é o suficiente para mostrar a grande importância desta habilidade para as espécies em seu meio.

Com isso em mente, a comunidade científica começou a perceber a utilidade de tal habilidade dentro de tecnologias que possam ser utilizadas em meios como: robôs, sistemas de segurança, vídeo conferências, meios de comunicação e várias outras tecnologias que envolvem interação humana.

Diante do exposto, neste trabalho propõe-se o desenvolvimento de um algoritmo capaz de indicar a localização de uma fonte sonora, utilizando um hardware composto de três microfones como dispositivo de entrada.

1.1 OBJETIVOS DO TRABALHO

O objetivo deste trabalho é implementar um algoritmo capaz de detectar a localização de fontes sonoras desconhecidas.

O objetivo específico do trabalho é definir modelo matemático para estimar direção, posicionamento e distância de diferentes fontes sonoras.

1.2 ESTRUTURA DO TRABALHO

Este trabalho está subdividido em capítulos que serão abordados a seguir.

O primeiro capítulo apresenta a justificativa para o desenvolvimento do objetivo do

trabalho e alguns dados acerca do tema escolhido.

O segundo capítulo trata de conceitos gerais de processamento de sinais e fundamentos matemáticos, esclarecendo tópicos importantes que devem ser considerados para que seja alcançado êxito nos objetivos propostos.

O terceiro capítulo exhibe o modelo matemático desenvolvido, a metodologia adotada durante o desenvolvimento do trabalho, assim como os resultados obtidos pelo sistema.

As conclusões são expostas no quarto capítulo, onde sugestões para trabalho futuros também podem ser encontradas.

2 FUNDAMENTAÇÃO TEÓRICA

Nas próximas seções são detalhados conceitos, técnicas e métodos utilizados na detecção e localização de fontes sonoras. Na seção 2.1 são abordados conceitos e técnicas de processamento de sinais, as relações físicas serão tratadas na seção 2.2. A fundamentação matemática utilizada na elaboração deste trabalho será tratada na seção 2.3. Por fim, a seção 2.4 apresenta alguns trabalhos relacionados e seus resultados.

2.1 PROCESSAMENTO DIGITAL DE SINAIS

De acordo com Punsakaya(2009), processamento digital de sinais (DSP – *Digital Signal Processing*) se preocupa com a representação dos sinais de tempo discretos por uma sequência de números ou símbolos e com o processamento destes sinais. DSP inclui subáreas tais como:

- a) áudio e reconhecimento de sinais de voz;
- b) processamento de sinais sonares e de radares;
- c) processamento de sinais de sensores;
- d) estimativas espectrais;
- e) processamento de sinal estatístico;
- f) processamento de imagens digitais;
- g) processamento de sinais para comunicação;
- h) controle de sistemas;
- i) processamento de sinais biomédicos (i.e. eletro cardiograma);
- j) processamento de dados sísmicos.

O objetivo de DSP é normalmente medir, filtrar e/ou comprimir sinais analógicos contínuos do mundo real.

O primeiro passo normalmente é converter um sinal de sua forma analógica para uma forma digital, por um processo chamado *sampling*¹. Os passos seguintes é trabalhar com o

¹ Processo pelo qual se utiliza um conversor analógico – digital, que torna o sinal analógico em uma sequência de números que o representa, sendo esta sequência de números o sinal digital.

signal já convertido, que dependem da aplicação desejada, que no contexto deste trabalho, é trabalhado com o sinal para encontrar atrasos entre os microfones.

2.1.1 Formato do sinal digital sonoro

Esta seção introduz conceitos de processamento digital de sinais sonoros, o foco não será os formatos de arquivos de áudio existentes, mas sim no formato geral de sinais de áudio para processamento após a leitura de arquivos de áudio, ou seja, trabalhar com o sinal de áudio em memória para processamento.

Considere o sinal da Figura 1.

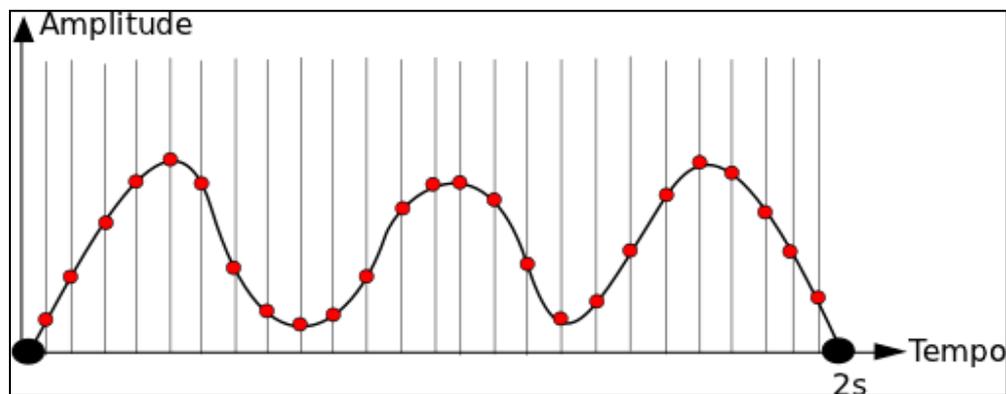


Figura 1 - Formato do sinal digital de áudio

O sinal da Figura 1 é um sinal hipotético que representa o formato do sinal digital de áudio. O formato de sinal digital tem várias propriedades, algumas é em relação ao formato de arquivo, para trabalhar com o sinal digital, a seguinte estrutura de formato é utilizada:

- sample*: é o valor da amplitude do sinal em um instante do tempo. Cada divisão do sinal representado na Figura 1 é um *sample*, ou seja, um *sample* é uma parte do sinal, o valor destes samples são as bolinhas vermelhas representadas na figura, que nada mais são como um ponto no plano cartesiano (x, y) no qual se utiliza (tempo, amplitude);
- frame rate*: a Figura 1 representa um sinal sonoro em um intervalo de tempo de dois segundos, no qual o *frame rate* são quantos samples por segundo um sinal sonoro tem;
- sample count*: são quantos samples um sinal tem, a figura acima tem o total de 26 samples como pode ser observado pelas bolinhas em vermelho;
- tempo do sinal: é tempo de captura de um sinal sonoro emitido, no qual a Figura 1 acima tem dois segundos de tempo.

Todas as propriedades acima citadas estão fortemente relacionadas, são diretamente proporcionais. Assim podemos mostrar as relações representadas no Quadro 1.

<p>a) $\text{framerate} = \text{samplecount} / \text{tempo}$ b) $\text{samplecount} = \text{tempo} * \text{framerate}$ c) $\text{tempo} = \text{samplecount} / \text{framerate}$</p>

Quadro 1 - Relações entre as propriedades do formato do sinal digital sonoro

Para a relação a representado no Quadro 1, conseguimos obter o *framerate* do sinal. Se um sinal tem 2 segundos e uma quantidade de *frames* de 10, o *frame rate* deste sinal é 10 *frames* / 2 segundos, ou seja, 5 *frames* por segundo. Já na relação b representada no Quadro 1, é possível obter a quantidade de *samples*. Um sinal com 2 segundos e um *frame rate* de 5 *frames* por segundo tem o tempo como $2 * 5$ o que resulta em 10 *frames*; A relação c representada no Quadro 1 mostra como obter o tempo do sinal, um sinal com *frame rate* de 5 *frames* por segundo e um *sample count* de 10 *frames*, resulta no tempo $10 / 5$ que é 2 segundos.

A estrutura descrita acima representa o formato do sinal digital e suas relações, ao qual serão necessárias no momento de aplicar as técnicas para medição de atraso. Tais técnicas serão abordadas nas próximas seções.

2.1.2 Localização da fonte sonora

A habilidade de conseguir encontrar a direção de uma fonte sonora está relacionada com a sobrevivência e é executada de maneira inconsciente. Esta tarefa já é executada desde muito cedo, de acordo com Bresloff et al.(2004, p. 67) os seres humanos, começam a localizar o som em média desde os seis meses de idade.

Na grande maioria das vezes para descobrir a localização da fonte sonora, o cérebro discrimina diferenças muito pequenas. Estas diferenças são em relação à intensidade, frequência e tempo de chegada de um evento auditivo que ocorre simultaneamente em nossos ouvidos. Como os ouvidos são separados por uma distância, é de fato muito simples observar que as ondas sonoras que atingem uma delas possuem características físicas diferentes das que atinge a outra. Assim existem duas principais diferenças consideradas para cada onda sonora que atinge cada ouvido: a diferença de fase, ou seja, a diferença do tempo de chegada da onda com relação aos ouvidos, chamado de ITD (*Interaural Time Difference*); e a diferença do

“volume”, ou seja, da intensidade do som, chamada de ILD (*Interaural Level Difference*) (FORNARI; MAIA JUNIOR; MANZOLLI, 2006, p. 275-276).

2.1.2.1 ILD

A diferença das intensidades (ou de volume) dos sons que atingem o ouvido esquerdo e o ouvido direito é conhecida como ILD, no qual é expressa em decibéis. O ILD é uma função que cobre uma grande parte do espectro audível (20 – 20.000Hz), isto porque ondas sonoras são difratadas quando seu comprimento de onda é maior que o tamanho da cabeça. Na frequência de 500Hz, por exemplo, o tamanho de onda é aproximadamente 69 cm, quatro vezes maior que o diâmetro médio da cabeça. O ILD é, no entanto, pequeno para frequências abaixo de 500Hz para fontes que estejam mais longe que um metro. Contudo a dispersão pela cabeça aumenta rapidamente com frequências crescentes, já em 4000Hz a cabeça gera uma diferença significativa na intensidade do som quando estiver no caminho entre a fonte sonora e um dos ouvidos (HARTMANN, 1999).

2.1.2.2 ITD

O ITD é a diferença de fase das ondas sonoras que chegam aos ouvidos, definida como a diferença do intervalo de tempo entre as ondas para chegar a um ouvido e depois na outra (SHINN-CUNNINGHAM; KOPCO; MARTIN, 2005). Essa diferença esta em função do tamanho da cabeça (distância de um ouvido à outro), da velocidade do som e do ângulo do azimute². Baseado nestes parâmetros e na largura da cabeça de um ser humano, que possui em média 8.75cm, fisicamente só é distinguível a direção de um evento auditivo pelo ITD quando a diferença de fase da onda não ultrapassa 0.66ms (WANG, 1990, p. 137). O ITD geralmente é mais útil quando a frequência do som não ultrapassa 1500Hz. Existem também estudos (BLAUERT, 1996, p. 183) que explicam a dificuldade de localizar a fonte sonora entre o intervalo de 1500Hz e 3000Hz, porque nesta faixa o comprimento de onda fica próximo ao diâmetro da cabeça, fazendo com que as diferenças de intensidade e do tempo de chegada do

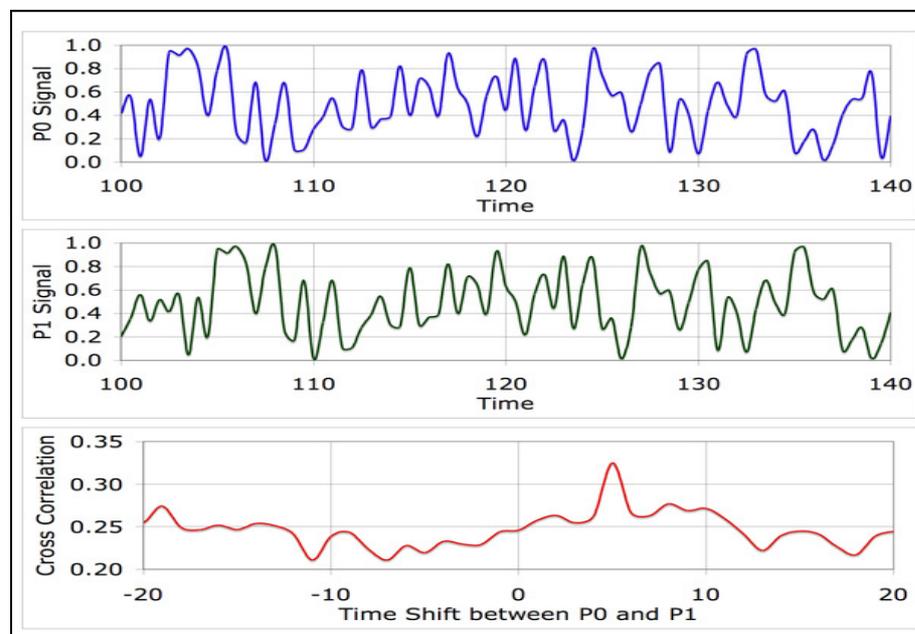
2 Ângulo formado entre a fonte sonora e uma linha horizontal imaginária que corta perpendicularmente o ponto médio entre os dois ouvidos – plano sagital (plano que passa através do eixo mais longo do corpo, o dividindo em lado direito e esquerdo)

evento auditivo reconhecido por ambas os ouvidos sejam insignificantes (OLIVEIRA et al., 2008, p. 8). Uma das principais técnicas para encontrar o ITD, é utilizar a correlação cruzada.

2.1.2.2.1 CORRELAÇÃO CRUZADA

De acordo com SHIESSER (2009, p. 1) correlação cruzada em processamento de sinais é a medição de similaridade entre dois sinais, normalmente utilizada para achar características de um sinal desconhecido comparando com um conhecido. É uma função do tempo relativo entre dois sinais.

Um exemplo de correlação cruzada pode ser visto na Figura 2.



Fonte: Siddhant (2011).

Figura 2 - Dois sinais de entrada e sua correlação cruzada

A correlação cruzada entre os dois sinais (P0 e P1 *signal* da Figura 2) tem o tamanho maior que os sinais, mas para fins de esclarecimentos a correlação da Figura 2 mostra apenas a faixa relevante. Isso se deve porque o sinal menor “atravessa” o sinal maior gerando um coeficiente para cada instante desta passagem. Observando a correlação cruzada gerada na Figura 2 (terceiro sinal representado em vermelho), verifica-se que o instante que tem o maior pico (parte do sinal mais parte do topo do gráfico) da correlação cruzada é o instante em que os sinais têm maior similaridade.

Para melhor entendimento desta técnica considere a Figura 3.

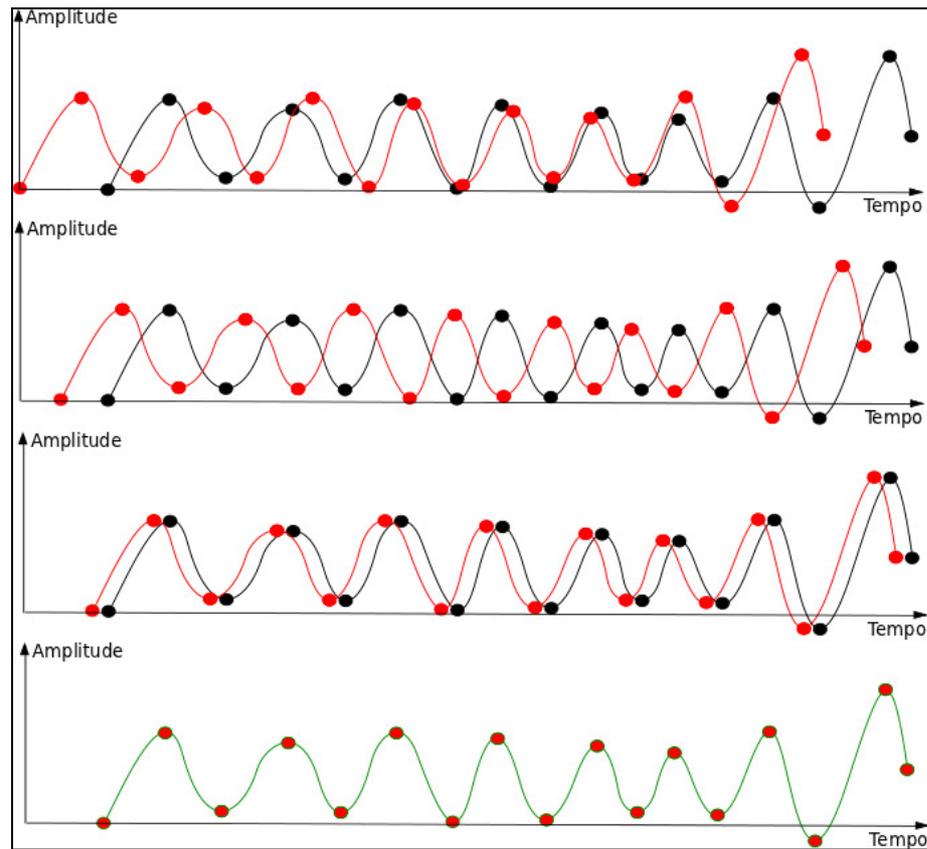


Figura 3 - Representação do deslocamento dos vetores para o cálculo de correlação cruzada

As quatro faixas representadas na Figura 3 são os mesmos sinais sendo comparados via correlação cruzada, mas em passos distintos da correlação e na ordem de comparação.

O sinal representado em vermelho varia pelo tempo passando pelo sinal representado em preto, como pode se observar na Figura 3, para cada instante desta passagem é calculado o produto dos sinais para saber o quão similar as ondas são naquele instante.

A maior correlação entre os sinais representados em preto e vermelho acontecem no instante representado pela última faixa da Figura 3, no qual os dois sinais passam a ser representados por um único sinal, mas com a cor verde, para indicar que são iguais naquele instante.

Este cálculo do produto é conhecido como *dot product* que é representado pela fórmula no Quadro 2.

$$X.Y = \sum_i X_i * Y_i$$

Quadro 2 - Equação do *dot product*.

Através do cálculo do *dot product* consegue-se uma razão de similaridade para cada atraso, desta forma compara-se a similaridade de cada atraso. O atraso que possuir maior *dot product* é o atraso de chegada da onda sonora nos microfones. Desta forma conseguiu-se saber a quantidade de *frames* que um sinal difere de outro.

2.2 RELAÇÕES FÍSICAS

Nas próximas seções serão abordadas algumas relações da física utilizadas no desenvolvimento deste trabalho.

2.2.1 MRU (Movimento Retilíneo Uniforme)

Na física existe uma relação direta entre velocidade, distância e tempo chamada MRU, no qual é representada no Quadro 3.

$$Velocidade\ média = \frac{Distância}{Tempo}$$

Quadro 3 - Relação física MRU

Esta relação implica que se qualquer um destas grandezas relacionadas alterarem, ao menos uma das outras duas deve-se alterar para satisfazer essa equação.

Analogamente podemos dizer que, se um objeto se movimentar mais rápido, vai precisar de menos tempo para trafegar em uma distância, ou irá trafegar por uma distância maior durante o mesmo tempo.

2.2.2 Velocidade do som

Como visto na seção anterior (seção 2.2.1), para converter os parâmetros de tempo para parâmetros de distância utiliza-se a relação física MRU.

No entanto, para encontrar a distância, temos o parâmetro de tempo e precisamos encontrar o parâmetro de velocidade.

A velocidade do som pode variar dependendo de alguns fatores ambientais como: temperatura, material, pressão. No caso da velocidade do som no ar, pode variar ainda com a umidade. No entanto para calcular a velocidade do som, são necessárias várias medições destes parâmetros, além de conhecer as propriedades do material que serve como meio de transmissão sonoro.

Como o objetivo deste trabalho é encontrar a localização da fonte sonora com

aproximação, tentando imitar a capacidade humana, não será utilizado o cálculo da velocidade do som, pois aumenta a complexidade e o hardware necessário, portanto é utilizado um valor médio de ambiente previamente calculado, conforme descrito a seguir.

A constante utilizada é um cálculo baseado no ar seco (sem umidade) em uma temperatura de 20 graus Celsius (ou 68 Fahrenheit), então a velocidade do som é de 343.2 m/s ou 1236 km/h, que é uma constante baseado em fatores de ambiente médios.

2.3 FUNDAMENTOS MATEMÁTICOS

Nesta seção aborda-se a fundamentação matemática para entendimento da solução proposta. Os assuntos abordados serão: teorema de Tales (seção 2.3.1), propriedades do triângulo equilátero (seção 2.3.2), relações trigonométricas no triângulo retângulo (seção 2.3.3), linha equações e relações (seção 2.3.4), sistemas lineares (2.3.5) e, por fim, relações hiperbólicas (seção 2.3.6).

2.3.1 Teorema de Tales

O teorema de Tales envolve várias relações e proporcionalidades, tais como: ângulos opostos pelo vértice e retas paralelas cortadas por uma transversal. Estas relações são abordadas nas próximas seções.

2.3.1.1 Ângulos opostos pelo vértice

O teorema de Tales diz que duas retas concorrentes, ou seja, se cruzando, independente da abertura do ângulo formado por esta reta, os ângulos opostos pelo vértice serão sempre iguais (WEISSTEIN, 2010). A Figura 4 ilustra melhor esta relação.

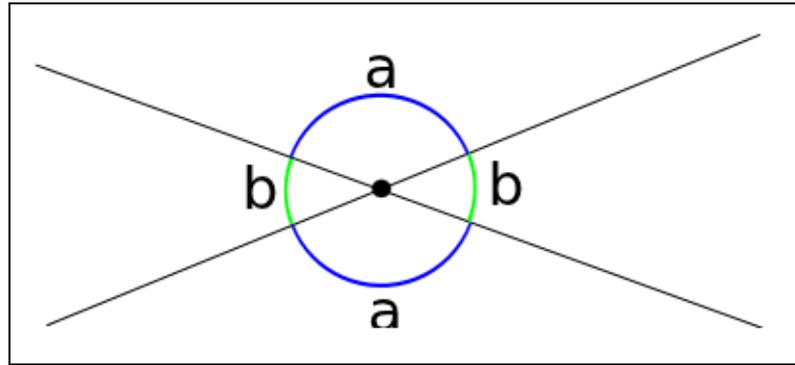


Figura 4 - Representação de ângulos iguais opostos pelo vértice

Os ângulos representados por a (em azul) tem o mesmo valor, pois são opostos pelo vértice, assim como os ângulos formados por b (em verde).

2.3.1.2 Retas paralelas cortadas por uma transversal

Quando as retas paralelas são cortadas por uma reta transversal, os ângulos formados numa destas retas paralelas são correspondentes e iguais aos ângulos da outra (OZAN, 2005) (Figura 5).

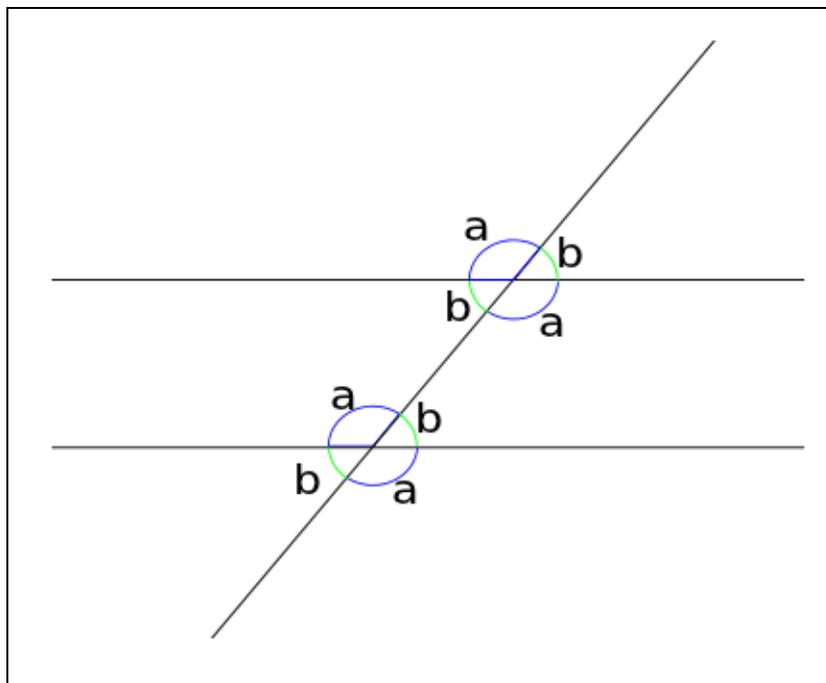


Figura 5 - Representação da relação de ângulos de uma reta transversal cortando duas paralelas

Conforme pode ser visto na Figura 5, os ângulos são correspondentes nos dois cortes, inclusive os ângulos opostos pelo vértice abordados na seção anterior. Os ângulos representados por a (em azul) são iguais entre si, assim como os representados por b (em verde).

2.3.2 Triângulo equilátero

Um triângulo equilátero, é um triângulo cujos lados possuem a mesma medida, e conseqüentemente seus ângulos internos. Como a soma dos ângulos internos de qualquer triângulo tem que ser 180 graus, o triângulo equilátero tem três ângulos iguais de 60 graus (WEISSTEIN, 2010).

2.3.3 Relações trigonométricas no triângulo retângulo

Um triângulo retângulo é um triângulo no qual tem um dos ângulos com a medida 90 graus. Os lados a , b e c deste triângulo satisfazem o teorema de Pitágoras: $a^2 + b^2 = c^2$, onde o maior lado é convencionalmente denotado como c e é chamado de hipotenusa (WEISSTEIN, 2010).

2.3.4 Reta

Uma reta é uma linha unidimensional sem nenhuma espessura e se estendendo infinitamente nas duas direções.

Uma reta é unicamente definida por dois pontos, com a linha passando pelos dois pontos vindo do infinito e indo para o infinito na direção oposta.

Considere apenas linhas em um plano, duas linhas que não se interceptam são consideradas linhas paralelas.

Toda reta pode ser definida na forma: $ax + by + c = 0$.

2.3.4.1 Coeficiente angular

O coeficiente angular de uma reta não perpendicular aos eixos das abscissas (eixo x) é o número real que expressa a tangente trigonométrica de sua inclinação (WEISSTEIN, 2010). Observe a Figura 6.

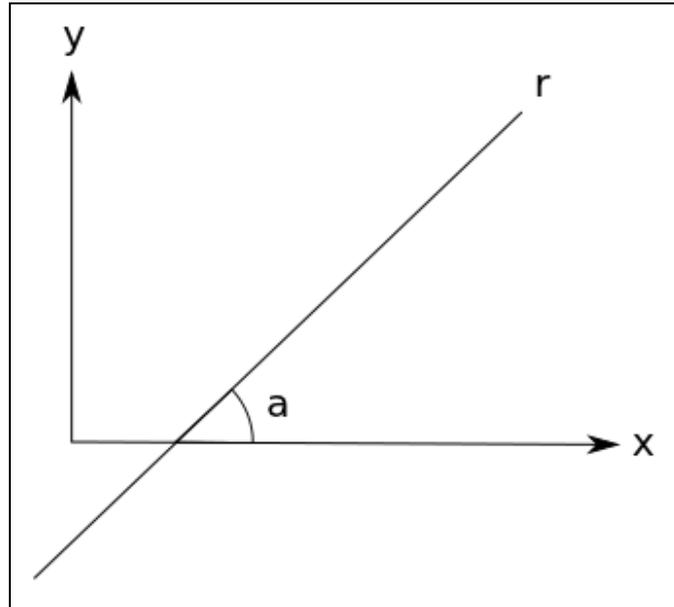


Figura 6 - Representação das características de uma reta r , com uma abertura de ângulo a

A equação reduzida da reta utilizando o coeficiente angular pode ser representada como: $y - y_1 = \text{ang} * (x - x_1)$, onde x e y representam os valores de x e y dos pontos pertencentes a esta reta, x_1 e y_1 representam um ponto qualquer pertencente a esta reta, e ang representa o coeficiente angular da reta, no qual é a tangente do ângulo a representado na Figura 6. Desta forma, com um ponto qualquer e a inclinação obtêm-se a fórmula da reta.

2.3.5 Sistemas lineares

Um sistema linear de equações é um conjunto de equações lineares com variáveis desconhecidas. Sistemas lineares servem para achar soluções para estas variáveis, dependendo do sistema, as variáveis podem ou não ter solução (WEISSTEIN, 2010).

Um exemplo de sistema linear é apresentado no Quadro 4.

$$\begin{array}{l} y = x - 10 \\ y = 2x + 9 \end{array}$$

Quadro 4 - Sistema linear de exemplo

O sistema representado no Quadro 4 é um sistema de equações dois por dois, pois existem duas equações sendo relacionadas e duas variáveis desconhecidas.

Para a solução do sistema linear dois por dois, será utilizado o método de substituição. Este método consiste em isolar uma das variáveis em uma das equações e substituir na outra equação a mesma variável. O Quadro 5 mostra a solução deste sistema, como no Quadro 4 o y já estava isolado, basta substituir em uma das equações o y da outra equação do sistema.

$$\begin{aligned}
 x - 10 &= 2x + 9 \\
 x &= 2x + 9 + 10 \\
 x - 2x &= 9 + 10 \\
 -x &= 19 \\
 x &= -19
 \end{aligned}$$

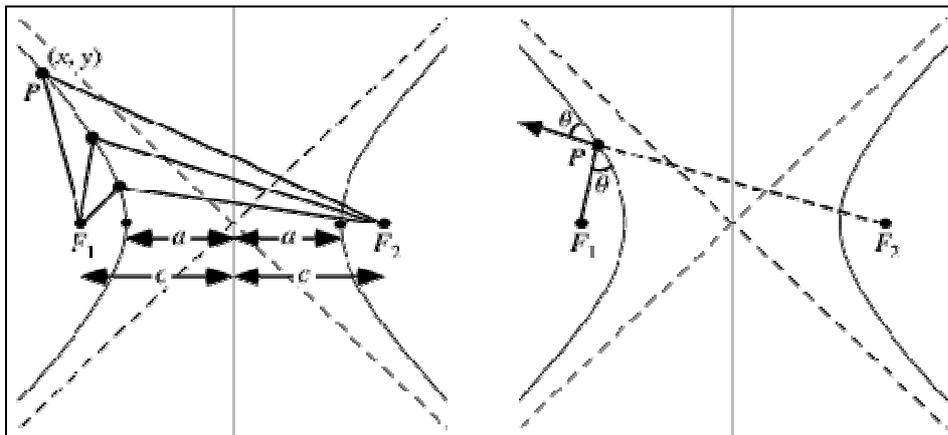
$$\begin{aligned}
 y &= x - 10 \\
 y &= -19 - 10 \\
 y &= -19 - 10 \\
 y &= -29
 \end{aligned}$$

Quadro 5 - Resolução do sistema linear representado no Quadro 4

A primeira linha do Quadro 5 é a equação para encontrar x no qual a solução é desenvolvida até encontrar $x = -19$. Uma vez encontrado o valor de x no sistema linear, substitui-se o valor de x em uma das equações do sistema linear do Quadro 4, para assim obter o valor de y conforme solução apresentada no Quadro 5.

2.3.6 Hipérbole

Uma hipérbole é uma seção cônica definida por uma faixa de pontos P no plano cuja diferença das distâncias entre F_1 , P e F_2 , P , é igual a $2a$, onde F_1 e F_2 são dois pontos fixos, chamados de foco, separados por uma distância $2c$ (WEISSTEIN, 2010). Dado o exemplo representado pela Figura 7.



Fonte: WEISSTEIN (2010).

Figura 7 - Representação de hipérboles e suas relações

Na Figura 8, está representado o ângulo a .

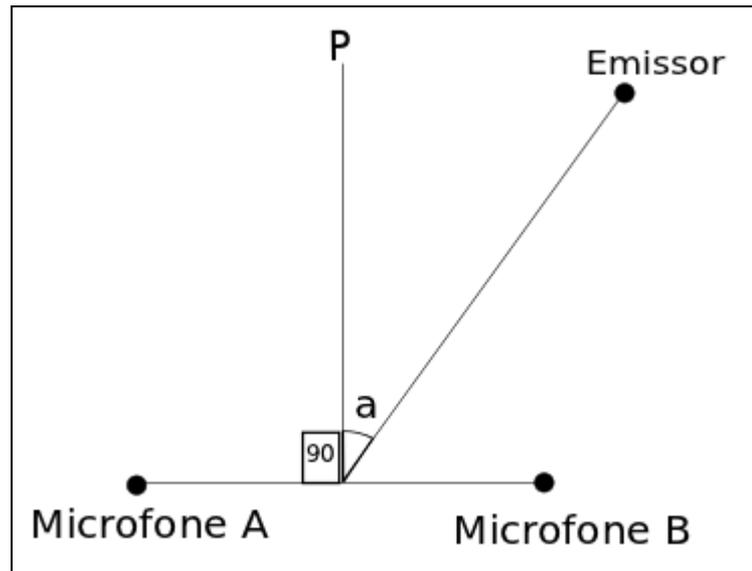


Figura 8 - Representação do ângulo a formado entre emissor e receptores

Em analogia a representação hiperbólica da Figura 7, os Microfones A e B serão os focos F_1 e F_2 , e o *Emissor* o ponto P .

Dadas as relações hiperbólicas, é possível deduzir a fórmula da Figura 9 conforme Goodridge (1997).

$$\theta \cong \sin^{-1} \left(\frac{D_I - D_R}{D_M} \right)$$

Fonte: Goodridge (1997).

Figura 9 - Fórmula para encontrar a

Na Figura 9, o \sin^{-1} representado, é o seno inverso também conhecido como *arcsin*, D_I e D_R são as dimensões dos lados do triângulo, que representam a distância entre P , F_1 e P , F_2 na Figura 7, no qual esta diferença representa o mesmo que $2a$, mas na fórmula é representado como a diferença entre as distâncias para o melhor entendimento da seção 3 onde é abordado o desenvolvimento deste trabalho. Conforme Goodridge (1997) esta relação representada na Figura 9, não tem precisão nenhuma para casos onde D_I e D_R representados na fórmula são maiores que D_M , o que para este trabalho seriam os pontos muito próximos do modelo de captura.

Verifica-se que o valor Theta representado na Figura 9, é ambíguo, pois como verificado na definição da hipérbole, há mais de um ponto que define a diferença $D_I - D_R$.

2.4 TRABALHOS CORRELATOS

Há um interesse cada vez maior, por diferentes comunidades de pesquisa em descobrir a direção de fontes sonoras. Por isso, esta seção destina-se a investigar na literatura alguns dos principais trabalhos que envolvem localização de fontes sonoras, sendo descritos na sequência.

Vieira et al. (2001) descreve como foi implementada a localização de fonte sonora sobre um robô para que este consiga seguir a fonte sonora. Este robô possui como unidade central de processamento um DSP C30 da Texas *Instrument*, sendo que este realiza todo o processamento dos sinais e ainda controla os motores elétricos. A localização descrita em Vieira et al. (2001, p. 243) utiliza dois microfones e possui algumas particularidades, tais como a localização de uma fonte sonora conhecida e de uma desconhecida. Para encontrar a localização de uma fonte sonora conhecida foi utilizado um esquema baseado em *matched filter*³, que se revelou muito eficaz e imune ao ruído. Já para a localização de uma fonte sonora desconhecida utilizou-se o método da correlação cruzada, entre os dois canais de recepção de áudio. Outra particularidade é que este sistema é capaz de descobrir a fonte sonora mesmo quando esta apresenta-se na retaguarda do robô. Para isso ele usa um algoritmo de localização, sendo que este algoritmo realiza um pequeno movimento angular em direção a fonte sonora para que o robô possa sempre encontrar a direção real.

A proposta de Zhang (2007) é implementar o sistema auditivo humano. A principal característica de seu trabalho é que ele não utiliza somente a diferença de tempo entre os dois ouvidos, mas também utiliza o princípio das mudanças que a cabeça e ouvidos causam na onda sonora (HRTF). Para realizar a implementação do sistema, Zhang (2007, p. 2) utiliza apenas dois microfones e defende que esta é a melhor maneira de alcançar as aplicações citadas acima, pois utilizando apenas dois microfones, a estrutura torna-se mais simples e a implementação de um sistema equivalente em miniatura fica mais fácil. Com apenas dois microfones a quantidade de sinais que devem ser processados é menor e com isso este processamento ocorre de maneira mais rápida. No método que ele propõe, a direção da fonte sonora é baseada em dois ângulos; o ângulo horizontal e o ângulo vertical. O primeiro é estimado com base no princípio do ITD, utilizando a diferença entre os tempos de chegada da onda sonora aos ouvidos, no caso os microfones. Já o ângulo vertical é estimado pela análise

3 A função do filtro casado (*matched filter*) é encontrar uma estimativa otimizada para o tempo de chegada do eco, tendo como informação a priori um modelo do sinal esperado.

do espectro de correlação usando HRTF.

Balconi (2008) apresenta uma solução utilizando três microfones, no qual o objetivo além de encontrar a direção do som no espaço 2D, propõe estimar a localização exata da fonte sonora (coordenadas x e y no espaço 2D). Para encontrar a posição da fonte sonora foi utilizado os conceitos de correlação cruzada e do ITD, conseguindo assim as direções (ângulos) das fontes sonoras. Utilizando então três microfones, são calculadas três direções e, baseado no conceito de triangulação, a posição da fonte sonora é definida (BALCONI, 2008, p. 30).

O Quadro 6 mostra de forma resumida as principais características destes sistemas tendo como base critérios considerados importantes, extraídos a partir dos conceitos descritos no decorrer desta seção.

características / trabalhos relacionados	Balconi (2008)	Vieira et al. (2001)	Zhang (2007)
localização 2D	X	X	X
localização 3D	-	-	X
cálculo de ângulo da fonte sonora	X	X	X
cálculo da distância da fonte sonora	X	-	-
algoritmo em tempo real	-	X	X
ITD	X	X	-
ILD	-	-	X
HRTF	-	-	X
quantidade de microfones	3	2	2

Quadro 6 - Características dos trabalhos relacionados

As informações foram dispostas em colunas, representando na vertical os trabalhos analisados e as linhas apresentam as características de cada sistema, indicando semelhanças e/ou diferenças.

A partir do Quadro 6, tem-se o estado da arte sobre localização de fonte sonora. É possível observar que em todos os trabalhos preocupam-se com a quantidade de sinais processados para estabelecer a localização da fonte sonora (redução do custo computacional).

Neste trabalho tem-se por objetivo a localização da fonte sonora em um espaço 2D, no qual os trabalhos acima serviram como base para pesquisa de técnicas e possíveis problemas e soluções relacionados ao objetivo.

3 DESENVOLVIMENTO DO ALGORITMO

Nesta seção são apresentados os requisitos do trabalho proposto (seção 3.1), a especificação na seção 3.2, o desenvolvimento do algoritmo na seção 3.3 e os resultados e discussões na seção 3.4.

3.1 REQUISITOS DO TRABALHO PROPOSTO

O algoritmo de localização de fontes sonoras proposto deverá:

- a) implementar um modelo computacional capaz de medir o atraso entre três *streams* de áudio, capturados pelos dispositivos de entrada (microfones) (Requisito Funcional – RF);
- b) implementar um modelo matemático para encontrar os ângulos direcionais (ângulo horizontal e vertical), disponibilizando como saída as coordenadas, ângulo e distância que indiquem a localização da fonte sonora desconhecida (RF);
- c) ser modularizado, para que seja possível incluir outras etapas no processamento para eventuais aplicações e usos externos, como por exemplo, filtros específicos que desconsideram apenas determinados tipos de sons ocorridos no ambiente (Requisito Não-Funcional – RNF);
- d) ser implementado utilizando o ambiente de desenvolvimento Eclipse (RNF);
- e) ser implementado utilizando a linguagem de programação Java (RNF);
- f) ser multiplataforma (RNF).

3.2 ESPECIFICAÇÃO

A especificação do algoritmo foi realizada através da metodologia de orientação a objetos. Ela foi representada em diagramas da *Unified Modeling Language* (UML), utilizando a ferramenta Top Coder UML Tool. Inicialmente, apresenta-se o diagrama de casos de uso e em seguida o diagrama de classes.

3.2.1 Diagrama de casos de uso

A Figura 10 exibe o diagrama de casos de uso.

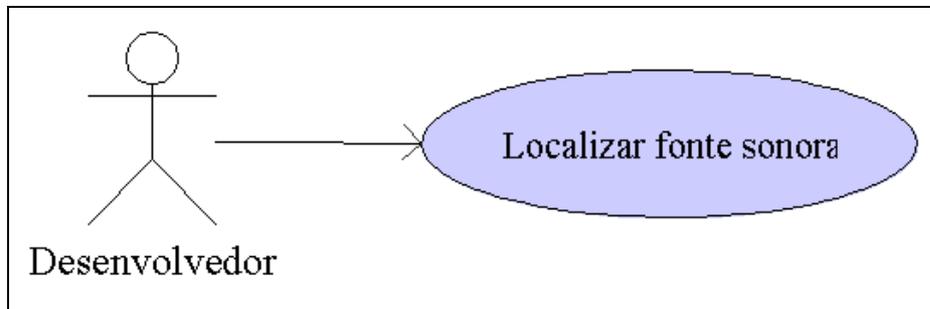


Figura 10 - Diagrama de caso de uso

O caso de uso da aplicação é descrito a seguir: UC01 – localizar fonte sonora: estima a localização da fonte sonora desconhecida.

3.2.2 Diagrama de classes

Responsável por fornecer a visão macro do relacionamento entre as classes do projeto, o diagrama de classes é apresentado na Figura 11.

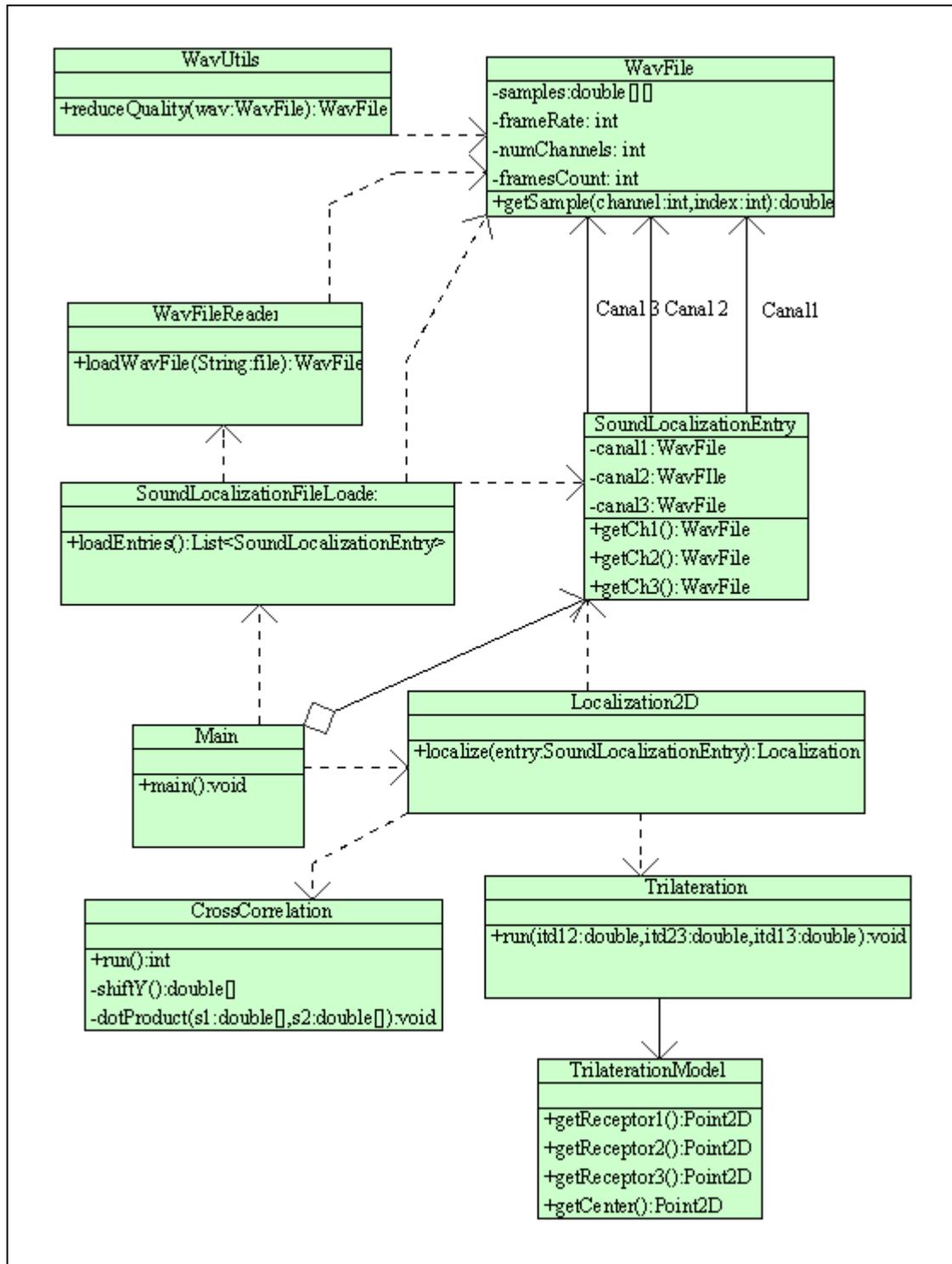


Figura 11 - Diagrama das classes desenvolvidas

Para melhor entendimento do diagrama de classes apresentado na Figura 11, as classes foram divididas em três categorias: formato de arquivo, carga dos arquivos sonoros em memória e algoritmo de localização.

As classes da categoria formato de arquivo são: **WavFile**, **WavFileReader**. Conforme mencionado na seção 2.1.1, o arquivo de som tem um formato específico, após carregá-lo, se trabalha com o arquivo sonoro em um formato de som genérico, independente do formato de

arquivo. A classe `WavFileReader`, conhece esse formato de arquivo, lendo-o a partir do formato WAV (*Waveform Audio Format*), com o método `loadWavFile`. Após a leitura feita pelo `WavFileReader`, é gerado um arquivo de som genérico `WavFile`, portanto a classe `WavFile` tem-se o sinal digital no formato genérico de processamento. A classe `WavUtils`, tem operações para reduzir o *frame rate* do sinal (ver seção 2.1.1), com o método `reduceQuality`.

As classes de carga de arquivo sonoros em memória são: `SoundLocalizationFileLoader`, `SoundLocalizationEntry`. A classe `SoundLocalizationFileLoader`, tem a lógica para carregar os arquivos em memória separando por canais 1, 2 e 3 corresponde ao sinal de cada microfone do modelo de captura de som, com o método `loadEntries`. Após ler e separar os arquivos em diferentes canais, a classe `SoundLocalizationEntry`, agrupa um exemplo de som nos três canais para processamento, no qual representa a entrada de dados que será processada pelo algoritmo de localização sonora, que podem ser acessados pelos métodos `getCh1`, `getCh2`, `getCh3`. Então a classe `SoundLocalizationFileLoader` vai gerar uma lista de `SoundLocalizationEntry`, representando várias entradas para processamento.

As classes de algoritmo são: `Localization2D`, `CrossCorrelation`, `Trilateration`, `TrilaterationModel`. A classe `TrilaterationModel` representa o modelo matemático definido pelo posicionamento dos microfones, portanto esta classe tem os três pontos no plano cartesiano que representam os microfones como centro do plano, que podem ser acessados pelos métodos `getReceptor1`, `getReceptor2`, `getReceptor3`, e também possui o método `getCenter`, que retorna o centro dos três receptores. A classe `CrossCorrelation` contém o algoritmo de correlação cruzada (ver seção 2.1.2.2.1) que é abordado na seção 3.3.1.1, que é executado pelo método `run`. A classe `Trilateration` contém o algoritmo de triangulação que é abordado na seção 3.3.2, que é executado pelo método `run`. A classe `Localization2D` unifica os processos, com o método `localize`, pegando a correlação resultante da classe `CrossCorrelation` e envia para a classe `Trilateration`, no qual faz toda a operação de localização.

A classe `Main`, é a classe onde tudo começa, no método `main`. A classe `Main` é responsável por utilizar a classe `SoundLocalizationFileLoader` para adquirir as entradas e encaminhar para o algoritmo de localização, na classe `Localization2D`, cada uma destas entradas.

3.3 DESENVOLVIMENTO DO ALGORITMO PARA LOCALIZAÇÃO DA FONTE SONORA

Nas próximas seções detalham-se os passos realizados no desenvolvimento do algoritmo de localização da fonte sonora. A explicação será dividida em duas partes: encontrar o atraso dos sons entre os microfones e em seguida, por triangulação, encontrar a fonte sonora baseado no atraso do som.

3.3.1 Cálculo do ITD

Considere o cenário da Figura 12.

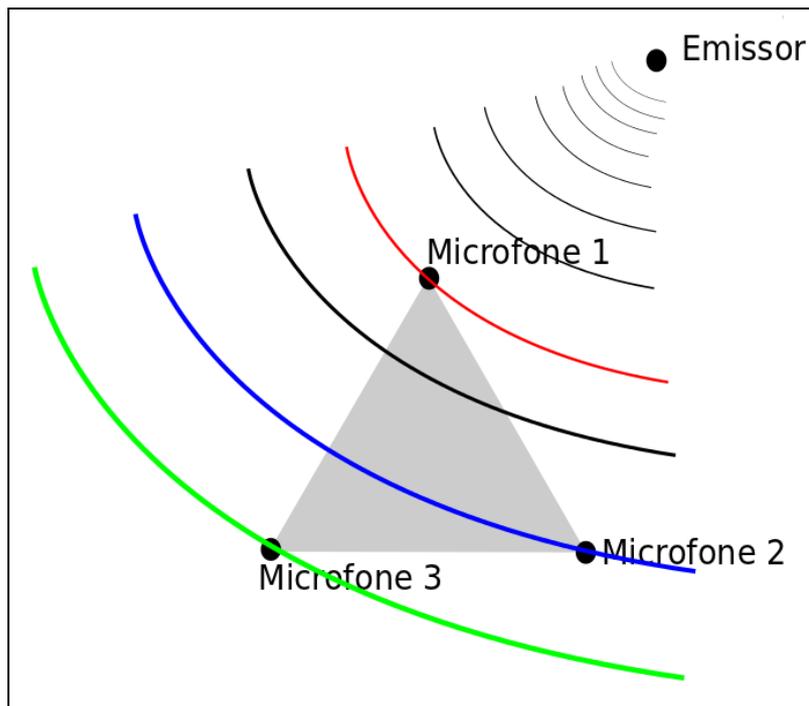


Figura 12 - Modelo de três microfones exemplificando um cenário de captura, e o desenho de um emissor, representando o som chegando nos três microfones

Os três microfones estão posicionados de forma que formem um triângulo equilátero (ver seção 2.3.2), com uma distância fixa de 30 cm entre cada um dos microfones.

Conforme percebe-se na Figura 12, o som emitido pela fonte sonora chega em momentos diferentes nos microfones e para conseguir estimar a posição da fonte sonora é necessário calcular ITD (ver seção 2.1.2.2).

As seções seguintes irão abordar o desenvolvimento do algoritmo para o cálculo do atraso da fonte sonora.

3.3.1.1 Correlação cruzada

Primeiramente é necessário encontrar a diferença física do som. O som como abordado na 2.1.1, é formado por *frames*, então utilizando a técnica de correlação cruzada é possível encontrar quantos *frames* de atraso um sinal sonoro está do outro.

O modelo de captura representado pela Figura 12 mostra o som emitido por uma fonte sonora chegando aos três microfones em momentos diferentes.

Vamos chamar o sinal capturado pelo microfone 1 de s_1 , do microfone 2 de s_2 e do microfone 3 de s_3 .

Desta forma aplicamos a correlação cruzada (ver seção 2.1.2.2.1) para cada par de sinais capturados pelos microfones conforme o Quadro 7.

```
frameDiff12 = CrossCorrelation(s1, s2)
frameDiff23 = CrossCorrelation(s2, s3)
frameDiff13 = CrossCorrelation(s1, s3)
```

Quadro 7 - Operações de correlação cruzada efetuadas

Aplicando as operações conforme Quadro 7 são obtidos três variáveis com as diferenças em *frames* entre sinais sonoros, que serão utilizados para calcular as diferenças no domínio de tempo na seção a seguir.

O Quadro 8, mostra o algoritmo para calcular a correlação cruzada da fonte sonora.

```

// Faz a variação do sinal data no tempo em shiftCount frames
private static double[] shift(double[] data, int shiftCount) {
    double c[] = new double[data.length];
    for (int i = 0; i < data.length; i++) {
        if ((shiftCount + i >= 0) && (shiftCount + i < data.length)) {
            c[i] = data[shiftCount + i];
        }
    }
    return c;
}
// Correlação cruzada entre os sinais a e b, retorna diferença em frames
public static int run(double[] a, double[] b) {
    double correlation = Double.NEGATIVE_INFINITY;
    int frame = 0;
    for (int i = -b.length + 1; i < b.length; i++) {
        double slidingY[] = shift(b, i);
        double res = dot(a, slidingY);
        if (res > correlation) {
            frame = i;
            correlation = res;
        }
    }
    return frame;
}
private static double dot(final double[] data, final double[] slide) {
    double sum = 0;
    for (int i = 0; i < Math.min(data.length, slide.length); i++){
        sum += slide[i] * data[i];
    }
    return sum;
}
}

```

Quadro 8 - Código fonte representando implementação da correlação cruzada por *dot product*

O algoritmo basicamente contém três métodos, o `run`, `shift` e `dot`. O método `run` é o que contém o *loop* principal, que vai varrer um dos sinais a partir de `-tamanho` até `tamanho`, ou seja, essa é a parte que faz um sinal “atravessar” o outro (ver seção 2.1.2.2.1). Para cada iteração deste *loop* principal, é chamado o método `shift`, para fazer a variação do sinal para aquele índice (escorregar o sinal). Desta forma, para cada uma destas variações feitas dentro do *loop* principal é calculado o *dot product* (ver seção 2.1.2.2.1) chamando o método `dot` para o sinal base e o sinal que está variando no tempo. No fim de cada *loop* é comparado a correlação corrente com a maior. Dessa forma, sabe-se em que variação a correlação foi maior. Ao final do algoritmo, sabe-se qual foi a variação ou índice dentro do *loop* principal que resultou na maior correlação. Este índice vai indicar a diferença de *frames* entre os dois sinais da entrada.

3.3.1.2 Cálculo da diferença no domínio de tempo

Sabendo algumas características dos sinais, como o *frame rate* e o resultado da correlação cruzada conseguiu-se estimar o tempo de diferença entre os sinais captados pelos

microfones.

Utilizando a relação acima introduzida, entre o *frame count* e o *frame rate*, é substituído o *frame count* pela quantidade de *frames* de diferença entre os sinais. Então haverá um cálculo para cada diferença encontrada conforme Quadro 9.

$$\begin{aligned} \text{secondsDiff12} &= \text{frameDiff12} / \text{frameRate} \\ \text{secondsDiff23} &= \text{frameDiff23} / \text{frameRate} \\ \text{secondsDiff13} &= \text{frameDiff13} / \text{frameRate} \end{aligned}$$

Quadro 9 - Cálculos para encontrar atrasos em segundos

Lembrando que o *frameRate* são quantos *frames* por segundo um sinal sonoro tem, então o resultado do cálculo será em segundos. Mais uma característica a ser percebida, é que é utilizado a mesma variável de *frameRate* no cálculo, isso é válido pois nesta parte do algoritmo, já passou-se pela correlação cruzada e como a correlação cruzada estabelece que os sinais tenham o mesmo *frameRate*, o *frameRate* aqui utilizado é o mesmo para todos os cálculos.

Estes atrasos, ou diferenças no domínio de tempo, serão utilizados para descobrir alguns ângulos que são abordados nas seções seguintes.

3.3.2 Triangulação para encontrar a fonte sonora

As seções seguintes abordam os algoritmos e cálculos utilizados para estabelecer a triangulação da fonte sonora, estimando a posição da mesma. Primeiramente são abordados as transformações dos parâmetros de entrada para estimar a localização sonora e depois as técnicas de triangulação.

3.3.2.1 Primeiro passo, transformar parâmetros para domínio de espaço

Os atrasos calculados estão no domínio de tempo, como o objetivo é estimar a localização de uma fonte sonora, a localização é relativa à posição, portanto, no domínio de espaço.

Desta forma, iremos utilizar fundamentações abordadas para fazer esta conversão: a relação física MRU (seção 2.2.1); e a velocidade do som (seção 2.2.2).

A Figura 13, demonstra o que deseja-se encontrar.

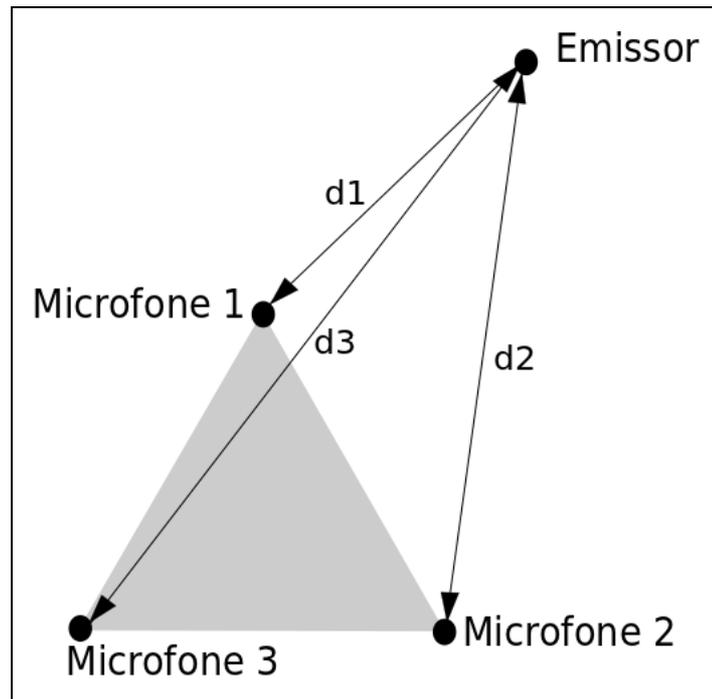


Figura 13 - representação dos parâmetros de distância

Observando a Figura 13, percebe-se que o emissor se encontra a uma distância não definida de cada microfone. E o que é conhecido até o momento são apenas as diferenças dos sinais no domínio de tempo. Quando estes parâmetros forem convertidos para domínio de espaço, vamos obter os parâmetros representados no Quadro 10.

$d1 - d2$
$d2 - d3$
$d1 - d3$

Quadro 10 - Parâmetros obtidos após conversão para domínio de espaço

Os parâmetros do Quadro 10 são exatamente a diferença entre as distâncias do emissor com os microfones representados na Figura 13.

Então os parâmetros conhecidos para conversão estão representados no Quadro 11.

<p>Velocidade = Constante da velocidade sonora V_s Tempo = secondsDiff12, secondsDiff23 e secondsDiff13</p>

Quadro 11 - Parâmetros de velocidade e tempo conhecidos

O que é procurado é a distância, então é necessário alterar a fórmula original (ver seção 2.2.1) conforme Quadro 12.

<p>Velocidade = Distância / Tempo Velocidade * Tempo = Distância</p>

Quadro 12 - Alteração na fórmula MRU

Desta forma é obtido as três diferenças conforme Quadro 13.

$\text{distanceDiff1 2} = V_s / \text{secondsDiff1 2}$ $\text{distanceDiff2 3} = V_s / \text{secondsDiff2 3}$ $\text{distanceDiff1 3} = V_s / \text{secondsDiff1 3}$
--

Quadro 13 - Diferenças de distância encontradas

Assim, obtêm-se os parâmetros no domínio de espaço.

3.3.2.2 Ângulos dos pontos médios

Para ambas as tentativas abordadas nas seções 3.3.2.3 e 3.3.2.4, são necessários os ângulos representados na Figura 14.

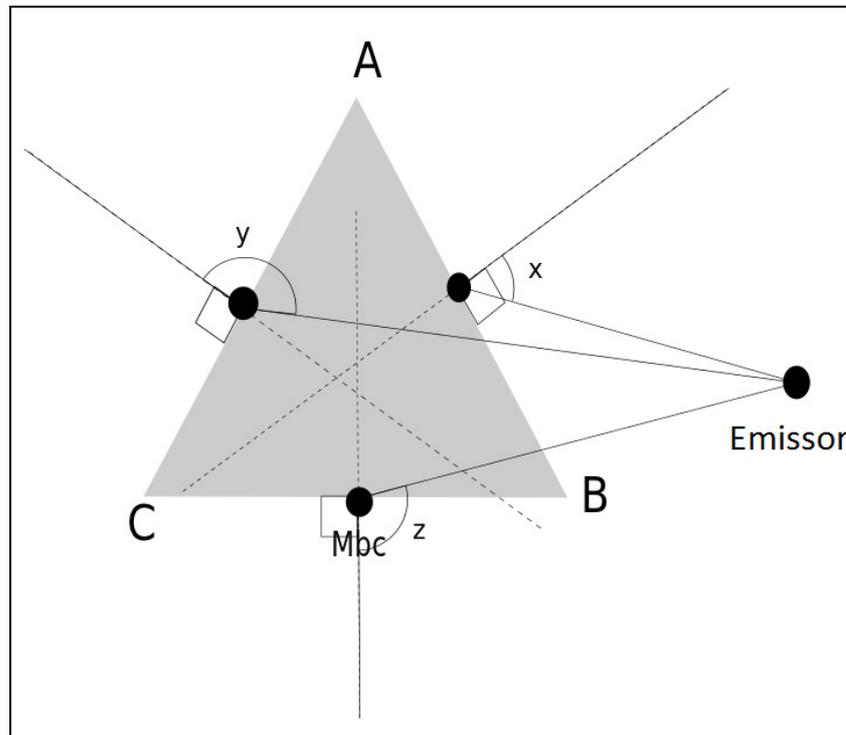


Figura 14 - ângulos entre perpendiculares dos lados e o emissor

Os ângulos representados na Figura 14 são obtidos utilizando a relação representada pela Figura 9 na seção 2.3.6. Nessa relação, existe a diferença $D_l - D_r$, que para este trabalho, é a diferença entre as distâncias do emissor com os microfones. Estas diferenças são obtidas conforme a seção 3.3.2.1, no qual são aqui denominadas como: distanceDiff12 , distanceDiff23 e distanceDiff13 .

Desta forma, é possível encontrar x , y e z representados na Figura 14 conforme Quadro 14.

$$\begin{aligned}
 x &= \arcsin \left(\frac{\text{distanceDiff1 } 2}{\text{distância}(A,B)} \right) \\
 y &= \arcsin \left(\frac{\text{distanceDiff2 } 3}{\text{distância}(B,C)} \right) \\
 z &= \arcsin \left(\frac{\text{distanceDiff1 } 3}{\text{distância}(A,C)} \right)
 \end{aligned}$$

Quadro 14 - Obtenção de ângulos

No entanto, esta relação hiperbólica é ambígua, como pode ser vista na seção que é representado o modelo da hipérbole (seção 2.3.6), assim encontramos também os ângulos suplementares⁴ de x , y e z representados no Quadro 15.

$$\begin{aligned}
 180 - x \\
 180 - y \\
 180 - z
 \end{aligned}$$

Quadro 15 - Ângulos suplementares encontrados

3.3.2.3 Tentativa 1, relações angulares e trigonometria básica

O objetivo desta seção é tentar encontrar a fonte sonora, por meio de trigonometria básica. A Figura 15 mostra a abordagem utilizada nesta tentativa.

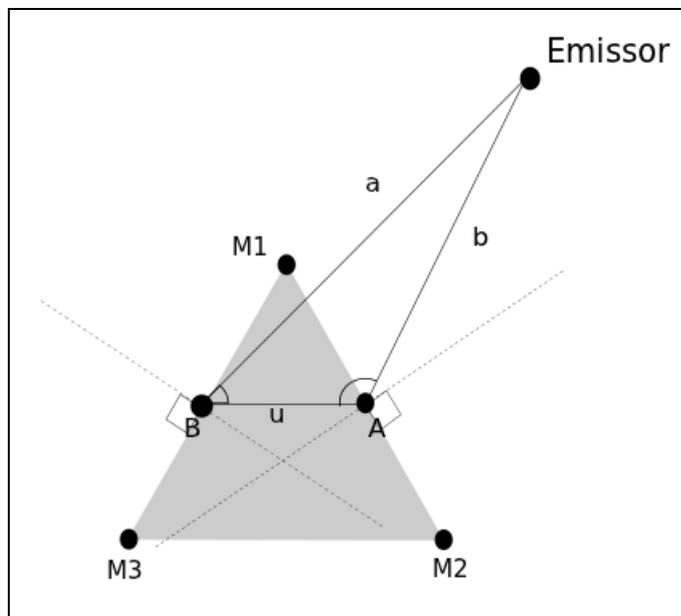


Figura 15 - Modelo matemático da primeira tentativa

O modelo matemático representado na Figura 15 visa encontrar a fonte sonora por meio de trigonometria básica.

Para tanto, os parâmetros já conhecidos são:

- a) $M1$, $M2$ e $M3$: os microfones;

- b) A e B: pontos médios entre os microfones M1, M2 e M1, M3 respectivamente;
- c) u: distância entre A e B;
- d) distanceDiff12, distanceDiff23, distanceDiff13: diferenças de distâncias, na abordagem desta solução seria apenas necessário distanceDiff12 e distanceDiff13.

O que deseja-se encontrar:

- a) a e b: distância entre os pontos médios e o emissor;
- b) Emissor (x, y): a posição do emissor de fato.

No entanto, percebeu-se que como os ângulos x , y e z são ambíguos, ficaria impossível encontrar ângulos internos do triângulo. Em todas as relações em que se tentou formar, poderiam dar ângulos internos ou externos do triângulo, ou mesmo ângulos que não fariam sentido.

Neste momento, percebeu-se que este caminho não era um caminho possível para encontrar a fonte sonora. A técnica não chegou a ser desenvolvida, pois durante as demonstrações matemáticas foi constatado que não há como remover as ambiguidades existentes apenas combinando os ângulos.

3.3.2.4 Tentativa 2, estimar posição baseado em interseção de retas

Para melhor entendimento da técnica empregada, é apresentado na Figura 16, o objetivo final, que é encontrar o emissor representado com as intersecções das retas formadas pelos ângulos x , y e z .

4 Ângulos que somados resultam em 180, são chamados de ângulos suplementares.

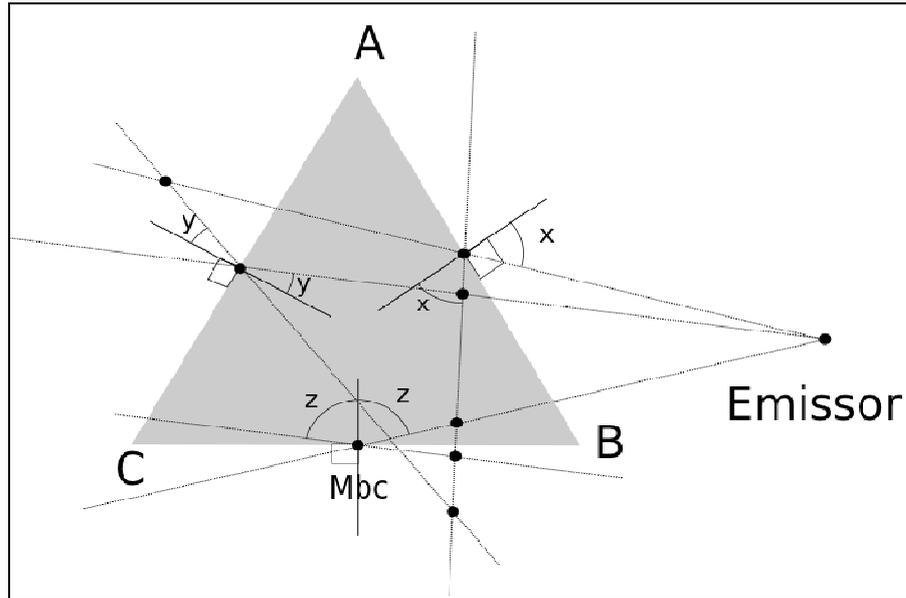


Figura 16 - Objetivo final da técnica empregada

Na Figura 16, os ângulos representados por x , y e z são os ângulos e seus suplementares encontrados utilizando a relação hiperbólica. As retas que interceptam com o centro dos lados são as retas formadas pelas aberturas destes ângulos e os pontos são as interseções destas retas, sendo dois deles, o mesmo ponto no qual representa o emissor.

Nas próximas seções serão detalhados o funcionamento e desenvolvimento desta técnica.

3.3.2.4.1 Encontrando os ângulos

Os ângulos x , y e z representados na Figura 16 (seção anterior), já são conhecidos, no entanto, estes ângulos são entre às perpendiculares dos lados do triângulo e o emissor.

Para poder estimar a posição utilizando esta técnica, é necessário que todos os ângulos sejam normalizados em relação ao eixo das abscissas. Para atingir este objetivo em específico, será utilizado o teorema de Tales (ver seção 2.3.1) e as propriedades do triângulo equilátero (ver seção 2.3.2).

Algumas características importantes são representadas na Figura 17.

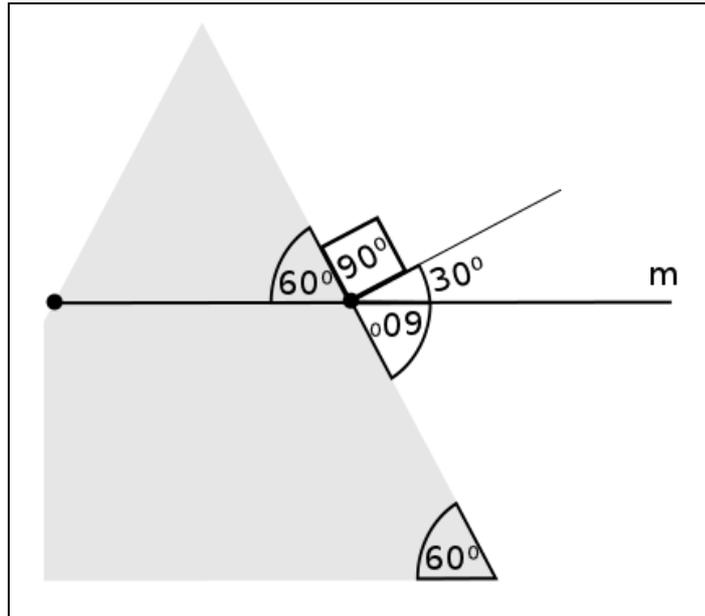


Figura 17 - características e relações utilizadas para encontrar os ângulos.

A reta representada por m na Figura 17 é a reta entre os pontos médios dos dois lados do triângulo equilátero, além disso, esta reta é paralela ao eixo das abscissas. Como um triângulo equilátero tem os três ângulos internos medindo sessenta graus, é possível utilizar o teorema de Tales, para deduzir que a abertura entre a reta do lado do triângulo e a reta m tem a medida de sessenta graus também, sendo que a reta m é paralela, respeitando assim a condição da relação. Utilizando novamente o teorema de Tales, é possível encontrar o outro ângulo, que é o ângulo oposto pelo vértice do primeiro ângulo encontrado, que mede também sessenta graus, representado pelo ângulo de sessenta graus externo ao triângulo conforme Figura 17.

Todas essas relações e ângulos quando encontrados servem para encontrar a diferença de ângulo da perpendicular saindo do lado do triângulo para se alinhar com o eixo das abscissas.

A reta perpendicular em relação ao lado, por definição tem o ângulo de noventa graus de abertura. Como a reta m que é paralela ao eixo das abscissas tem uma abertura de sessenta graus em relação a lado do triângulo, então a diferença necessária para alinhar a perpendicular é $90 - 60 = 30$, onde noventa é a abertura da perpendicular, e sessenta a abertura da reta m .

Desta forma é encontrado a primeira diferença de ângulo para um dos lados do triângulo. Equivalentemente é encontrado também a diferença de trinta graus do lado esquerda do triângulo, pelas mesmas relações, conforme observa-se na Figura 17.

Para a lado inferior do triângulo, é mais simples encontrar esta diferença. Como a lado inferior é paralela ao eixo das abscissas e a reta perpendicular a este lado tem uma abertura de noventa graus, pelo teorema de Tales, teremos esta mesma abertura em relação ao eixo das

abscissas, desta forma a diferença de ângulos para alinhamento da perpendicular no lado inferior é de noventa.

Com estas diferenças, é possível alinhar todos os ângulos já encontrados com o eixo das abscissas, conforme a Figura 16 que representa o objetivo desta técnica.

Para alinhar os ângulos x , y e z , será utilizado as diferenças de alinhamento referentes aos seus ângulos dos lados correspondentes. Os cálculos que irão encontrar os alinhamentos são apresentados no Quadro 16.

$$\begin{aligned} x_{\text{Aligned}} &= x + 30^\circ \\ y_{\text{Aligned}} &= y - 30^\circ \\ z_{\text{Aligned}} &= 90^\circ - z \end{aligned}$$

Quadro 16 - Normalização de ângulos

Onde x_{Aligned} é correspondente a lado um, somando o ângulo necessário para alinhar com o eixo das abscissas. Já y_{Aligned} é correspondente ao lado oposto, portanto como o lado oposto do triângulo é em direção oposta, o ângulo tem que ser normalizado para a direção oposta, então é subtraído o ângulo necessário. O ângulo z_{Aligned} é correspondente ao lado inferior, e o lado inferior tem sua própria diferença sendo que já está alinhado com o eixo das abscissas. O que deve ser calculado é, o ângulo z encontrado, representado na Figura 16, é em relação a perpendicular do lado inferior. Assim a normalização a ser feita, é achar o ângulo correspondente ao próprio lado inferior, por isso é subtraído z de 90 , onde é obtido z_{Aligned} .

Uma vez encontrado os ângulos normalizados, é necessário encontrar os suplementares destes ângulos. Pois como os ângulos iniciais x , y e z encontrados são ambíguos, é necessário considerar estas ambiguidades para poder estimar a fonte sonora.

O Quadro 17 representa as definições dos ângulos suplementares.

$$\begin{aligned} x_{\text{AlignedSup}} &= 180 - x_{\text{Aligned}} \\ y_{\text{AlignedSup}} &= 180 - y_{\text{Aligned}} \\ z_{\text{AlignedSup}} &= 180 - z_{\text{Aligned}} \end{aligned}$$

Quadro 17 - Ângulos suplementares

3.3.2.4.2 Encontrando as equações lineares

Foi necessário definir uma equação de reta (ver seção 2.3.4) para cada ângulo normalizado ao eixo das abscissas encontrado, para então fazer as interseções abordadas na próxima seção. Então é necessário utilizar a definição da reta que é dada pela seguinte

equação: $y - y_1 = a * (x - x_1)$.

No entanto, são necessários dois parâmetros para definir esta equação: coeficiente angular e um ponto pertencente à reta. Estes dois parâmetros já foram obtidos pelos passos anteriores. Um deles é o coeficiente angular (ver seção 2.3.4) calculado utilizando os ângulos normalizados no qual foram abordados na seção 3.3.2.4.1. O outro é um ponto qualquer pertencente a reta, no qual são utilizados os pontos médios entre os lados, pois são os vértices que continham os ângulos inicialmente.

Será utilizado a equação no formato reduzido da reta, assim basta substituir os parâmetros encontrados, conforme Quadro 18.

$$\begin{aligned} y - y_1 &= a * (x - x_1) \\ y - y_1 &= a * x - a * x_1 \\ y &= a * x - a * x_1 - y_1 \end{aligned}$$

Quadro 18 - Isolando variáveis na equação da reta

Onde a representa o coeficiente angular e (x_1, y_1) representam um ponto da reta no Quadro 18.

Para ficar mais claro, serão utilizados nomes para os pontos médios conforme mostra a Figura 18.

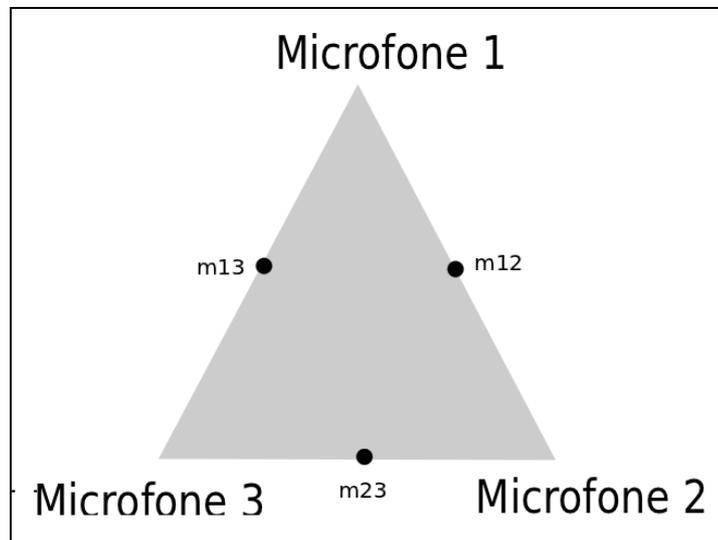


Figura 18 - Modelo pontos médios e microfones

E então é possível encontrar as seis equações que definem as retas, apenas substituindo os parâmetros conforme Quadro 19.

$$\begin{aligned}
 y &= \tan(x\text{Aligned}) * x - \tan(x\text{Aligned}) * m12.x - m12.y \\
 y &= \tan(x\text{AlignedSup}) * x - \tan(x\text{AlignedSup}) * m12.x - m12.y \\
 y &= \tan(y\text{Aligned}) * x - \tan(y\text{Aligned}) * m13.x - m13.y \\
 y &= \tan(y\text{AlignedSup}) * x - \tan(y\text{AlignedSup}) * m13.x - m13.y \\
 y &= \tan(z\text{Aligned}) * x - \tan(z\text{Aligned}) * m23.x - m23.y \\
 y &= \tan(z\text{AlignedSup}) * x - \tan(y\text{AlignedSup}) * m23.x - m23.y
 \end{aligned}$$

Quadro 19 - Encontrando equações das retas

3.3.2.4.3 Encontrando as interseções

Para encontrar as interseções, basta resolver um sistema de equações simples para cada par de retas, resultando num sistema dois por dois (ver seção 2.3.5).

Utilizando novamente a equação simplificada da reta, é possível formar um sistema genérico e achar a solução para este sistema e então apenas substituir os valores das equações encontradas com as equações do sistema. A equação reduzida da reta é definida por: $y = a * x + b$, como será montado um sistema entre duas equações de retas quaisquer, é possível definir este sistema genérico dois por dois conforme Quadro 20.

$$\begin{aligned}
 y &= a1 * x + b1 \\
 y &= a2 * x + b2
 \end{aligned}$$

Quadro 20 - Sistema linear genérico entre duas equações de retas

O sistema é resolvido por substituição conforme Quadro 21.

$$\begin{aligned}
 a1x + b1 &= a2x + b2 \\
 a1x - a2x &= b2 - b1 \\
 x * (a1 - a2) &= b2 - b1 \\
 x &= \frac{b2 - b1}{a1 - a2}
 \end{aligned}$$

Quadro 21 - Resolução de um sistema de equações genéricas de reta

No Quadro 21 $a1$ e $a2$ são os coeficientes angulares, $b1$ e $b2$, os coeficientes lineares. O coeficiente angular das equações é o valor que está multiplicando o x e o coeficiente linear é todo o resto da fórmula no qual valores já são conhecidos. Uma vez encontrado x basta substituir o valor de x em qualquer uma das fórmulas do sistema, para encontrar y .

Assim, é encontrado um ponto de interseção para cada de equações definidas. A única possibilidade de não haver uma intersecção é quando há duas retas paralelas, no qual esta exceção é tratada.

3.3.2.4.4 Estimando a posição da fonte sonora

Como foi encontrada a lista de interseções, será necessária estimar qual destas é a fonte sonora. Teoricamente três interseções iguais seriam o ponto onde a fonte sonora se encontra, mas na prática, todos estes cálculos das operações efetuadas vão acumulando pequenos erros em operações de ponto flutuante. Desta forma, esta técnica não previu os erros possíveis nos cálculos e como poderiam ser contornados ou amenizados. E também na prática seriam coincidentes nos casos onde os três pontos seriam iguais. Então foi necessário desenvolver uma heurística para poder encontrar os pontos mais prováveis que representariam a fonte sonora.

Primeiro elimina-se os pontos encontrados muito próximos do modelo de captura, para desconsiderar os resultados não válidos de acordo com a relação de hipérbole apresentado na seção 2.3.6. Então a partir do centro do modelo de captura, todos as interseções que estiverem no diâmetro do tamanho de uma dos lados do modelo, serão eliminadas. A Figura 19 representa as interseções de um exemplo hipotético, no qual as interseções circundados por um círculo tracejado em vermelho (ao redor do modelo) são as interseções eliminadas.

Para decidir entre as interseções restantes, para cada três pontos das interseções é calculada a distância entre cada par destes três pontos e somado, a menor distância entre os três pontos serão considerados os três pontos mais próximos entre si. A média destes três pontos selecionados será considerada a fonte sonora como pode ser visto na Figura 19.

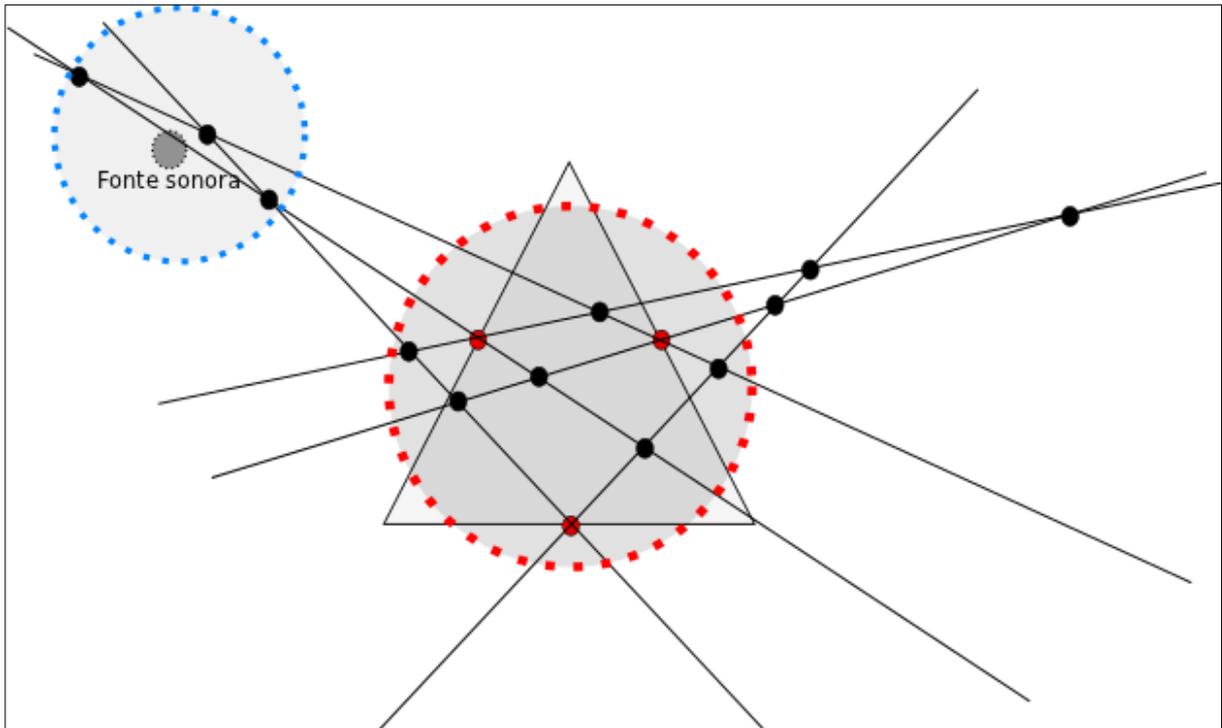


Figura 19 - Representação das interseções encontradas e a fonte sonora

Na Figura 19, pode ser observar os pontos em cima das retas, todas as interseções. As interseções em vermelho são também os pontos médios entre os lados do modelo. Como já explanado acima, as interseções dentro do círculo pontilhado em vermelho são as interseções desconsideradas. Dentre as interseções que restaram, as que estão circundadas por um círculo tracejado em azul, são as três interseções selecionadas. A média das interseções selecionadas será considerado a localização da fonte sonora, representada pelo ponto maior em cinza com a borda tracejada.

3.3.3 Otimizações

O algoritmo de correlação cruzada utilizado é de complexidade $O(n^2)$ enquanto o cálculo da triangulação é de complexidade $O(1)$. Portanto todo o gargalo está no algoritmo de correlação cruzada utilizado.

Portanto foi feita uma otimização para diminuir o *frame rate* dos sinais utilizados, no qual em um algoritmo de complexidade $O(n^2)$ qualquer quantidade reduzida da entrada oferece um ganho considerável no resultado final.

Considerando a diminuição de dados para estimar os atrasos do som entre os microfones, a precisão das estimativas dos atrasos e da triangulação diminuirá.

Desta forma o algoritmo é parametrizável para utilizar um *frame rate* definido como qualidade para estimar as posições, porém, dependendo da aplicação para o qual o algoritmo será utilizado, a qualidade pode ser menor ou até bem menor.

Por padrão será utilizado o *frame rate* dos sinais, se os sinais tiverem um *frame rate* diferente será padronizado entre os sinais o menor dos *frame rates*.

Se um *frame rate* for definido, será utilizado o menor dos *frame rates* entre o definido pelo parâmetro de entrada e os *frame rates* dos sinais de entrada.

Para diminuir o *frame rate* utiliza-se da seguinte heurística:

- a) entre os *frame rates* definidos, encontra-se o menor;
- b) dividi-se o *frame* do sinal, pelo *frame rate* encontrado no passo anterior, assim encontra-se o coeficiente de remoção C_r ;
- c) se C_r é 1, acabe o algoritmo, os sinais tem o mesmo *frame rate*;
- d) varre-se o sinal, jogando os valores do sinal em uma nova estrutura;
- e) para cada C_r *frames* durante a varredura do sinal, um *frame* é ignorado e não é adicionado na nova estrutura do sinal.

Desta forma diminui-se a qualidade do sinal de entrada. Este pré-processamento pode parecer pesado, no entanto sua complexidade é de $O(n)$. Levando isto em consideração e que o algoritmo de correlação é de complexidade $O(n^2)$, para cada *frame* reduzido da entrada, aproximadamente $O(n)$ iterações serão reduzidas do algoritmo de correlação cruzada, pois $O(n^2) = O(n*n)$ e que se $O(n*(n-x))$, sendo x a quantidade de *frames* reduzidos, $O(x*n)$ iterações foram reduzidas do algoritmo original.

3.3.4 Técnicas e ferramentas utilizadas

Como ambiente de desenvolvimento foi utilizado a IDE eclipse com a linguagem de programação Java 1.6.

Para captação dos sons, foi utilizado microfones do modelo *MD441* com o foco de direção de sessenta graus de abertura em frente do microfone.

Para medição dos atrasos do som emitido entre os microfones foi utilizada a técnica de correlação cruzada com *dot product*.

Para montagem e edição das imagens foram utilizadas as ferramentas: InkScape e GIMP.

3.3.5 Operacionalidade da implementação

A operacionalidade do algoritmo dependerá do tipo da aplicação que será utilizada.

Para a correlação cruzada, conforme o código apresentado no desenvolvimento, basta chamar o método passando como parâmetro os dois sinais que estão sendo correlacionados, para se encontrar a diferença em *frames* destes sinais.

Para a triangulação, basta passar as diferenças em tempo como parâmetro e o modelo de captura (posição dos microfones) seguindo algumas regras: o algoritmo se baseia no modelo que representa um triângulo equilátero, então o triângulo pode ter o tamanho que for, mas deve ser equilátero, além disto, uma das bases deste modelo deve estar posicionadas em paralelo ao eixo das abscissas.

3.4 RESULTADOS E DISCUSSÃO

Os resultados serão divididos em duas seções, resultados da correlação cruzada e resultados da triangulação.

3.4.1 Resultados da correlação cruzada

Aqui é apresentado alguns sinais de teste para validar a implementação da correlação cruzada. O algoritmo implementado tem a complexidade de $O(n^2)$.

No Quadro 22 é apresentado como exemplo dois vetores representando sinais hipotéticos, que claramente estão a 1 *frame* de distância um do outro.

Sinal hipotético	Operação	Resultado
$a = \{ 1, 2, 3, 4, 5, 6, 7 \}$	CrossCorrelation (a, b)	-1
$b = \{ 2, 3, 4, 5, 6, 7, 0 \}$	CrossCorrelation(b, a)	1

Quadro 22 - Teste de correlação cruzada em sinais hipotéticos.

O resultado *-1*, representado pela primeira linha do Quadro 22, significa que o sinal hipotético *a* deve variar 1 *frame* para a esquerda para atingir o instante de maior correlação entre os sinais *a* e *b*.

A segunda linha do Quadro 22, aplica a mesma operação entre os sinais, mas invertendo o sinal base de comparação, mudando a ordem dos parâmetros, para que seja o sinal. Assim o resultado l representa que o sinal b deve variar 1 *frame* para a direita para atingir o instante de maior correlação entre os sinais b e a .

O teste representado no Quadro 23 demonstra que o momento de maior correlação entre os sinais, é o momento no qual o deslocamento do sinal de comparação resulta no maior *dot product*.

Sinal hipotético	Operação	Resultado
a = { 1, 2, 3, 4, 5, 6, 7 }	CrossCorrelation (a, b)	-3
b = { 4, 5, 6, 7, 1, 1, 1 }	CrossCorrelation(b, a)	3

Quadro 23 - Teste de correlação cruzada em sinais hipotéticos.

No Quadro 23 é claramente visível que quando a parte {4, 5, 6, 7} dos sinais estiverem alinhados é o momento onde a operação *dot product* (ver seção 2.1.2.2.1) será maior. O resultado -3, significa que o sinal a deve variar para a esquerda para ter o maior *dot product*, assim como o segundo resultado mostra que b deve variar 3 frames para a direita.

Para os testes reais, foi utilizado o modelo representado na Figura 20.

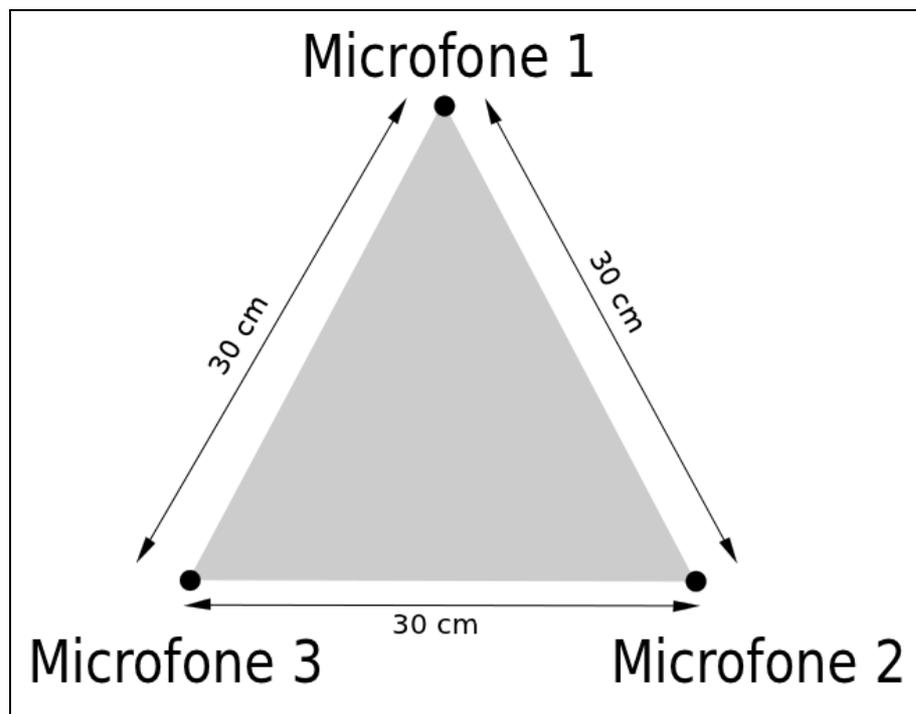


Figura 20 - Representação das interseções encontradas e a fonte sonora

A partir da Figura 20 tem-se a representação do modelo de captação, definido por três microfones, com uma distância entre si de 30cm formando um triângulo equilátero, com os três canais do microfones sendo capturados simultaneamente em tempo real com uma taxa de captura de 44.100 Hz.

O Quadro 24 mostra as entradas de sinais e as correlações cruzadas e o máximo de

correlação fisicamente possível.

Entrada(x, y)	Microfones envolvidos	Frames de diferença	Máximo de diferença de frames possível
(0, 40)	1 e 2	-29	38
(0, 40)	1 e 3	-29	38
(0, 40)	2 e 3	-2	38
(35, 35)	1 e 2	-4	38
(35, 35)	1 e 3	-426	38
(35, 35)	2 e 3	423	38

Quadro 24 - Correlação cruzada dos sinais, em relação a posição sonora conhecida.

A coluna “máximo de diferença possível” mostra a maior quantidade de *frames* fisicamente possível, considerando a distância entre os microfones e a velocidade do som.

De acordo com o Quadro 24 é possível verificar que há algumas correlações acima do possível, estes resultados estão ressaltados com o fundo azul. Isto significa que estas entradas, não serviriam para fazer triangulação, porque simplesmente não fazem sentido. Esses vários “erros” indicam algum tipo de problema na própria gravação dos sons, sendo que conforme abordado na seção anterior, os microfones utilizados são direcionais, indicados para entrevistas e gravação em um ângulo de 60° a frente do microfone (conforme especificação). E como o modelo utilizado é composto por três microfones, os três estão virados em direções opostas. Portanto no momento da gravação quando um som é emitido, os microfones virados para o lado oposto, não irão gravar o sinal imediatamente, além disto, podem gravar muito mais os sons ecoados e ruídos. Portanto as entradas gravadas não são confiáveis para uso.

3.4.2 Resultados da triangulação

Como as entradas dos microfones não são confiáveis para uso, foi desenvolvido um programa de teste que simula o disparo de um sinal na velocidade do som, desta forma como o ambiente é controlado, os resultados para avaliar a técnica são mais confiáveis.

O objetivo deste trabalho é principalmente desenvolver e validar uma técnica para triangular a fonte sonora, os testes serão focados na simulação de disparos de sinais e na detecção destas fontes sonoras simuladas.

O testes representados nas Figuras: 21, 22 e 23, contém a aplicação desenvolvida com uma representação gráfica dos resultados e as coordenadas da fonte sonora real e a fonte

sonora estimada. Essas figuras representam um caso de teste cada: os eixos x e y do plano cartesiano, modelo de captura representados pelo triângulo, retas e intersecções das retas, área de corte de intersecções, fonte sonora real e fonte sonora estimada. O modelo virtual de captura é um triângulo equilátero conforme os requisitos do algoritmo desenvolvido, com os lados (distância entre microfones) medindo 500 cm. O círculo em vermelho posicionado no centro, sobrepondo o modelo de captura representado nas figuras é a área de corte de intersecções. As intersecções são representadas em azul e são as intersecções das retas representadas em cinza.

A Figura 21 representa um caso no qual a fonte sonora estimada é aproximada a fonte sonora real.

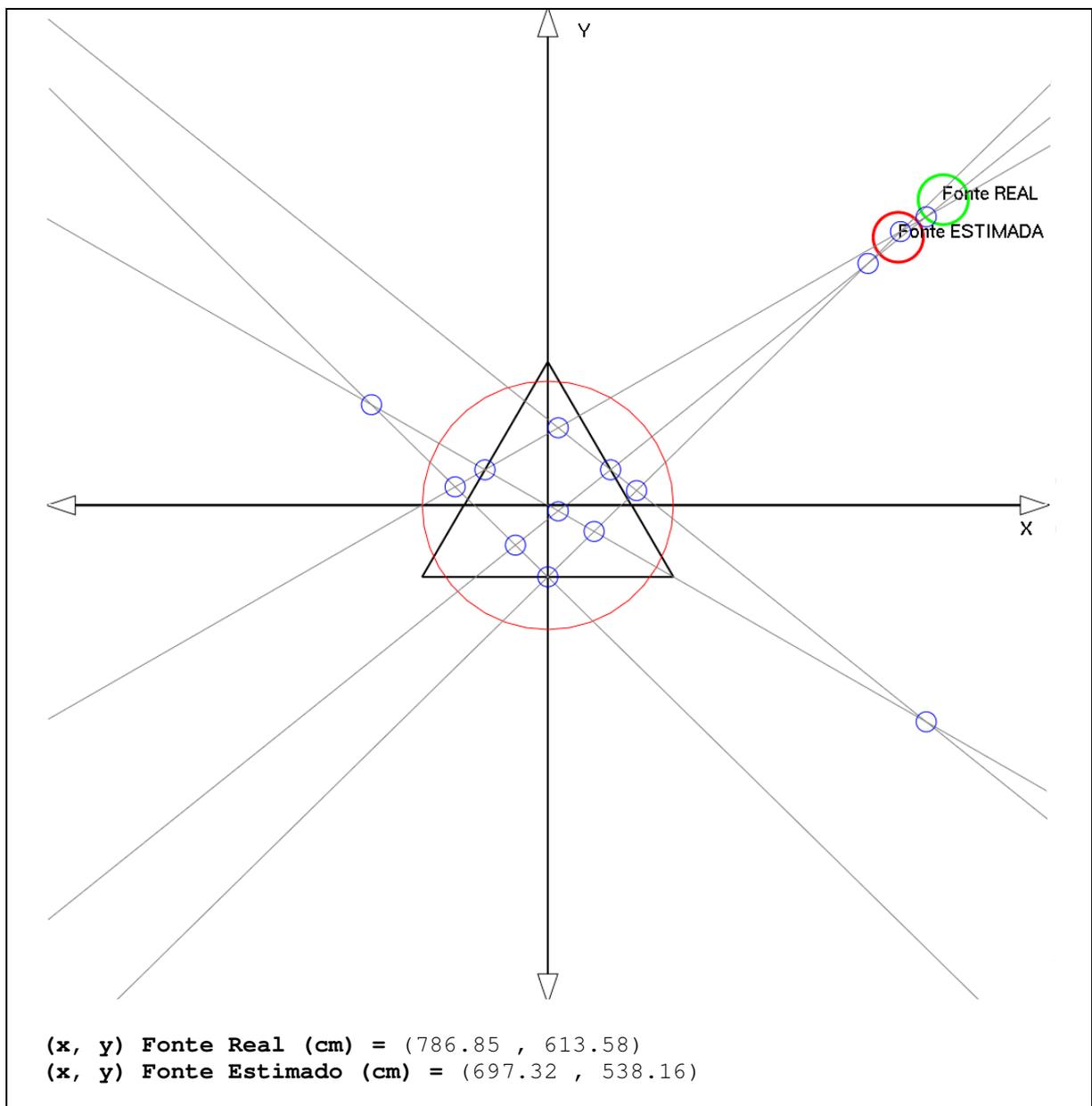


Figura 21 - Fonte sonora estimada com erro tolerável

padrão de casos de teste no qual as intersecções encontradas estejam próximas a fonte sonora real ou não. Este caso também representa que a área de corte está sendo considerada mediante precisão da função hiperbólica utilizada de Goodridge (1997), e como não há pelo menos três intersecções fora da área de corte, é considerado como um caso que não foi possível estimar.

Já na Figura 23 é representado um caso de teste no qual a fonte sonora estimada não está aproximada da fonte sonora real.

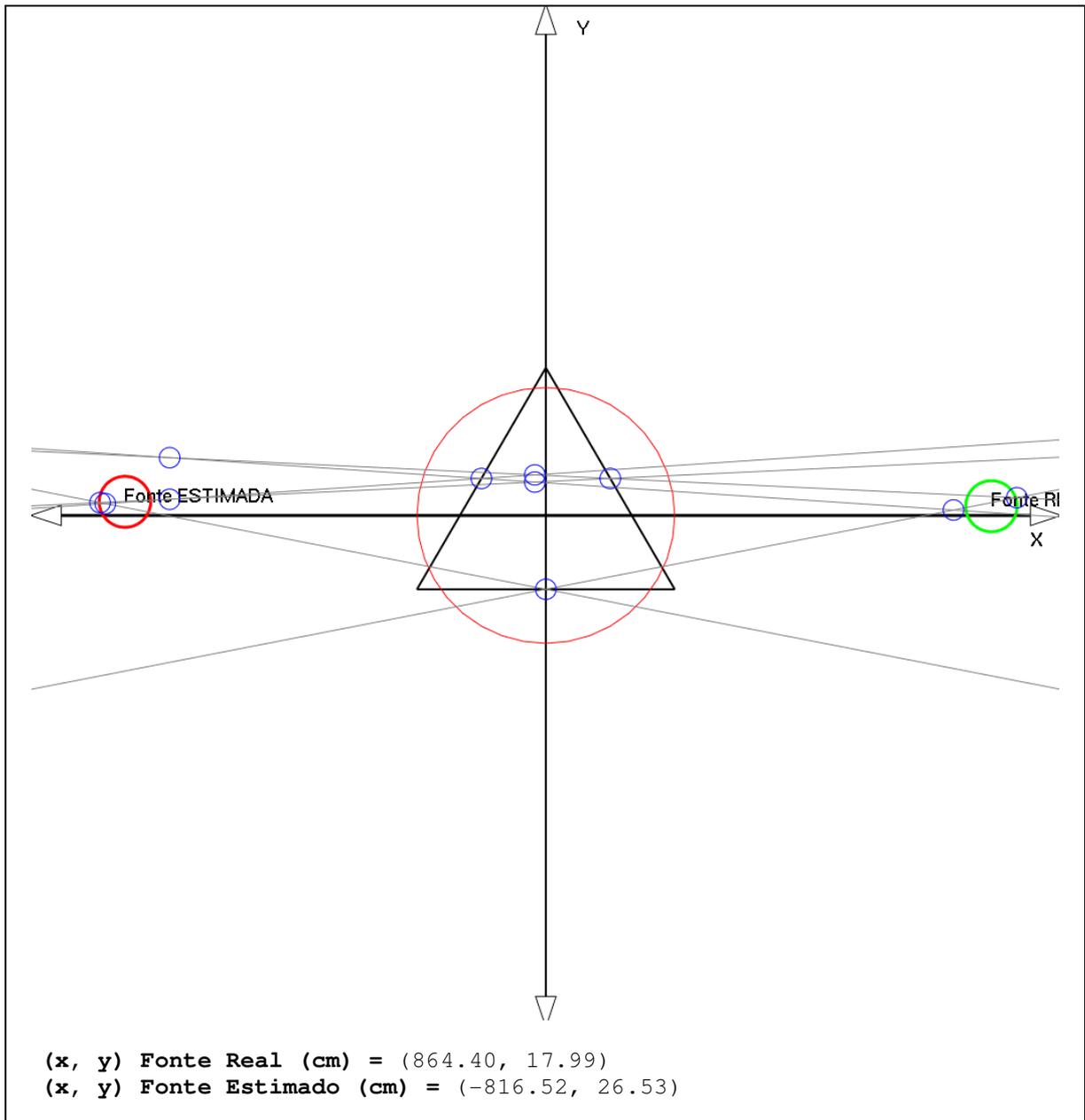


Figura 23 - Fonte sonora estimada com erro não tolerável

Em casos semelhantes à Figura 23, no qual a fonte sonora real está próxima aos eixos do plano cartesiano, percebeu-se que a fonte sonora estimada tem erros muito grandes, pequenas variações nos cálculos para encontrar os ângulos, atrasos, podem gerar grandes variações ao encontrar as intersecções das retas geradas.

4 CONCLUSÕES

Como pode ser observado neste trabalho, é possível localizar a fonte sonora por meio de captura dos sons por algum modelo e dispositivos adequados. No decorrer do desenvolvimento deste trabalho, não houve preocupação com os dispositivos de captura, pois as fontes pesquisadas preocupavam-se principalmente com os modelos matemáticos e as técnicas envolvidas, mas os dispositivos de captura devem ser considerados.

Para usar correlação cruzada, os dispositivos devem ter o máximo de precisão possível, e não podem ser direcionais, pois este tipo de trabalho envolve que os sensores estejam “abertos” a todas as direções.

O método de correlação cruzada mostrou-se precisa nos testes não envolvendo ambientes, como as fontes pesquisadas já confirmavam, no entanto, há outros meios de correlação que consideram ruídos e também poderiam ser utilizados.

Além disto, não se preocupou com o tempo de processamento da técnica de correlação cruzada, que no decorrer do processo inteiro para localizar a fonte sonora, é a parte que mais consome processamento, então para uma aplicação real, este método deve ser considerado, em questões de desempenho e ruídos.

O modelo para localizar a fonte sonora, após obter os parâmetros, prova-se funcional, mas não para uma aplicação real. Computacionalmente, os cálculos envolvidos vão acumulando muitos erros, no qual o método utilizado não considera. Desta forma, ao aprimorar os algoritmos para localização de fonte sonora, devem-se considerar técnicas que prevêm erros e formas de amenizar estes erros.

Este trabalho, serve como uma boa introdução em processamentos de sinais, algumas técnicas utilizadas, bem como base matemática que normalmente são utilizados em outras técnicas de triangulação. Também mostra vários problemas que provavelmente se encontram em qualquer tipo de aplicação nesse sentido.

Além disso, com a base de processamento de sinais deste trabalho, também foram desenvolvidos os algoritmos que fizeram a separação dos sinais de entrada para cada som de teste utilizado, pois mostrou-se muito mais produtivo do que utilizar alguma ferramenta existente de som, no entanto foram desenvolvidos exclusivamente para este trabalho.

Em relação aos objetivos, inicialmente ao que o trabalho propunha-se a detectar uma fonte sonora em um espaço 3D e com apenas dois microfones, no qual foi cortado do escopo do trabalho. As técnicas existentes para atender este objetivo com apenas dois microfones são

de uma complexidade muito grande, no qual é necessária uma base matemática e de processamento de sinais muito maior e mesmo tentando utilizar de três à quatro microfones, a complexidade ainda é muito alta para o tempo e conhecimento disponíveis para o desenvolvimento deste trabalho.

4.1 EXTENSÕES

Algumas sugestões podem se basear nos problemas e limitações atuais deste trabalho.

Uma boa continuação deste trabalho seria o processamento em tempo real.

O algoritmo de correlação cruzada atual é baseado no *dot product* onde a ordem de complexidade é de $O(n^2)$. Existem algoritmos melhores para correlação cruzada que podem diminuir o algoritmo para $O(n \log n)$, no qual já seria um ganho realmente considerável. Estes algoritmos tem como base a transformada de Fourier, no qual trabalha no domínio de frequência, enquanto os algoritmos convencionais utilizados trabalham no domínio de tempo. Em conjunto com estes algoritmos é possível utilizar de programação dinâmica e algumas heurísticas para otimizar ainda mais o algoritmo.

Uma outra sugestão seria incluir a dimensão 3d para estimar a posição sonora em um espaço e não apenas no plano, existem técnicas utilizadas fortemente por aeroportos para descobrir o posicionamento dos aviões baseado na emissão desses sinais, no qual pode ser adaptado para este trabalho, sendo que o som também é um tipo de sinal, a técnica utilizada geralmente é a técnica de multilateração no qual é baseado em interseções de hipérbolas formadas pelos sinais, no entanto esta técnica é bem complexa.

É possível também melhorar o algoritmo de correlação cruzada para utilizar um algoritmo de correlação com maior tolerância a ruídos no som como o GCC-PHAT (*Generalized Cross Correlation – Phase Transform*), que também é muito complexo.

Conforme mencionado anteriormente, este algoritmo é possível de ser utilizado para detecção sonora dentro de água, que no caso implicam no mar ou lagoas, com a inclusão do algoritmo de cálculo de velocidade do som para dentro da água ou um algoritmo genérico de cálculo parametrizável da velocidade som, para poder ser utilizado em mais situações. Claro que mudando a aplicação muda-se o hardware necessário para captação dos sinais.

Baseado na sugestão anterior este algoritmo pode ser estendido ou modificado para tratar qualquer tipo de aplicação que envolva sinais, sendo necessário hardwares para

captação destes sinais nas aplicações desejadas, tais como posicionamento de celulares, rádio, ondas sísmicas (ponto de origem de terremotos medindo com sismógrafos), posicionamento por imagem (transformando a imagem para o domínio de frequência), enfim qualquer tipo de sinais digitais que tenham os equipamentos certos para captura.

O principal foco da continuação deste trabalho seria a técnica de triangulação, para melhorar a precisão e as estimativas obtidas. Uma sugestão que se verificou interessante seria a técnica de multilateração, que também são utilizados em aeroportos nas torres de controle para triangulação de sinais emitidos pelos aviões.

Cada uma destas sugestões geram um trabalho com complexidade e custos provavelmente maiores do que este, possibilitando muito mais aplicações.

REFERÊNCIAS BIBLIOGRÁFICAS

- BALCONI, Mateus R. **Modelo triaural para encontrar a posição de uma fonte sonora**. 2008. 39 f. Trabalho de Conclusão de Curso (Bacharelado em Ciências da Computação) - Departamento de Computação, Universidade Estadual de Londrina. Londrina. Disponível em: <<http://www2.dc.uel.br/nourau/document/?down=677>>. Acesso em: 17 nov. 2010.
- BLAUERT, Jens. **Spatial hearing: psychophysics of human sound localization**. 2nd ed. USA: MIT Press, 1996.
- BRESLOFF, Furst I. et. al. Interaural time coincidence detectors are present at birth: evidence from binaural interaction. **Hearing Research**, Tel-Aviv, v. 187, n. 1, p. 63-72, Jan. 2004. Disponível em: <<http://www.sciencedirect.com/science/article/B6T73-4B2CRDK-1/2/5bc7ef033d0b6d4b9acb17d61dd3ac30>>. Acesso em: 17 nov. 2010.
- OZAN, Ariadine C. **Ensino fundamental - geometria: conceitos básicos**. [S.l.], 2005. Disponível em: <<http://pessoal.sercomtel.com.br/matematica/fundam/geometria/geo-basico.htm>>. Acesso em: 12 de jun. 2011.
- FORNARI, José; MAIA JUNIOR, Adolfo; MANZOLLI, Jônatas. A síntese evolutiva guiada pela espacialização sonora. In: CONGRESSO DA ASSOCIAÇÃO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM MÚSICA, 16. 2006, Brasília. **Anais...** Brasília: ANPPOM, 2006. p. 270-284. Disponível em: <http://www.anppom.com.br/anais/anaiscongresso_anppom_2006/CDROM/COM/03_Com_InfMus/sessao01/03COM_InfMus_0105-036.pdf>. Acesso em: 17 nov. 2010.
- GOODRIDGE, Steven G. **Multimedia sensor fusion for intelligent camera control and human-computer interaction**. Raleigh, 1997. Disponível em: <<http://www.ise.ncsu.edu/kay/msf/sound.htm>>. Acesso em: 02 jun. 2011.
- HARTMANN, William M. **How we localize sound**. [S.l.], 1999. Disponível em: <<http://www.aip.org/pt/nov99/locsound.html>>. Acesso em: 09 nov. 2010.
- LYON, Dick. **Head-related transfers functions**. Silicon Valley, 2010. Disponível em: <http://en.wikipedia.org/wiki/Head-related_transfer_function>. Acesso em: 20 de set. 2010.
- LYON, Douglas. **The discrete fourier transform, part 6: cross-correlation**. [S.l.], 2009. Disponível em: <<http://www.scribd.com/doc/25334405/The-Discrete-Fourier-Transform-Part-6-Cross-Correlation>>. Acesso em: 21 maio. 2011.
- MACKENSEN, Philip. **Auditive localization: head movements, an additional cue in localization**. 2004. 101 f. Doktorarbeit (Diplom Physiker, Magister) - Von der Fakultät I Geisteswissenschaften, Technischen Universität Berlin, Berlin. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.131.6323&rep=rep1&type=pdf>>. Acesso em: 17 nov. 2010.

OLIVEIRA, Aline C. et al. Localização de fontes sonoras: a importância das diferenças dos limiares auditivos interaurais. **Revista Sociedade Brasileira de Fonoaudiologia**, [S. l.], v. 13, n. 1, p. 7-11, mar 2008. Disponível em: <<http://www.scielo.br/pdf/rsbf/v13n1/04.pdf>> Acesso em: 17 nov. 2010.

PUNSKAYA, Elena; **Introduction to digital signal processing (DSP)**. [S.l.], 2009. Disponível em: <http://www-sigproc.eng.cam.ac.uk/~op205/3F3_1_Introduction_to_DSP.pdf>. Acesso em: 12 de jun. 2011.

SCHIESSER, Georg. **GNU/Linux based TDOA localization using COTS hardware**. Mistellbach, 2009. Disponível em: <<http://www.opentech.at/papers/tdoa.pdf>>. Acesso em: 22 maio. 2011.

SHINN-CUNNINGHAM, Barbara G.; KOPCO, Norbert.; MARTIN, TaraJ. **Localizing nearby sound sources in a classroom**: binaural room impulse responses. Massachusetts, 2005. Disponível em: <http://cns.bu.edu/~shinn/pages/pdf/JASA2005_BRIRs.pdf>. Acesso em: 18 set. 2010.

SIDDHANT, Joseph. **Multilateration**. [S.l.], 2011. Disponível em: <<http://en.wikipedia.org/wiki/Multilateration>>. Acesso em: 26 de maio. 2011.

VIEIRA, José M. N. et al. GoodEars: robot orientado por farol acústico. In: JORNADAS HISPANO-LUSAS DE INGENIERÍA ELÉTRICA, 7., 2001. Madrid. **Anais...** Madrid: [s.n.], 2001. p. 242-247. Disponível em: <<http://www.ieeta.pt/~vieira/Curriculum/papers/Vieira2001.pdf>>. Acesso em: 17 nov. 2010.

WANG, Guy J. **Computational auditory scene analysis**: principles, algorithms and applications. [S.l.]: IEEE Press Wiley Interscience, 1990.

WEISSTEIN, Eric W. **MathWorld - a wolfram web resource**. [S. l.], 2010. Disponível em: <<http://mathworld.wolfram.com> >. Acesso em: 02 jun. 2011.

ZHANG, Kazuaki. Estimation of sound source direction using a binaural model. **International Journal of Innovative Computing, Information and Control**, Toyohashi, v. 3, n. 3, p. 551–564, jun. 2007. Disponível em: <<http://www.ijicic.org/06-127-1.pdf>>. Acesso em: 17 nov. 2010.