

INSTITUTO TECNOLÓGICO DE AERONÁUTICA



Rahyan Azin

SÍNTESE DE SONS ESPACIAIS

Trabalho de Graduação
2019

Curso de Engenharia da Computação

Rahyan Azin

SÍNTESE DE SONS ESPACIAIS

Orientador

Prof. Manish Sharma (ITA)

ENGENHARIA DA COMPUTAÇÃO

SÃO JOSÉ DOS CAMPOS
INSTITUTO TECNOLÓGICO DE AERONÁUTICA

Dados Internacionais de Catalogação-na-Publicação (CIP)
Divisão de Informação e Documentação

Azin, Rahyan
Síntese de Sons Espaciais / Rahyan Azin.
São José dos Campos, 2019.
35f.

Trabalho de Graduação – Curso de Engenharia da Computação– Instituto Tecnológico de Aeronáutica, 2019. Orientador: Prof. Manish Sharma.

I. Instituto Tecnológico de Aeronáutica. II. Título.

REFERÊNCIA BIBLIOGRÁFICA

AZIN, Rahyan. **Síntese de Sons Espaciais**. 2019. 35f. Trabalho de Conclusão de Curso (Graduação) – Instituto Tecnológico de Aeronáutica, São José dos Campos.

CESSÃO DE DIREITOS

NOME DO AUTOR: Rahyan Azin

TÍTULO DO TRABALHO: Síntese de Sons Espaciais.

TIPO DO TRABALHO/ANO: Trabalho de Conclusão de Curso (Graduação) / 2019

É concedida ao Instituto Tecnológico de Aeronáutica permissão para reproduzir cópias deste trabalho de graduação e para emprestar ou vender cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte deste trabalho de graduação pode ser reproduzida sem a autorização do autor.

Rahyan Azin
Rua H8A, 420
12.228-460 – São José dos Campos–SP

SÍNTESE DE SONS ESPACIAIS

Essa publicação foi aceita como Relatório Final de Trabalho de Graduação

Rahyan Azin
Autor

Manish Sharma (ITA)
Orientador

Prof. Dr. Inaldo Costa
Coordenador do Curso de Engenharia da Computação

São José dos Campos, 26 de Novembro de 2019.

Resumo

Este artigo aborda o desenvolvimento de um sintetizador de áudio binaural a partir de uma fonte monoaural colocada em uma determinada posição para uma pessoa posicionada no centro de um ambiente anecóico. Espera-se que o sintetizador simule as variações do nível de pressão em cada tímpano causadas pelo estímulo de uma onda sonora e sua interação com o corpo humano. Seguindo o Modelo Estrutural da Cabeça Esférica, será desenvolvida uma ferramenta de sintetização binaural. A ferramenta possuirá outras funcionalidades de análise e processamento de áudio.

Abstract

This paper addresses for the development of a binaural audio synthesizer from a monaural source virtually localized to a person positioned in the center of an anechoic environment. It is expected the synthesizer simulates the changes in pressure level for each tympanum caused by the stimulation of a sound wave and its interaction with the human body. Following the Spherical Head Structural Model, a binaural synthesis tool will be developed. The tool will have other audio analysis and digital signal processing features.

Lista de Figuras

FIGURA 1.1 – Regiões de Confusão Frente-costas	11
FIGURA 2.1 – À esquerda: estímulo. À direita: sensação	13
FIGURA 2.2 – Representação da onda sonora no meio atmosférico	14
FIGURA 2.3 – Sistema auditivo humano	14
FIGURA 2.4 – Representação do Sistema de Coordenadas Polares Interaurais	16
FIGURA 2.5 – Comportamento HRIR CIPIC	19
FIGURA 2.6 – Modelo Estrutural da Cabeça Esférica	23
FIGURA 2.7 – Modelo Estrutural da Orelha	24
FIGURA 2.8 – Modelo Estrutural da Orelha	24
FIGURA 2.9 – Modelo Estrutural da Orelha acoplado na Cabeça Esférica	25
FIGURA 4.1 – Sintetizador de Sons Espaciais	29
FIGURA 4.2 – Head Model ativado	29
FIGURA 4.3 – CIPIC Model ativado	30
FIGURA 4.4 – Controlador da fonte sonora	30
FIGURA 4.5 – Azimute Cabeça Esférica: comportamento esperado	31
FIGURA 4.6 – Azimute Cabeça Esférica: mau funcionamento	31
FIGURA 4.7 – Azimute CIPIC	32
FIGURA 4.8 – Cabeça Esférica: distância	33

Lista de Tabelas

TABELA 2.1 – Coeficiente de 2.10	21
--	----

Sumário

1	INTRODUÇÃO	10
1.1	Motivação	10
1.2	Proposta	10
1.3	Estado da Arte	10
2	CONTEXTO	13
2.1	Som: Estímulo vs Sensação	13
2.2	A Onda Sonora	13
2.3	Sistema Auditivo	13
2.4	Síntese Sonora Digital	14
2.5	Processamento de Sinais Digitais	14
2.6	Sistema de Coordenadas Polares Interaurais	15
2.7	Percepção Sonora	16
2.7.1	Localização Horizontal/Azimutal	16
2.7.2	Localização Vertical, Elevação	17
2.7.3	Distância	17
2.7.4	Impressões Acústicas do Ambiente: Externalização	18
2.7.5	Som Espacial através de Fones de Ouvidos	18
2.8	Bases de Dados CIPIC	19
2.9	Modelagem Estrutural	19
2.9.1	Modelo da Cabeça Esférica	20
2.9.2	Modelo do Boneco de Neve	23
2.9.3	Modelo da Orelha	24

3	METODOLOGIA	26
3.1	Desenvolvimento da Ferramenta	26
3.2	Modelo da Cabeça Esférica	26
3.3	Modelo CIPIC	26
3.4	Regiões do Espaço	26
4	RESULTADOS E ANÁLISES	28
4.1	Ambiente de Desenvolvimento	28
4.2	Ferramenta: Visão Geral	28
4.2.1	Spatial Sound	29
4.2.2	Synthesis	30
4.3	Análise de resultados	30
4.3.1	Direcção Azimutal	31
4.3.2	Distância	32
4.3.3	Direção Vertical (Elevação)	32
4.3.4	Conclusões	33
5	CONCLUSÕES GERAIS	34
	REFERÊNCIAS	35

1 Introdução

1.1 Motivação

A forma que consumimos mídia audiovisual tem mudado drasticamente ao longo dos anos. Não há muito tempo que nos comunicávamos através de cartas, enquanto hoje temos transmissões ao vivo de imagem e som em escala global. Com o avanço da tecnologia o conteúdo audiovisual tem crescido substancialmente. Para diversos fins, tentamos simular nossa realidade através de dispositivos eletrônicos e recentemente obtivemos grandes avanços. Possuímos óculos de realidade aumentada que nos proporcionam uma grande imersão em ambientes virtuais. Tal imersão está intimamente relacionada com a simulação fiel dos estímulos dos nossos órgãos sensoriais que nos fazem perceber o mundo o tal como é. Torna-se parte desse processo simular fielmente os estímulos do nosso sistema auditivo: as ondas sonoras que chegam ao nosso tímpano. Nesse contexto, propõe-se o desenvolvimento de uma ferramenta que nos permita simular e analisar as reações do nosso sistema auditivo estimulado por ondas sonoras geradas através de técnicas de síntese de áudio binaural.

1.2 Proposta

Propõe-se nesse artigo o desenvolvimento de uma ferramenta para síntese de áudio binaural a ser reproduzido em fones de ouvido. As entradas do modelo serão as coordenadas da fonte sonora. O usuário será capaz de escolher dentre algumas opções de osciladores, como também poderá ajustar sua frequência. A aplicação conterá, também, ferramentas de visualização do sinal da onda sonora e seu espectro de frequência em tempo real.

1.3 Estado da Arte

A habilidade do ser humano de estimar a posição de fontes de sonoras tem papel relevante em várias situações do cotidiano: detecção de perigo no ambiente (um animal

que está prestes a atacar ou um carro que passa despercebido), localização de alvos que não estão no campo de visão, entre outros. Além disso, a sensação das características espaciais do som tem um grande potencial de engajamento em ambientes de realidade virtual e jogos eletrônicos.

Dessa forma, muitos estudos foram realizados ao longo dos anos com a intenção de compreender quais são os fatores que permitem o ser humano a se capaz de perceber tais características direcionais do som. As ondas sonoras que ouvimos diariamente estão sujeitas a diversas transformações ao longo do caminho entre sua criação e chegada aos tímpanos. A abordagem que melhor interpreta essas transformações envolve o uso de Resposta ao Impulso Binaural (BRIR). Uma fonte sonora irradia um som, que sofre modificações temporais e espectrais causadas pelo ambiente e pelo contato com o corpo humano. As propriedades que envolvem o ambiente estão contidas Resposta ao Impulso da Sala (RIR), enquanto que a Resposta ao Impulso da Cabeça (HRIR) contém as propriedades relacionadas ao corpo do ouvinte (KLEINER M.; SVENSSON, 1993). A função de transferência relacionada à HRIR é chamada de Função de Transferência da Cabeça (HRTF). A percepção das características espaciais do som pelo sistema auditivo é dividida em quatro categorias: localização horizontal, localização vertical, estimativa de distância e externalização (STRUMILLO, 2011).

A diferença de tempo e diferença de intensidade interaural foram caracterizadas no contexto da Teoria Duplex de Localização de Lord Rayleigh (STRUTT J., 1904).

A movimentação da cabeça tem papel importante na localização do som. Se o ouvinte é incapaz de movimentar a cabeça, torna-se mais difícil diferenciar sons que vem de trás da cabeça devido a existência de regiões em que os valores ITD e ILD são constantes (1.1).

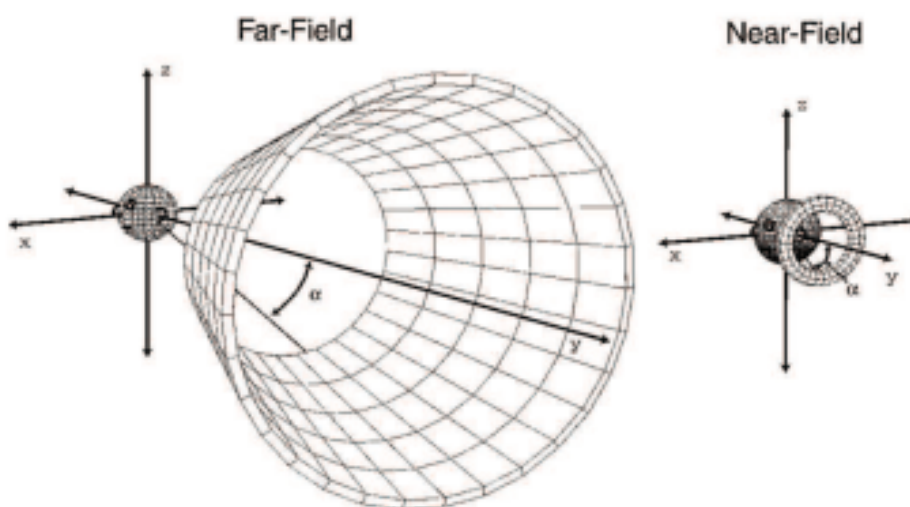


FIGURA 1.1 – Regiões de Confusão Frente-costas

A localização vertical da fonte sonora está relacionada fortemente com a presença da

orelha (GARDNER; GARDNER, 1973). Sabe-se também que é necessária a presença de componentes de frequências altas para uma localização vertical precisa do som (ASANO F.; SONE, 1990). Sob condições anecoicas, a percepção de distância está relacionada com a intensidade do sinal que chega aos ouvidos. Por outro lado, em um ambiente reverberante, a sensação de distância está relacionada com a relação de intensidade entre os sinais que chegam diretamente aos ouvidos com os refletidos pelo ambiente (KOLARIK A.; PARDHAN, 2013).

O termo externalização é usado para indicar quando um estímulo é percebido dentro da cabeça (lateralizado) ou fora da cabeça (localizado). Os fatores que contribuem para tal fenômeno são: modificações no espectro frequência do sinal, acústica do ambiente, movimentação da cabeça e ergonomia do sistema que transmite o sinal (LOOMIS J.; GOLLEDGE, 1999).

Muitas técnicas foram desenvolvidas ao longo dos anos para o design de HRTF e são classificadas em três grupos (BROWN; DUDA, 1998): design de filtros digitais através de técnicas de redes neurais, expansões em série para modelagem da HRTF através de dados experimentais coletados e modelos estruturais que isolam a contribuição dos efeitos causados pela cabeça, orelhas, ombros e tronco. Sendo a última, o foco desse artigo.

2 Contexto

2.1 Som: Estímulo vs Sensação

Do ponto de vista físico, define-se som como a oscilação de pressão (ou outra grandeza física, como velocidade, deslocamento e/ou velocidade de partículas) causada pelas interações de forças internas de um meio material. Entende-se por som, também, a sensação auditiva causa pelo estímulo da oscilação definida anteriormente.

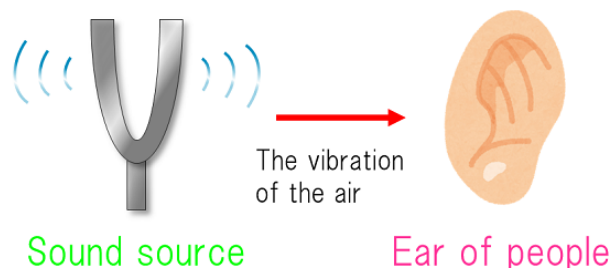


FIGURA 2.1 – À esquerda: estímulo. À direita: sensação

2.2 A Onda Sonora

Os sons que vemos na natureza são compostos por vários sinais, mas um som monotônico, representado por uma senóide, possui uma frequência de oscilação e uma amplitude. Sendo o meio de propagação um gás 2.2, os pontos de máxima compressão das partículas representam o pico da senóide e os pontos de máxima rarefação, os vales.

2.3 Sistema Auditivo

Nosso sistema auditivo 2.3 é responsável por captar as ondas sonoras e transformar as oscilações de pressão em impulsos nervosos e levá-los até o cérebro, que nos faz ter a sensação sonora.

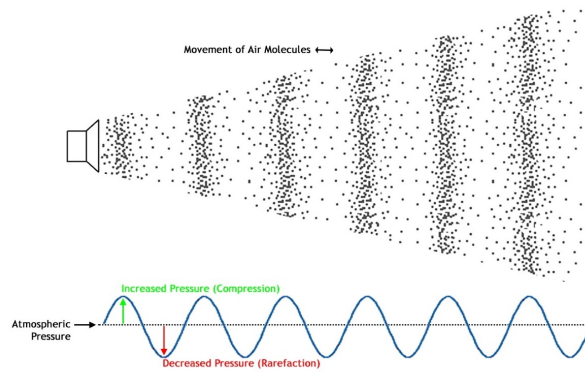


FIGURA 2.2 – Representação da onda sonora no meio atmosférico

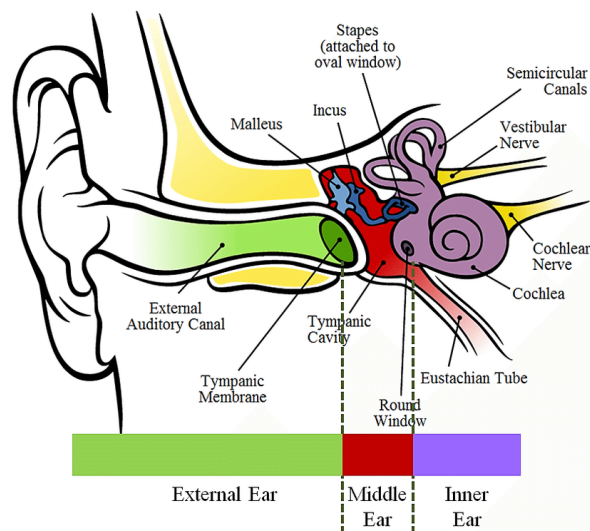


FIGURA 2.3 – Sistema auditivo humano

2.4 Síntese Sonora Digital

Entende-se por Síntese de Sons Digitais o conjunto de técnicas onde se definem sequências de sinais em que cada amostra será transmitida e reproduzida por uma placa de som a uma taxa amostragem definida, 44.1 kHz, por exemplo.

2.5 Processamento de Sinais Digitais

O processamento digital de sinais é um campo de estudo que define algoritmos computacionais com o objetivo de realizar operações específicas sobre dados digitais

Uma dessas operações é a filtragem digital de sinais, que dá origem ao Filtro Digital.

No contexto sinais que representam sons, um filtro digital é responsável por realizar modificações no domínio do tempo e frequência do sinal original. Um filtro linear e invariante no tempo (LTI) relaciona o sinal de saída com o sinal de entrada com uma equação

da seguinte forma:

$$y(n) = b_0x(n) + b_1x(n-1) + \dots + b_Mx(n-M) - a_1y(n-1) - \dots - a_Ny(n-N) \quad (2.1)$$

A função de transferência $H(z)$ de um filtro digital é a representação da relação entre os sinais de entrada e os de saída do sistema no domínio da frequência. Para um filtro LTI, temos:

$$H(z) = \frac{b_0 + b_1z^{-1} + \dots + b_Mz^{-M}}{1 + a_1z^{-1} + \dots + a_Nz^{-N}} \quad (2.2)$$

Um filtro de resposta ao impulso finita (FIR) é um tipo de filtro digital caracterizado por uma resposta ao impulso que se torna nula após um tempo finito.

A resposta ao impulso $h(n)$ é obtida quando o sinal de entrada é o impulso $\delta = [1, 0, 0, \dots]$. Para um FIR causal, quando a resposta ao impulso é nula para $n < 0$, sendo b_k a k -ésima amostra da resposta ao impulso, temos que:

$$h(n) = \begin{cases} 0, & n < 0 \\ b_n, & 0 \leq n \leq M \\ 0, & n > M \end{cases} \quad (2.3)$$

Se o sinal de entrada for definido como $x(n)$, temos que a saída do filtro é obtida através da convolução da resposta ao impulso com o sinal de entrada:

$$y(n) = b_0x[n] + b_1x[n-1] + \dots + b_Mx[n-M] \quad (2.4)$$

2.6 Sistema de Coordenadas Polares Interaurais

Para efeito de posicionamento da fonte sonora, precisamos definir um sistema de coordenadas que simplifique a representação no espaço. Desse modo, utiliza-se o Sistema de Coordenadas Polares Interaurais, onde temos um sistema de coordenadas esféricas em que o centro coincide com o centro da cabeça esférica.

O ponto em que azimutal e elevação são ambos 0° está posicionado diretamente de frente à cabeça. O ponto em que azimutal é 90° e elevação é 0° está posicionado diretamente à direita da cabeça. O ponto em que azimutal é 0° e elevação é 90° está posicionado diretamente à cima da cabeça. Abaixo uma representação gráfica de tal sistema.

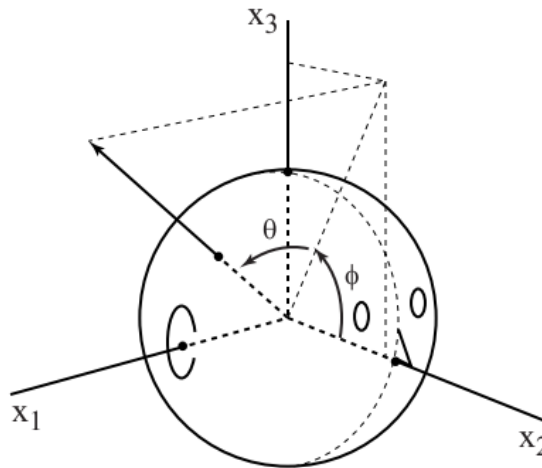


FIGURA 2.4 – Representação do Sistema de Coordenadas Polares Interaurais

2.7 Percepção Sonora

Os sons que ouvimos diariamente sofrem diversas transformações desde sua fonte até chegar aos nossos ouvidos. Um sinal acústico irradiado no espaço sofre modificações temporais e espectrais causada pela interação com ambiente e o corpo do ouvinte.

A percepção de eventos sonoros é dividida em três grupos elementares de sentidos:

- Temporal: ritmo, durabilidade e reverberação
- Qualidade: volume, altura e timbre
- Espacial: direção, distância e externalização.

A audição humana é capaz de perceber cada um desses elementos. Em particular, o sistema auditivo consegue interpretar a localização de uma fonte sonora em termos de direção e distância e também ter uma impressão acústica do ambiente em que está inserido. A posição e a movimentação do ouvinte têm papel determinante na percepção dessas características.

A percepção de características espaciais do som pode ser classificada como uma das quatro categorias descritas nas subseções seguintes.

2.7.1 Localização Horizontal/Azimutal

Habilidade de se perceber diferenças na direção de uma fonte sonora ao longo do plano horizontal. É caracterizada por uma diferenciação mínima do ângulo azimutal de aproximadamente 1° a 2° para uma grande variedade de sons. Para posições laterais, esse

valor cresce para 6° a 8° . As ondas sonoras percorrem distâncias diferentes até chegar aos ouvidos, o que acarreta em uma diferença temporal entre os sinais que chegam a cada canal auditivo, causando o que é conhecido como Interaural Time/Phase Difference (ITD/IPD).

A cabeça pode funcionar como um obstáculo ao interagir com a onda sonora. O que gera uma diferença de intensidade entre os sinais de cada canal auditivo. Conhecido como Interaural Level/Intensity Difference (ILD/IID).

As ondas de baixa frequência sofrem difração ao interagir com a cabeça, justamente por terem a ordem de grandeza do comprimento de onda próxima do raio da cabeça, nesse caso o efeito ITD é mais proeminente. À medida que a frequência aumenta, o efeito de difração deixa de existir e a cabeça se torna um obstáculo ao interagir com a onda, nesse caso o efeito ILD é mais proeminente.

2.7.2 Localização Vertical, Elevação

Habilidade de se perceber diferenças na direção de uma fonte sonora ao longo do plano vertical. No plano médio vertical (azimutal = 0°), a diferenciação mínima é 4° . Chegando a aproximadamente 17° para sons desconhecidos ao ouvinte.

A interação das ondas sonoras com as superfícies da orelha e canal auditivo gera reflexões e modos de ressonância que alteram o espectro de frequência no sinal que chega ao ouvinte. Tais efeitos têm papel fundamental na localização vertical da fonte sonora.

2.7.3 Distância

Um espaço anecoico é um local onde as ondas sonoras são absorvidas pelo ambiente, ou seja, o som não percorre múltiplos caminhos, pois não há reflexões. Num ambiente como esse a percepção de distância é correlacionada à intensidade do sinal, que é diretamente proporcional ao inverso do quadrado da distância.

Por outro lado, se o ambiente é reverberante, a diferenciação de distância é percebida através dos efeitos consequentes da interação das ondas sonoras com o espaço acústico.

Quando a fonte sonora está no Campo Distante (raio maior que $1 - 1.5m$), os efeitos que permitem a diferenciação direcional do som são quase independentes da distância. Entretanto, ao se aproximar do Campo Próximo (raio menor que $1 - 1.5m$), apesar de o ITD permanecer independente da distância, o efeito ILD é amplificado em todo espectro, principalmente para baixas frequências.

2.7.4 Impressões Acústicas do Ambiente: Externalização

Sinais estéreos ouvidos através de fones de ouvido podem ser percebidos como dentro ou fora da cabeça. No primeiro caso, as diferenças nos sinais são percebidas em apenas uma dimensão, através do eixo interaural (que passa pelos pontos de entrada de cada canal auditivo) e o som é tido como lateralizado. Já no outro caso, as diferenças de sinais são percebidas nas três dimensões do espaço e o som é tido como localizado, esse efeito é descrito como externalização do som.

As principais características que causam tal efeito são descritas a seguir:

- **Modificações no Espectro de Frequências:** as alterações que ocorrem no espectro decorrentes dos efeitos direcionais auxiliam na externalização sonora;
- **Acústica do ambiente:** a maioria das situações cotidianas ocorrem em ambientes com propriedades acústicas que alteram bastante os sinais que chegam aos ouvidos do ouvinte.
- **Movimentação da cabeça:** Sabe-se que a movimentação do cabeça tem papel fundamental na localização de sons para um ser humano, mas sua influência na externalização de sinais não está bem fundamentada atualmente.
- **Ergonomia do Sistema:** Se o ouvinte não tem ciência do sistema que está emitindo o sinal, ele irá perceber mais facilmente o efeito de externalização. Por exemplo, um par de fones minúsculos acoplados em cada canal auditivo tornaria a simulação do som localizado mais próxima da realidade.

2.7.5 Som Espacial através de Fones de Ouvidos

Para simular experiências sonoras espaciais realistas, deve-se, através de fones de ouvido, por exemplo, transmitir a cada tímpano o nível correto pressão.

A função de transferência da cabeça (HRTF) caracteriza o filtro que reproduz todas as transformações que uma onda sonora sofre desde a saída da fonte até a chegada aos tímpanos.

As transformações ocorridas desde a transmissão pelo fone de ouvido até a chegada no tímpano são caracterizadas pela função de transferência do fone de ouvido (HpRTF).

Fazendo a convolução de um sinal monofônico e anecoico com a resposta ao impulso equivalente à HRTF e, também, realizando a correta equalização dos fones de ouvido para compensar os efeitos causados pela HpRTF, pode-se realizar uma síntese que simula as características espaciais do som, especialmente quando movimentação da cabeça e propriedades acústicas do ambiente são consideradas.

2.8 Bases de Dados CIPIC

O Laboratório de Interface CIPIC na Universidade Davis da Califórnia realizou um experimento em um ambiente de alta resolução espacial com mais de 90 sujeitos (ALGAZI *et al.*, 2001). O microfone foi digitalizado em 44.1 kHz e para cada orelha foi capturada a resposta ao impulso em 1250 direções diferentes. A duração de cada HRIR foi de 200 amostras, correspondendo a uma duração de aproximadamente 4.5 ms. Foi-se disponibilizada publicamente uma base de dados com o resultado de 45 sujeitos desse experimento.

Dessa forma, a Base de Dados CIPIC foi utilizada nesse trabalho para fins comparativos com o modelo estrutural que foi desenvolvido.

A convolução dos componentes da resposta ao impulso do filtro com o sinal de entrada possui infinitos termos, entretanto, a HRIR possui um número finito de amostras. Entende-se, portanto, que o termos de maior grau do filtro tendem a zero. Podemos verificar tal comportamento através da figura abaixo (criado através da biblioteca Python Matplotlib), onde está representado graficamente a respostas ao impulso de todas as direções juntas.

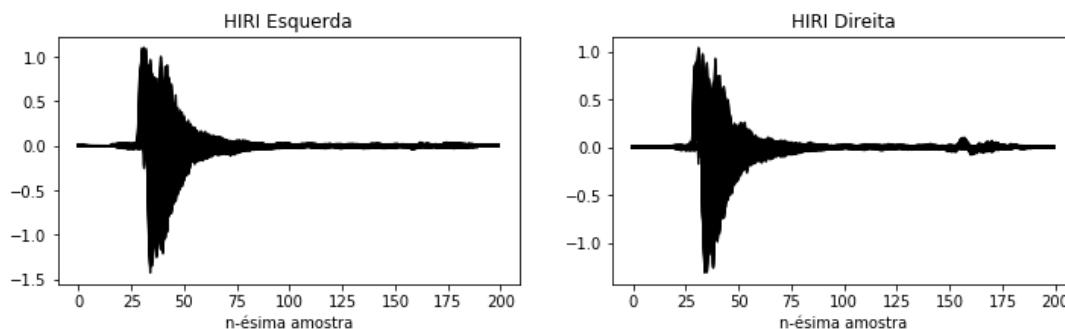


FIGURA 2.5 – Comportamento HRIR CIPIC

2.9 Modelagem Estrutural

Separando conceitualmente os efeitos da interação das ondas sonoras com corpo do ouvinte e associando cada efeito separadamente a um elemento de filtragem, a HRTF pode ser aproximada através de uma combinação dos filtros de cada efeito considerado.

Além disso, correlacionando os parâmetros de cada filtro com medidas antropométricas, pode-se, em princípio, obter um modelo estrutural das características espaciais do som personalizável para cada indivíduo.

2.9.1 Modelo da Cabeça Esférica

Assumindo a cabeça de um ser humano como uma esfera, define-se a Função de Transferência Esférica (STF) como a função que caracteriza os efeitos de tal modelo.

Em (Rabinowitz et al., 1993), a resposta relacionada a um ponto de observação fixo numa superfície esférica, é caracterizada através da seguinte função de transferência, baseada na fórmula de difração de Lord Rayleigh's (Strutt, 1904):

$$H(\rho, \mu, \theta_{inc}) = -\frac{\rho}{\mu} e^{-i\mu\rho} \sum_{m=0}^{\infty} (2m+1) P_m(\cos\theta_{inc}) \frac{h_m(\mu\rho)}{h'_m(\mu)}, \quad (2.5)$$

onde a é o raio da esfera, $p = r/a$ é a distância normalizada, θ_{inc} é o ângulo de incidência (o ângulo entre os raios que conectam o centro da esfera e a fonte sonora e o ponto de observação), e μ é a frequência normalizada, definida como:

$$\mu = f \frac{2\pi a}{c}, \quad (2.6)$$

onde c é a velocidade do som.

Pode-se avaliar a STF para $r \rightarrow \infty$, aplicando uma transformação bilinear na função de transferência descrita em 2.5:

$$H_{head}(z) = \frac{\frac{\beta + \alpha f_s}{\beta + f_s} + \frac{\beta - \alpha f_s}{\beta + f_s} z^{-1}}{1 + \frac{\beta + f_s}{\beta + f_s} z^{-1}}, \quad (2.7)$$

onde f_s é a frequência de amostragem, β é definido como $\beta = c/a$, e α é definido como:

$$\alpha(\theta_{inc}) = 1 + \frac{\alpha_{min}}{2} + (1 - \frac{\alpha_{min}}{2}) \cos(\frac{\theta_{inc}}{\theta_{min}} \pi), \quad (2.8)$$

θ_{inc} é o ângulo de incidência, e, assumindo o sistema de coordenadas polares interaurais, se relaciona com o azimuto θ como $\theta_{inc} = 90^\circ + \theta$, para o ouvido direito e $\theta_{inc} = 90^\circ - \theta$, para o ouvido esquerdo.

A função de transferência definida em 2.7 será referida como Função de Transferência do Campo Distante (FFTF).

Para analisar o efeito da distância no Campo Próximo, uma dada STF pode ser normalizada pelo Campo Distante correspondente, resultando em uma nova função de transferência, que será referida como Função de Transferência do Campo Próximo (NFTF):

$$H_{NF}(\rho, \mu, \theta_{inc}) = \frac{H(\rho, \mu, \theta_{inc})}{H(\infty, \mu, \theta_{inc})}, \quad (2.9)$$

TABELA 2.1 – Coeficiente de 2.10

θ_{inc}	p_{11}	p_{21}	q_{11}	q_{21}
0	12.97	-9.691	-1.136	0.219
10	13.19	234.2	18.48	-8.498
20	12.13	-11.17	-1.249	0.346
30	11.19	-9.035	-1.017	0.336
40	9.91	-7.866	-0.83	0.379
50	8.328	-7.416	-0.666	0.421
60	6.493	-7.312	-0.503	0.423
70	4.455	-7.278	-0.321	0.382
80	2.274	-7.291	-0.11	0.314
90	0.018	-7.484	-0.13	0.24
100	-2.242	-8.04	0.395	0.177
110	-4.433	-9.231	0.699	0.132
120	-6.488	-11.61	1.084	0.113
130	-8.342	-17.38	1.757	0.142
140	-9.93	-48.42	4.764	0.462
150	-11.29	9.149	-0.644	-0.138
160	-12.22	1.905	0.109	-0.082
170	-12.81	-0.748	0.386	-0.058
180	-13	-1.32	0.45	-0.055

Para determinar uma equação para a NFFT, primeiro é realizada uma análise para $\mu = 0$. Para cada ângulos de incidência de 0° a 180° , em passos de 10° , a 2.9 é calculada para um grande número distâncias diferentes.

O resultado é então modelado através uma função racional de segunda ordem. Para cada ângulo de incidência calculado, temos:

$$G_0(\theta_{inc}, \rho) = \frac{p_{11}(\theta_{inc})\rho + p_{21}(\theta_{inc})}{p^2 + q_{11}(\theta_{inc})\rho + q_{21}(\theta_{inc})}, \quad (2.10)$$

onde $\theta_{inc} = 0^\circ, 10^\circ, \dots, 180^\circ$.

Os coeficientes $p_{11}(\theta_{inc})$, $p_{21}(\theta_{inc})$, $q_{11}(\theta_{inc})$, $q_{21}(\theta_{inc})$ para cada um dos ângulos de incidência estão descritos na tabela abaixo:

Para analisar o comportamento da NFFT de acordo com a frequência, a NFFT é normalizada por $\frac{H_0}{2}$ e a frequência é normalizada por $\frac{f_c}{f_s}$.

$$H_{sh}(z) = 1 + \frac{H_0}{2} \left(1 - \frac{z^{-1} + a_c}{1 + a_c z^{-1}} \right), \quad (2.11)$$

$$a_c = \frac{V_0 \tan(\pi \frac{f_c}{f_s}) - 1}{V_0 \tan(\pi \frac{f_c}{f_s}) + 1}, V_0 = 10^{\frac{G_\infty}{20}}, \quad (2.12)$$

onde f_c é a frequência de amostragem e

$$G_\infty(\theta_{inc}, \rho) = \frac{p_{12}(\theta_{inc})\rho + p_{22}(\theta_{inc})}{p^2 + q_{12}(\theta_{inc})\rho + q_{22}(\theta_{inc})}, \quad (2.13)$$

$$f_c(\theta_{inc}, \rho) = \frac{p_{13}(\theta_{inc})p^2 + p_{23}(\theta_{inc})\rho + p_{33}(\theta_{inc})}{p^2 + q_{13}(\theta_{inc})\rho + q_{23}(\theta_{inc})}, \quad (2.14)$$

para cada $\theta_{inc} = 0^\circ, 10^\circ, \dots, 180^\circ$.

Os coeficientes para 2.13 e 2.14 estão descritos abaixo, respectivamente:

θ_{inc}	p_{12}	p_{22}	q_{12}	q_{22}
0	-4.391	2.123	-0.55	-0.061
10	-4.314	-2.782	0.59	-0.173
20	-4.18	4.224	-1.006	-0.021
30	-4.012	3.039	-0.563	-0.316
40	-3.874	-0.566	0.665	-1.129
50	-4.099	-34.74	11.39	-8.301
60	-3.868	3.271	-1.571	0.637
70	-5.021	0.023	-0.875	0.325
80	-6.724	-8.965	0.37	-0.083
90	-8.693	-58.38	5.446	-1.188
100	-11.17	11.47	-1.131	0.103
110	-12.08	8.716	-0.631	-0.12
120	-11.13	21.8	-2.009	0.098
130	-11.1	1.91	0.15	-0.401
140	-9.719	-0.043	0.243	-0.411
150	-8.417	-0.659	0.147	-0.344
160	-7.437	0.395	-0.178	-0.184
170	-6.783	2.662	-0.671	0.05
180	-6.584	3.387	-0.836	0.131

θ_{inc}	p_{13}	p_{23}	p_{33}	q_{13}	q_{23}
0	0.457	-0.668	0.174	-1.746	0.699
10	0.455	0.142	-0.115	-0.01	-0.348
20	-0.87	3.404	-1.699	7354	-5350
30	0.465	-0.913	0.437	-2.181	1.188
40	0.494	-0.669	0.658	-1.196	0.256
50	0.549	-1.208	2.02	-1.59	0.816
60	0.663	-1.756	6.815	-1.296	1.166
70	0.691	4.655	0.614	-0.889	0.76
80	3.507	55.09	589.3	29.23	59.51
90	-27.41	10336	16.818	1.945	1707
100	6.371	1.735	-9.389	-0.058	-1.118
110	7.032	40.88	-44.09	5.635	-6.18
120	7.092	23.86	-23.61	3.308	-3.392
130	7.463	102.8	-92.27	13.88	-12.67
140	7.453	-6.145	-1.809	-0.877	-0.19
150	8.101	-18.1	10.54	-2.23	1.295
160	8.702	-9.05	0.532	-0.96	-0.023
170	8.925	-9.03	0.285	-0.905	-0.079
180	9.317	-6.888	-2.082	-0.566	-0.398

Dessa forma, podemos definir uma estrutura geral para a STF ilustrada em 2.6.

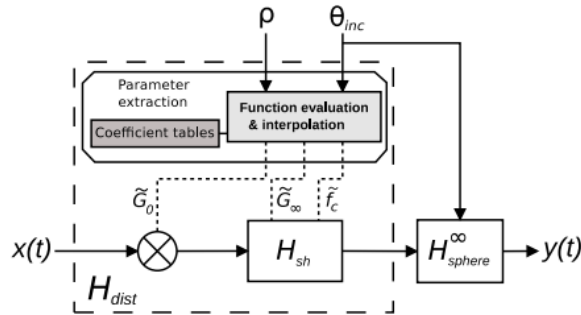


FIGURA 2.6 – Modelo Estrutural da Cabeça Esférica

Para se extrair resultados intermediários entre os valores discretizados, basta realizar uma interpolação entre os coeficientes das tabelas.

2.9.2 Modelo do Boneco de Neve

Um modelo que leva em consideração a cabeça e o tronco do ouvinte foi proposto em (Algazi et al., 2002b) como a superposição de duas esferas de raio a e b , separadas por

uma distância h , representando o tamanho do pescoço. A estrutura de tal modelo está ilustrada em 2.7.

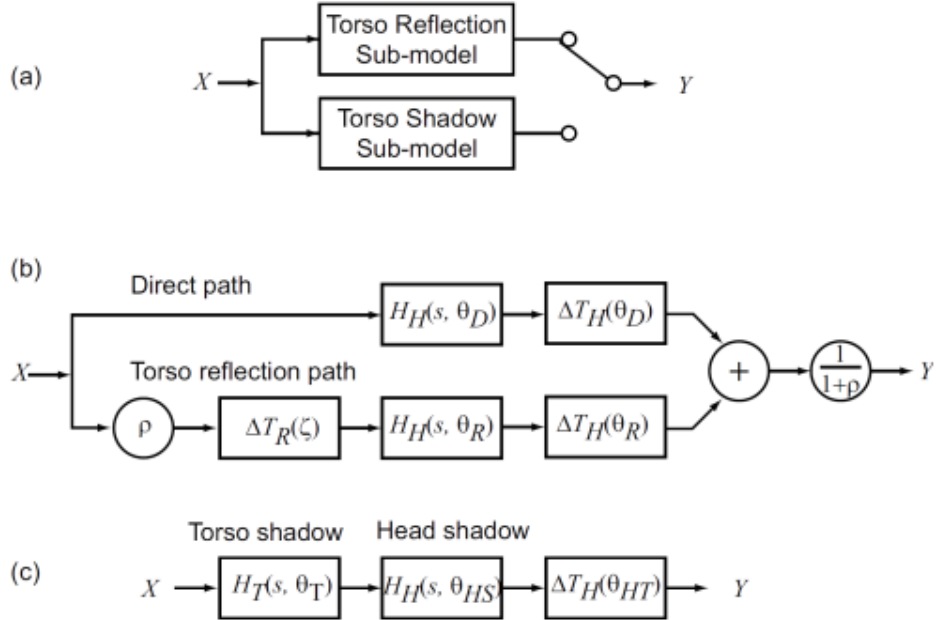


FIGURA 2.7 – Modelo Estrutural da Orelha

2.9.3 Modelo da Orelha

A seguir um modelo que engloba os efeitos causados pela Orelha (GERONAZZO, 2013), responsável dominante pela percepção da elevação do som.

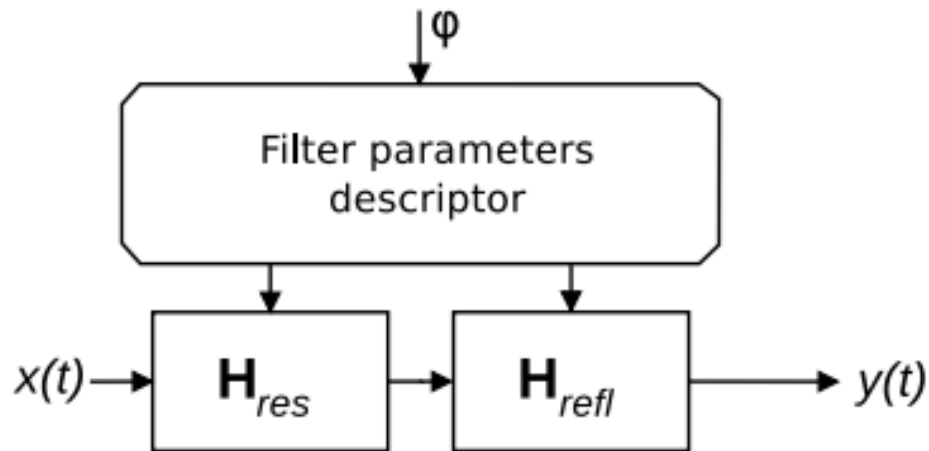


FIGURA 2.8 – Modelo Estrutural da Orelha

Nesse modelo, os efeitos causados pela orelha são divididos em dois blocos: efeitos de ressonância e efeitos de reflexão.

Os coeficientes que determinam os filtros de ressonância são bem parecidos para diferentes pessoas, dependem, portanto, apenas do grau de elevação. Já os coeficientes dos filtros de entalhe (que simulam os vales de causados pelas reflexões na orelha) variam muito de pessoas para pessoa.

A grande dificuldade para utilizar esse modelo é justamente implementar o algoritmo que determina os coeficientes de tais filtros. De qualquer maneira, o modelo acima poderia ser adicionado ao Modelo da Cabeça Esférica da seguinte forma:

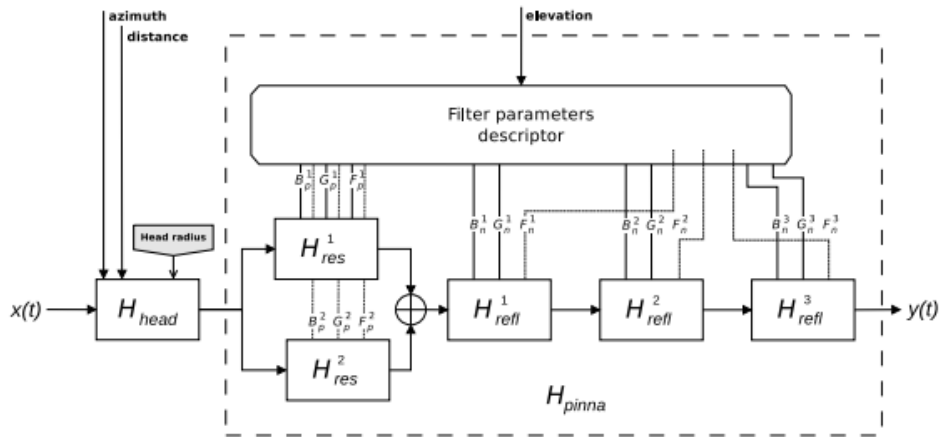


FIGURA 2.9 – Modelo Estrutural da Orelha acoplado na Cabeça Esférica

3 Metodologia

3.1 Desenvolvimento da Ferramenta

A IDE Qt Creator foi o ambiente escolhido para criar o sintetizador. Possui integração entre back e front end, tornando mais simples o desenvolvimento da interface gráfica e permite que o mecanismo de processamento de sinais seja feito em C++, linguagem padrão para desenvolvimento no contexto DSP. Além de ser Cross-Plataform, tornando o projeto mais escalável.

3.2 Modelo da Cabeça Esférica

Para a implementação do modelo da Cabeça Esférica foi aplicada a estrutura do filtro digital definida em (2.9.1) e os parâmetros foram calibrados como sugerido em (GERONAZZO, 2013) da seguinte forma: $\alpha_{min} = 0.1$ e $\theta_{min} = 180^\circ$ e $a = 8.75cm$.

O sinal de saída foi obtido através da equação descrita em 2.1

3.3 Modelo CIPIC

Também foi implementado um filtro de resposta ao impulso (FIR) utilizando a base de dados CIPIC.

O sinal de saída foi obtido através da equação descrita em 2.4.

3.4 Regiões do Espaço

Após a implementação dos modelos notou-se que os filtros paravam de funcionar para certas regiões do espaço. Para uma ferramenta inicial, decidiu-se, portanto, restringir as entradas aos pontos em que o modelo apresentava um comportamento aceitável (não

necessariamente de acordo com o esperado). Tais regiões são: -45° a 45° para o ângulo azimutal, 0° a 45° para elevação e 1 a 200 para ρ .

4 Resultados e Análises

4.1 Ambiente de Desenvolvimento

A ferramenta foi desenvolvida no Qt Creator, uma IDE que oferece desenvolvimento de aplicações multi-plataforma de maneira fácil e rápida, integrando a linguagem de programação compilada C++ com a linguagem de marcação QML. O programa utiliza a biblioteca RtAudio (<https://www.music.mcgill.ca/~gary/rtaudio/>), uma API para entrada/saída de áudio em tempo real que simplifica o processo de interação com o hardware de áudio do computador.

Para realizar uma análise no campo espectral do sinal transmitido para os fones de ouvido, foi utilizada uma outra biblioteca chamada fftw (<http://www.fftw.org>) que auxilia na computação da Transformada de Fourier Discreta. Para a implementação de um dos modelos, foi utilizado um banco de dados com arquivos de extensão .mat. Para simplificar o processo, foram desenvolvidos scripts em Python para visualização e tratamento de tais dados. O processo envolveu selecionar os dados pertinentes ao desenvolvimento do modelo e salvá-los na extensão .npy, que foram importados para o ambiente de desenvolvimento da aplicação através da bibliotec libnpy (<https://github.com/llohse/libnpy>).

4.2 Ferramenta: Visão Geral

Nessa seção iremos explicar as funcionalidades do Sintetizador. Abaixo uma foto do panorama geral da ferramenta. O código para o projeto encontra-se em <https://github.com/rahyanazin/SpatialSynthesis>

Pode-se observar que há três áreas na ferramenta, cada uma responsável por um aspecto diferente do sintetizador.

- Analysis (Left/Right): Visualização em tempo real do sinal e das componentes de frequência em cada lado do fone de ouvido.
- Spatial Sound: Visualização das funções de transferência; Seleção e controle dos

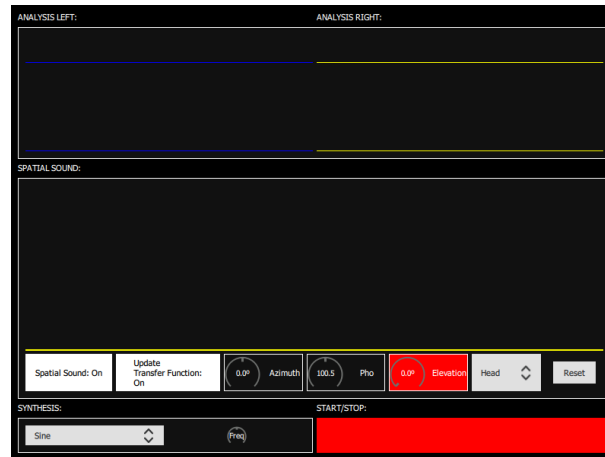


FIGURA 4.1 – Sintetizador de Sons Espaciais

modelos.

- Synthesis: Seleção e controle da fonte sonora; Botão liga/desliga.

4.2.1 Spatial Sound

Nessa seção há dois modos de operação: Head Model e CIPIC Model. Para cada caso temos funcionalidades diferentes.



FIGURA 4.2 – Head Model ativado

Quando ativamos o Head Model, estamos utilizando os filtros correspondentes ao modelo da cabeça esférica. Note que o botão que controla a elevação está vermelho, indicando que não altera o modelo. A um botão Reset, que coloca os valores iniciais do filtro para zero e deve ser utilizado quando o filtro parar de funcionar.

Quando ativamos o CIPIC Model, estamos utilizando a base de dados CIPIC para calcular a resposta ao impulso em um ponto específico do espaço. Note que o botão que controla a distância está vermelho, indicando que não altera o modelo. À direita do

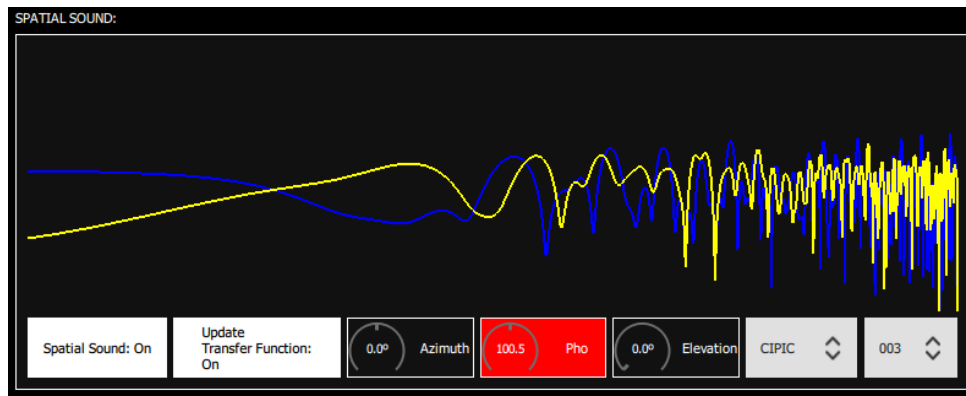


FIGURA 4.3 – CIPIC Model ativado

botão de selecionar o modelo, temos uma caixa para selecionar um dos 45 voluntários do experimento CIPIC, há nessa mesma caixa um opção com a média de todos os casos.

Nos dois modelos temos a presença de um botão para ligar e desligar o efeito do modelo. Temos também um botão que desliga a atualização do gráfico da função de transferência, essa funcionalidade é útil no modo CIPIC, pois para gerar o gráfico requer muito processamento e ao controlar as dimensões o programa fica bastante lento.

4.2.2 Synthesis

Nessa seção conseguimos ligar e desligar a passagem de sinal.

Também podemos escolher fonte sonora que iremos analisar:

- Sine ou Square: Nesse caso controlamos a frequência.
- Noise: Aqui controlamos a frequência de corte do filtro passa-baixo do ruído, como sua ressonância.



FIGURA 4.4 – Controlador da fonte sonora

4.3 Análise de resultados

Nessa seção será feita uma análise individual da performance da ferramenta para as direções disponíveis em cada modelo.

4.3.1 Direcção Azimutal

4.3.1.1 Cabeça Esférica

O modelo funcionou em boa parte das regiões como esperado. Ao deslocar o ângulo azimutal para esquerda, o sinal de saída do fone esquerdo era amplificado, enquanto o do lado direito era atenuado, indicando o efeito da ILD. Entretanto, o modelo demonstrou defeitos em regiões para $-7^\circ < \theta < 7^\circ$. Também, para certas frequências o filtro para de funcionar quando $\rho < 5$

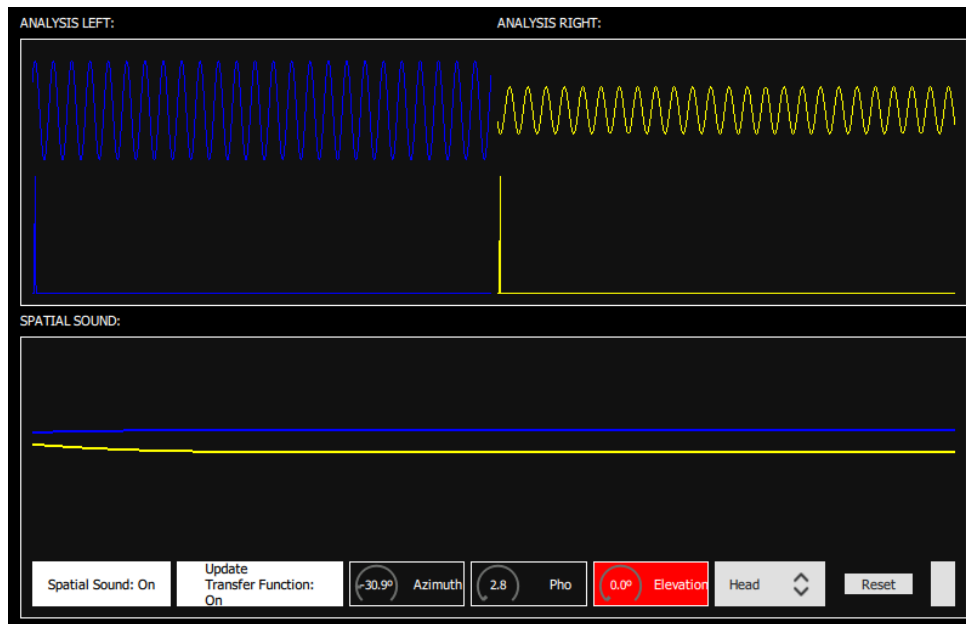


FIGURA 4.5 – Azimute Cabeça Esférica: comportando esperado

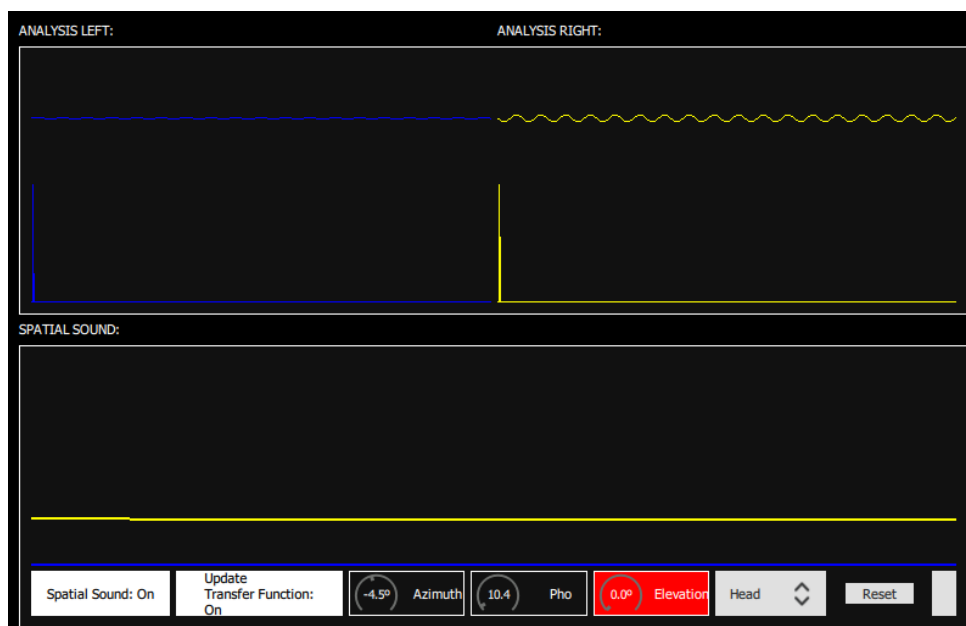


FIGURA 4.6 – Azimute Cabeça Esférica: mau funcionamento

4.3.1.2 Base de Dados CIPIC

Esse modelo apresentou resultado esperado para todas variações azimutais disponíveis. Porém, o resultado fica mais evidente com fontes sonoras ricas em vários componentes de frequência.

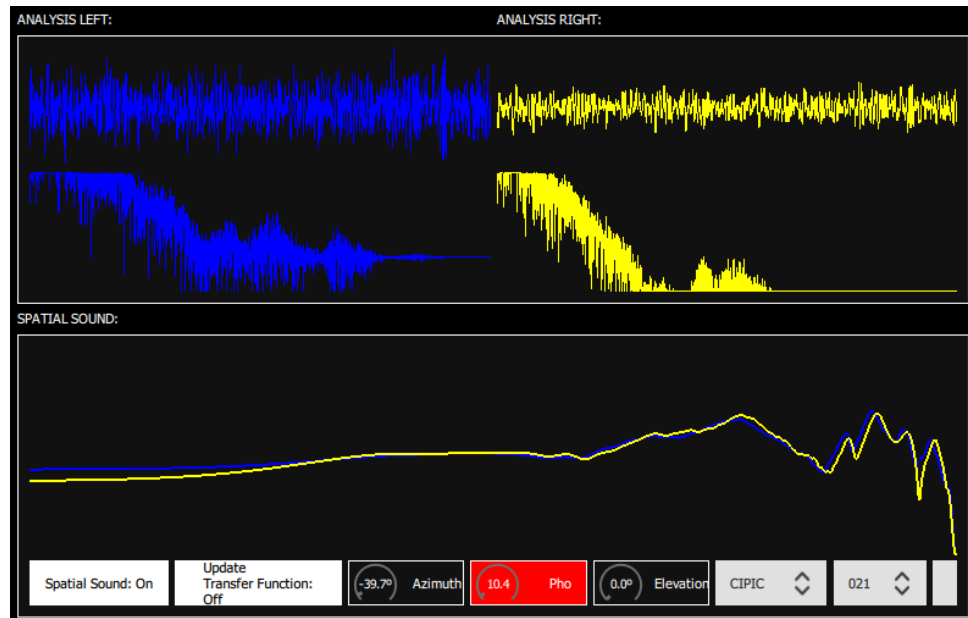


FIGURA 4.7 – Azimute CIPIC

4.3.2 Distância

4.3.2.1 Cabeça Esférica

A distância apresentou um bom comportamento, à medida que aumentava o sinal era atenuado. Para certas frequências o filtro para de funcionar quando $\rho < 5$

4.3.3 Direção Vertical (Elevação)

4.3.3.1 Base de Dados CIPIC

Finalmente, a elevação, a característica dimensional do som mais difícil de ser diferenciada. Alguns voluntários testaram e confirmaram perceber diferenças da elevação para algumas da HRTF's do modelo com dados do CIPIC.

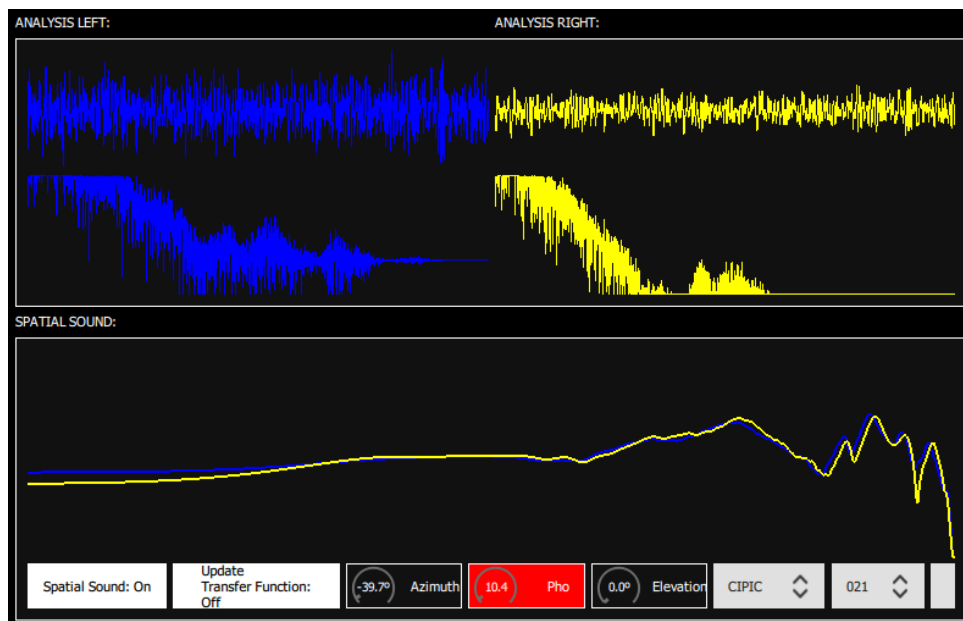


FIGURA 4.8 – Cabeça Esférica: distância

4.3.4 Conclusões

O modelo da Cabeça Esférica, apesar de apresentar regiões problemáticas, demonstrou um bom funcionamento. Talvez uma calibragem adequada dos parâmetros antropométricos torne a experiência mais próxima da realidade.

O modelo de resposta ao impulso utilizando o Banco de Dados CIPIC apresentou bom funcionamento para todas regiões, o que é de se esperar, uma vez que os dados foram obtidos em um experimento muito bem controlado.

5 Conclusões Gerais

Ao longo dos anos, muitos estudos a cerca das características espaciais do som foram realizados. Sabe-se que o sistema auditivo do ser humano é capaz de localizar a direção do som, como também reconhecer propriedades acústicas do ambiente em que está inserido. Muitos são os fatores que dão ao ser humano a capacidade de diferenciar essas características. A diferença de fase e intensidade dos sinais que chegam aos tímpanos auxilia na localização horizontal, a interação das ondas sonoras com a superfície da orelha é fundamental na localização da direção vertical e propriedades acústicas do ambiente nos dão a sensação de distância. A movimentação da cabeça nos permite diferenciar fontes sonoras que se encontram em pontos de confusão do cérebro. Nesse artigo foi implementado o Modelo Estrutural da cabeça para realizar a síntese sonora. O modelo apresentou bom funcionamento em muitas regiões. Os comportamentos não esperados em certas regiões podem ser associados tanto às simplificações e aproximações do modelo, como também a não existência da modelagem de todos os fatores que permite o sistema auditivo interpretar as características de uma onda sonora.

Como esperado, o modelo de resposta ao impulso utilizando o Banco de Dados CIPIC apresentou uma melhor representação da síntese binaural, uma vez que por sua própria natureza tem menos propensão ao erro e também porque utiliza dados coletados em um ambiente bem controlado de alta resolução acústica. As funcionalidades da ferramenta funcionam de acordo com a teoria. O processamento dos sinais ocorre de forma contínua e instantânea, menos para representação gráfica da função de transferência do modelo CIPIC, pois esse apresenta uma equação de diferenças de ordem 200, o que acarreta em uma taxa muita alta de cálculos por segundo.

Para trabalhos futuros, pode-se agregar ao modelo da Cabeça Esférica outras componentes que representam outras transformações que a onda sonora sofre desde de sua origem até chegar nos ouvidos, como os efeitos do tronco e ombros através do Modelo Boneco de Neve e também a contribuição da orelha com o Modelo da Orelha. Em relação às ferramentas de análise, pode-se implementar uma visualização da fase do sinal que chega em cada fone, que auxiliaria na verificação da ITD.

Referências

- ALGAZI, V. R.; DUDA, R. O.; THOMPSON, D. M. The cipic hrtf database. **IEEE Workshop on Applications of Signal Processing to Audio and Acoustic**, p. 4, 2001.
- ASANO F., S. Y.; SONE, T. Role of spectral cues in median plane localization. **The Journal of the Acoustical Society of America**, 1990.
- BROWN, C. P.; DUDA, R. O. A structural model for binaural sound synthesis. **IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING**, v. 10, n. 5, p. 476, 1998.
- GARDNER, M. B.; GARDNER, R. S. Problem of localization in the median plane: Effect of pinnae cavity occlusion. **J. Acoust. Spc. Am.**, 1973.
- GERONAZZO, M. **Mixed Structural Models For 3D Audio in Virtual Environments**. 242 p. Trabalho de Doutorado em Ciência da Tecnologia e Informação — Università Degli Studi Padova, 2013.
- KLEINER M., D. B.-I.; SVENSSON. Auralization - an overview. **Journal of the Audio Engineering Society**, 1993.
- KOLARIK A., C. S.; PARDHAN, S. Discrimination of virtual auditory distance using level and direct-to-reverberant ratio cues. **The Journal of the Acoustical Society of America**, 2013.
- LOOMIS J., K. R.; GOLLEDGE, R. Auditory distance perception in real, virtual and mixed environments. **Mixed Reality: Merging Real and Virtual Worlds.**, 1999.
- STRUMILLO, P. Advances in sound localization. **InTech**, 2011.
- STRUTT J., W. On the acoustic shadow of a sphere. **Phil. Trans.**, 1904.

FOLHA DE REGISTRO DO DOCUMENTO

1. CLASSIFICAÇÃO/TIPO TC	2. DATA 26 de Novembro de 2019	3. DOCUMENTO Nº DCTA/ITA/TC- 074/2019	4. Nº DE PÁGINAS 35
5. TÍTULO E SUBTÍTULO: Síntese de Sons Espaciais			
6. AUTOR(ES): Rahyan Azin			
7. INSTITUIÇÃO(ÕES)/ÓRGÃO(S) INTERNO(S)/DIVISÃO(ÕES): Instituto Tecnológico de Aeronáutica – ITA			
8. PALAVRAS-CHAVE SUGERIDAS PELO AUTOR: Desenvolvimento de software; Processamento digital de sinais; C mais mais (linguagem de programação); Sons musicais; Computação.			
9. PALAVRAS-CHAVE RESULTANTES DE INDEXAÇÃO: Desenvolvimento de software; Processamento digital de sinais; C mais mais (linguagem de programação); Sons musicais; Computação.			
10. APRESENTAÇÃO: (X) Nacional () Internacional ITA, São José dos Campos. Curso de Graduação em Engenharia da Computação. Orientador: Prof. Manish Sharma.			
11. RESUMO: <p>Este artigo aborda o desenvolvimento de um sintetizador de áudio binaural a partir de uma fonte monoaural colocada em uma determinada posição para uma pessoa posicionada no centro de um ambiente anecóico. Espera-se que o sintetizador simule as variações do nível de pressão em cada tímpano causadas pelo estímulo de uma onda sonora e sua interação com o corpo humano. Seguindo o Modelo Estrutural da Cabeça Esférica, será desenvolvida uma ferramenta de sintetização binaural. A ferramenta possuirá outras funcionalidades de análise e processamento de áudio.</p>			
12. GRAU DE SIGILO: (X) OSTENSIVO () RESERVADO () SECRETO			