Emilio Antonio Saievicz

Estudo da Aplicação do Áudio Binaural a uma Videoconferência

São José – SC
março / 2010

Emilio Antonio Saievicz

Estudo da Aplicação do Áudio Binaural a uma Videoconferência

Monografia apresentada à Coordenação do Curso Superior de Tecnologia em Sistemas de Telecomunicações do Instituto Federal de Santa Catarina para a obtenção do diploma de Tecnólogo em Sistemas de Telecomunicações.

Orientador:
Prof. Dr. Marcos Moecke

CURSO SUPERIOR DE TECNOLOGIA EM SISTEMAS DE TELECOMUNICAÇÕES
INSTITUTO FEDERAL DE SANTA CATARINA

São José – SC
março / 2010

Monografia sob o título “Estudo da Aplicação do Áudio Binaural a uma Videoconferência”, defendida por Emilio Antonio Saievicz e aprovada em 17 de março de 2010, em São José, Santa Catarina, pela banca examinadora assim constituída:

_____

Prof. Dr. Marcos Moecke
Orientador

_____

Prof. Dr. Emeson Ribeiro de Mello
Co-orientador

_____

Prof. Dra. Elen Macedo Lobato Merlin
IFSC

_____

Prof. Ms. Silviana Cirino
IFSC

Agradecimentos

Agradeço primeiramente ao professor Marcos Moecke, por ter me sugerido trabalhar com áudio binaural para meu trabalho de conclusão de curso, e não somente, a todo apoio que me forneceu durante a elaboração deste projeto. Agradeço também ao meu ex-professor de português, Vidomar Silva Filho, ao executar sua função como professor como nenhum outro igual, indiretamente me proporcionando a possibilidade de ingresso neste curso que encerro, graças ao conhecimento de português que adquiri com ele de forma perfeita.

Quer você pense que pode ou não fazer algo, você está certo.

Henry Ford

Resumo

Esta monografia apresenta o estudo do áudio binaural e sua aplicação para videoconferências. O áudio binaural é um tipo de áudio que possibilita a percepção de um som em um plano tridimensional, assim como o discernimento de vozes simultâneas. Este estudo tem por objetivo a geração de sinais de áudio binaural através de um sistema que não dependa de dispositivos extras específicos e caros, utilizando apenas fones de ouvido e um computador convencional. Essa técnica pode ser aplicada em videoconferências para seu aprimoramento. Os resultados deste trabalho mostram que a aplicação do áudio binaural a uma videoconferência é possível e que irá promover uma melhoria no uso de videoconferências.

Palavras Chave: Áudio Binaural, Videoconferência, Processamento de Áudio.

Abstract

This monograph presents the study of the binaural audio and it's application for teleconferencing. The binaural audio is a kind of audio processing which allows sound perception in a tridimensional plane, as also the identification of simultaneous voices. This study's objective is the generation of binaural audio signals through a system which doesn’t need additional specific expensive equipments, using only headphones and a conventional computer. This work's results shows the application of the binaural audio to teleconferencing possible, and it shall promote an improvement for the teleconferencing usage.

Keywords: Binaral Audio, Teleconferencing, Audio Processing.

Sumário

Lista de Figuras

Lista de Tabelas

1 Introdução

O áudio binaural é um tipo de sistema de áudio no qual permite a localização de fontes sonoras num espaço tridimensional apenas com dois receptores. Este sistema de áudio é comum na natureza, onde quase todos os animais, incluindo os seres humanos, utilizam de dois ouvidos como receptores, para o uso do áudio binaural. Na maioria dos sistemas de áudio são utilizados apenas dois canais de áudio (direito e esquerdo) produzindo o efeito estereofônico (frequentemente denominado simplesmente estéreo), ou um único canal monofônico. Nas videoconferências a mesma tecnologia também é empregada.

1.1 Objetivo

Este trabalho tem como objetivo aprimorar as videoconferências, abordando o discernimento de vozes simultâneas e identificação dos participantes através da posição na tela. Propõe-se portanto a criação de um sistema que permita de forma eficaz transformar o áudio da videoconferência para o áudio binaural de uma forma que não necessite de uma carga computacional muito alta nem a necessidade de adquirir um equipamento eletrônico em específico.

1.2 Justificativa

Alguns estudos demonstram que em videoconferências nas quais duas ou mais pessoas falam ao mesmo tempo o uso de canais mono ou estéreo torna difícil a compreensão das falas (CIPIC, 2010). Por outro lado, há estudos que mostram que o uso do áudio binaural aprimora a videoconferência em dois aspectos: facilita o discernimento de vozes simultâneas e possibilita a localização dos participantes (CIPIC, 2010).

1.3 Organização do texto

O texto está organizado como segue. No Capítulo 2 são apresentadas a fundamentação teórica do trabalho, as características físicas do som, o áudio binaural. Em relação ao áudio, é estudada a forma como este é captado pelos seres humanos, o uso de microfones binaurais e de banco de dados de áudio binaural. No Capítulo 3 um sistema projetado para o uso do áudio binaural e sua interface com o usuário é apresentada. O Capítulo 4 mostra os resultados obtidos com o sistema proposto e as conclusões deste trabalho e propostas de desenvolvimentos futuros são pontuadas no Capítulo 5.

2 Fundamentação teórica

Neste capítulo serão tratadas questões como as propriedades das ondas sonoras, de que forma existe o áudio binaural nos seres humanos e então o processamento do áudio binaural, como este áudio é captado, gerado, e armazenado.

2.1 Ondas Sonoras

O entendimento das propriedades das ondas sonoras e sua interação com o meio físico principalmente em relação à atenuação sonora são vitais para a compreensão de como o áudio é captado (ouvido) pelos seres humanos.

As ondas sonoras são produzidas por vibrações da matéria, e necessitam um meio físico (gás, líquido ou sólido) para se propagar. A sua propagação ocorre através da alteração do meio, gerando compressão e rarefação das partículas da matéria (ver Figura 1). As compressões são áreas de maior densidade de moléculas enquanto que as rarefações são áreas de menor densidade de moléculas.

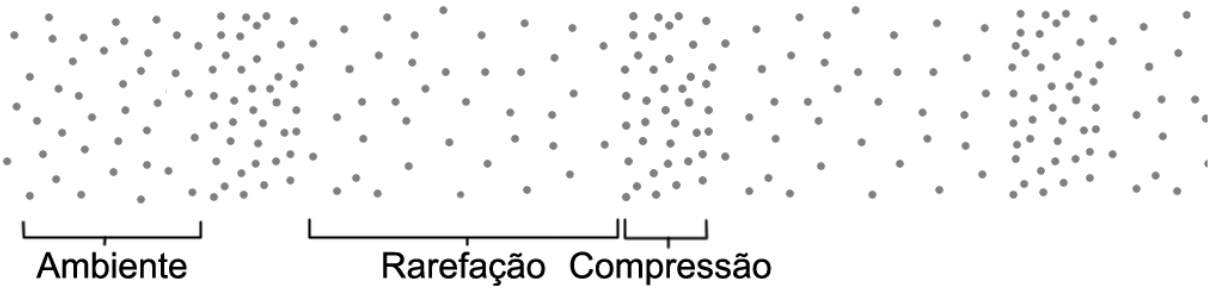

Figura 1 - Comportamento das moléculas na propagação do som.

A compressão pode ser chamada de “pico” e a rarefação pode ser chamada de “vale”. Como essa variação é relativa a um ponto zero, que seria a matéria em repouso, utiliza-se o termo “amplitude”. Para estudos mais simples da compressão e rarefação, é utilizada uma abordagem longitudinal: a amplitude é vista no eixo vertical e os picos e vales são observados ao longo do eixo horizontal (JANUS, 2004).

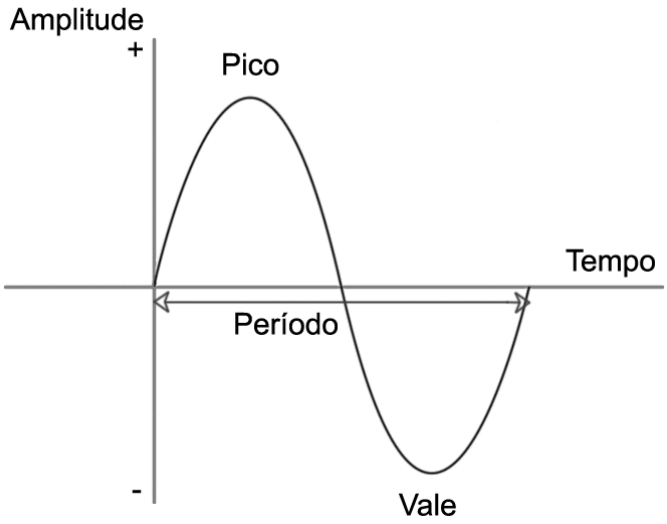

Figura 2 - Onda senoidal ou tom puro.

A Figura 2 ilustra uma onda sonora simples, chamada de “tom puro”. Um tom puro é uma onda senoidal. Para cada ciclo completo de uma onda é atribuído o nome de “período”, que é medido em segundos. A frequência de uma onda é a quantidade de vezes que ela se repete em um segundo, sendo medida em Hertz. Portanto o período é o inverso da frequência.

Independentemente da frequência ou período, podem ocorrer variações no tempo. Estas variações no tempo são denominadas de atrasos, conforme ilustrado na Figura 3. Como o estudo de ondas sonoras sempre se refere a um tempo inicial zero, as ondas sempre são tidas como atrasadas no tempo (JANUS, 2004).

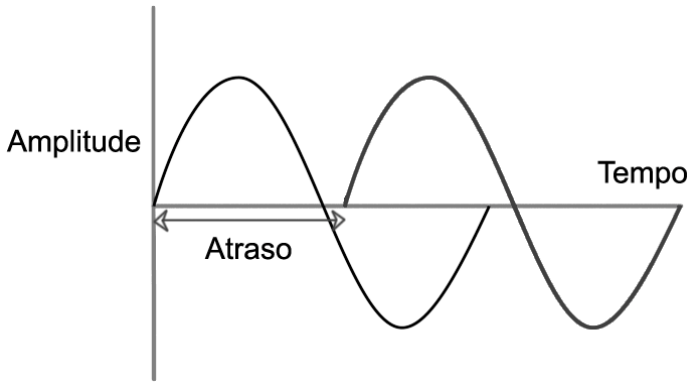

Figura 3 - Atraso demonstrado numa abordagem longitudinal.

Quando se refere às dimensões espaciais a abordagem longitudinal mencionada acima se refere apenas a duas das três dimensões espaciais, quando na realidade em um meio físico, as ondas sonoras se propagam em forma de casca esférica, como ilustra a Figura 4 (JANUS, 2004).

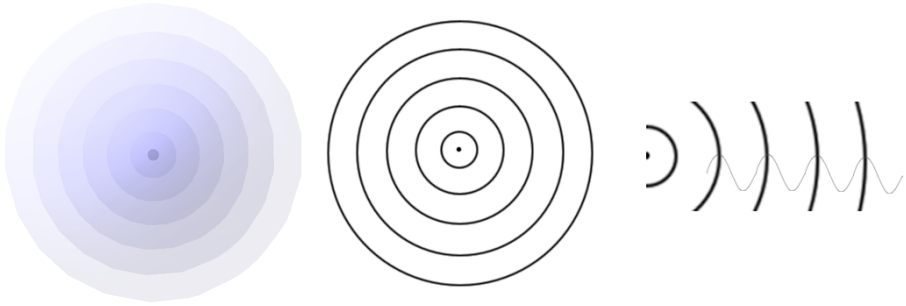

Figura 4 - Casca esférica: tridimensional, bidimensional, abordagem bidimensional.

A velocidade de propagação depende diretamente do meio na qual ela propaga. Para o caso de propagação no ar, que é de interesse neste estudo, a velocidade com que ela propaga depende das características físicas do ar, tais como, pressão atmosférica, temperatura, umidade, ou em termos mais específicos, temperatura, viscosidade dinâmica, viscosidade volumétrica e densidade (JANUS, 2004).

Com a frequência e a velocidade de propagação pode-se obter o comprimento de onda, que é a distância dentre dois períodos no espaço, medido em metros. Devido a propagação em casca esférica, o comprimento de onda permanece o mesmo tamanho durante a propagação, isso porque as ondas sonoras se propagam igualmente pelo espaço em todas as direções (JANUS, 2004).

Quando em contato com outro meio, uma onda sonora pode alterar sua trajetória. Se a onda não é transmitida para o outro meio, ocorre uma reflexão. Na reflexão de ondas sonoras, o ângulo de incidência e o ângulo de reflexão são iguais em relação a reta perpendicular à superfície do outro meio, conforme mostra a Figura 5.

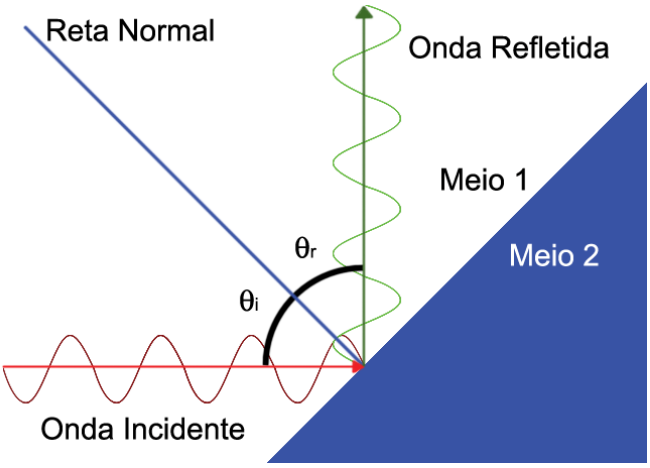

Figura 5 - Reflexão de uma onda sonora.

Na reflexão de ondas sonoras, os ângulos de incidência e reflexão são iguais em relação a uma reta perpendicular à superfície do outro meio, como mostra a Figura 5, no entanto ocorre a inversão da fase da onda refletida em relação a onda incidente. Se a onda sonora é transmitida para o outro meio, ocorre a refração, sendo neste caso, o ângulo de refração diferente do ângulo de incidência (ver Figura 6).

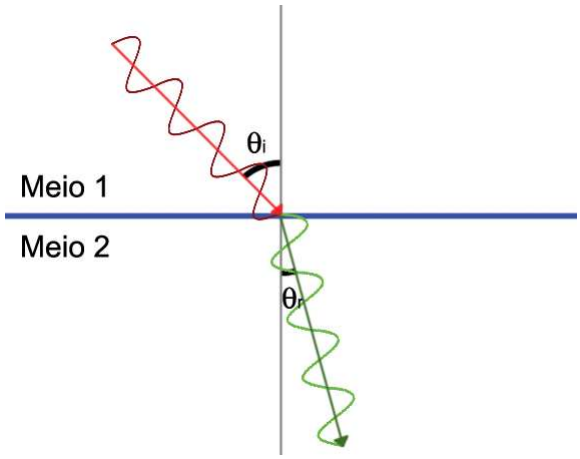

Figura 6 - Refração de uma onda sonora (não ocorre inversão de fase).

Uma onda sonora pode sofrer difração quando encontra um obstáculo, tendendo contornar o obstáculo (ver Figura 7). A curvatura desse contorno dependerá do tamanho do obstáculo e comprimento de onda (JANUS, 2004).

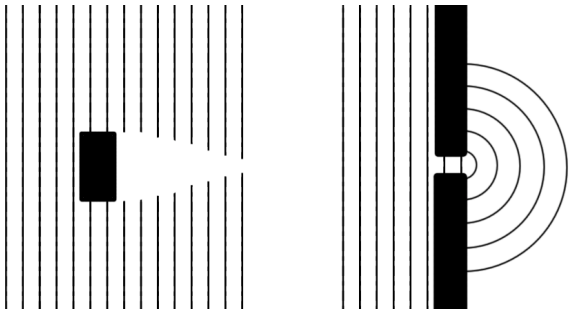

Figura 7 - Duas situações nas quais ocorre refração da onda sonora.

2.2 Coordenadas esféricas

Para a descrição da posição da fonte sonora em relação à cabeça do ouvinte, é necessário utilizar um sistema para referenciar cada posição no espaço em torno do ouvinte. Embora coordenadas cartesianas tridimensionais sejam suficientes para uma boa localização no plano tridimensional, como o áudio binaural é trabalhado com referência à cabeça, e como a cabeça pode ser dita como esférica, utilizam-se coordenadas esféricas para a localização. Ao invés de

se utilizar *x*, *y* e *z*, utilizam-se azimute (θ), elevação (ϕ) e distância (ρ). Para a conversão de coordenadas cartesianas para esféricas utiliza-se a Equação 1.

$$\begin{aligned} \theta &= \arctan 2(y, x) \\ \phi &= \arctan 2(z, \sqrt{x^2 + y^2} \\ \rho &= \sqrt{x^2 + y^2 + z^2} \end{aligned} \tag{1}$$

Para a conversão de coordenadas esféricas para cartesianas utiliza-se a Equação 2.

$$\begin{aligned} x &= \rho \cos(\phi) \cos(\theta) \\ y &= \rho \cos(\phi) \sin(\theta) \\ z &= \rho \sin(\phi) \end{aligned} \tag{2}$$

A representação em coordenadas esféricas pode ser realizada de duas formas distintas: vertical-polar e interaural-polar. O uso dessas formas de representação depende da escolha do usuário.

A representação vertical-polar é a mais utilizada em estudos do áudio binaural por ser de fácil compreensão e consiste em definir as posições variando primeiramente o azimute e depois a elevação. Conforme ilustra a Figura 8, as superfícies de mesmo azimute são planos que interceptam o eixo *z* das coordenadas cartesianas, enquanto que superfícies de mesma elevação são cones concêntricos com o eixo *z*. Nesta representação o azimute varia de 0° a 360° (ou -180° a 180°) e a elevação variar de -90° a 90°.

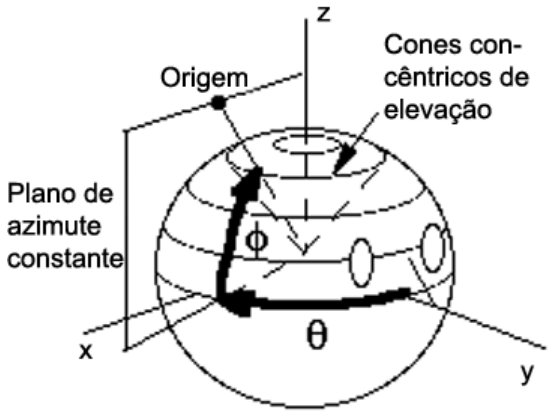

Figura 8 - Coordenadas esféricas vertical-polar.

Por outro lado, as coordenadas esféricas interaural-polar utilizam uma abordagem inversa, ou seja, primeiro é variada a elevação e depois o azimute. Neste caso as superfícies de elevação constante são planos que interceptam o eixo *x*, também denominado de eixo interaural, e as superfícies de azimute constante são cones concêntricos com o eixo interaural. Assim como na representação anterior, as variações de azimute e elevação também ocorrem

em ângulos, neste caso o azimute varia de -90° a 90° e a elevação de -180° a 180°, como mostra a Figura 9.

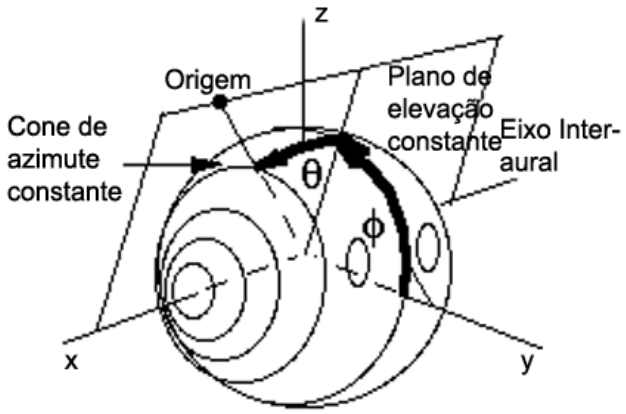

Figura 9 - Coordenadas esféricas interaural-polar.

Apesar de alguns estudos preferirem o uso das coordenadas interaural-polar, neste serão utilizadas as coordenadas vertical-polar por ser de mais simples compreensão.

2.3 A percepção do áudio binaural pelo homem

O áudio binaural consiste em simular a capacidade humana de perceber a origem de fontes sonoras em um plano tridimensional. A capacidade humana no aspecto de detectar fontes sonoras consiste em realizar comparações entre os sons recebidos por cada um dos ouvidos. Usando a diferença de atraso detectada por cada ouvido, a diferença de volume, a reverberação e o conhecimento prévio do som, o cérebro humano consegue deduzir a posição de onde o som é emitido.

2.3.1 Diferença Temporal Interaural

A diferença temporal interaural (*Interaural Time Difference* – ITD), como ilustra a Figura 10, ocorre quando um som percorre distâncias diferentes antes de alcançar cada ouvido, gerando uma defasagem entre os sons captados. Esta defasagem é calculada pelo cérebro, e permite determinar a posição do áudio no plano horizontal (CHENG, 2001). Tal detecção no plano horizontal requer, porém, a aprendizagem do cérebro, através do uso da visão para detecção da origem do som, e armazenamento da variação de defasagem detectada pelos ouvidos.

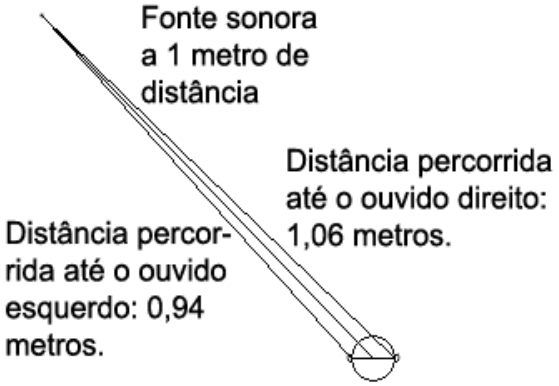

Figura 10 - Diferença temporal interaural.

2.3.2 Diferença de Nível Interaural

A diferença de nível interaural (*Interaural Level Difference* – ILD) resulta da variação da amplitude recebida pelos dois ouvidos, sendo dependente da frequência do som. Em baixas frequências, a cabeça humana refrata o som, portanto o ILD varia muito pouco, pois o ouvido recebe o áudio com uma amplitude quase a mesma sem a refração. Em frequências altas, a cabeça humana reflete o som, sendo este recebido pelo ouvido somente por reflexão de um outro objeto, o que causa uma atenuação destas frequências. Portanto dependendo da posição do som a cabeça humana age como uma sombra para o som, fazendo com que este seja recebido pelos ouvidos com maior variação de amplitude nas frequências mais altas. (CHENG, 2001).

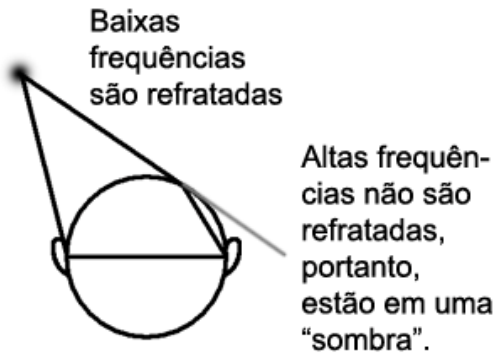

Figura 11 - Diferença de nível interaural, sombreamento do som.

2.3.3 Variação *de* fase

Para poder localizar a direção do som o cérebro humano também realiza a comparação da variação da fase do som em determinadas frequências. A Figura 12 mostra as diversas curvaturas que dão a forma ao ouvido humano. Os diferentes formatos e consistências de cartilagem no ouvido resultam em mudança nas propriedades acústicas de reflexão de acordo com a frequência do som. A cada reflexão da onda sonora a fase do som é invertida, possibilitando o cérebro determinar a origem do som relativo ao plano vertical através da detecção da variação de fase. Para esse poder realizar esse discernimento o cérebro também necessita de aprendizagem (CIPIC, 2010).

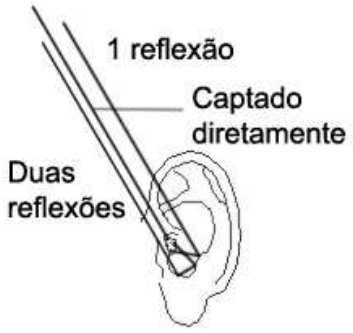

Figura 12 - Reflexões do som no ouvido humano

2.3.4 Refração

Além da ITD, ILD e da variação de fase, o cérebro também compara a presença ou ausência de certas frequências detectadas pelos ouvidos. Isso se deve ao fato de que certas frequências são refratadas pelo corpo humano. Se uma fonte sonora estiver atrás da pessoa, as orelhas irão tanto refratar como refletir determinadas frequências. Analisando a refração e a reflexão, o cérebro pode discernir a posição da fonte sonora detectando se está a frente ou nas costas (CIPIC, 2010).

2.3.5 Distância

Para a detecção da distância da fonte sonora, o cérebro não apresenta a mesma capacidade que demonstra para as posições do áudio. Nesse processo, o cérebro depende muito do aprendizado das amplitudes e característica de sons mais familiares. Por exemplo, o cérebro consegue discernir a diferença entre um sussurro e um grito, ou seja, ele tem registrado a diferença da amplitude sonora dentre um sussurro e de um grito. Devido a este

aprendizado é possível determinar a distância da fonte sonora com base na amplitude sonora esperada (CIPIC, 2010).

A paralaxe de movimento é um outro fator que pode auxiliar na detecção da distância da fonte. Essa característica consiste no movimento da cabeça em direção ao som ao ouvi-lo. Como em relação a posição da cabeça a posição da fonte sonora mudou, é possível determinar a posição do som com base nas variações de amplitude antes e após movimentar a cabeça (ver Figura 13). Se a fonte estiver próxima ao ouvinte, tal como em (c) e (d), o movimento de rotação da cabeça resultará uma grande variação da posição relativa, no entanto, se a fonte estiver mais distante, tal como em (a) e (b), o movimento de rotação da cabeça resultará em uma pequena variação da posição relativa (CIPIC, 2010).

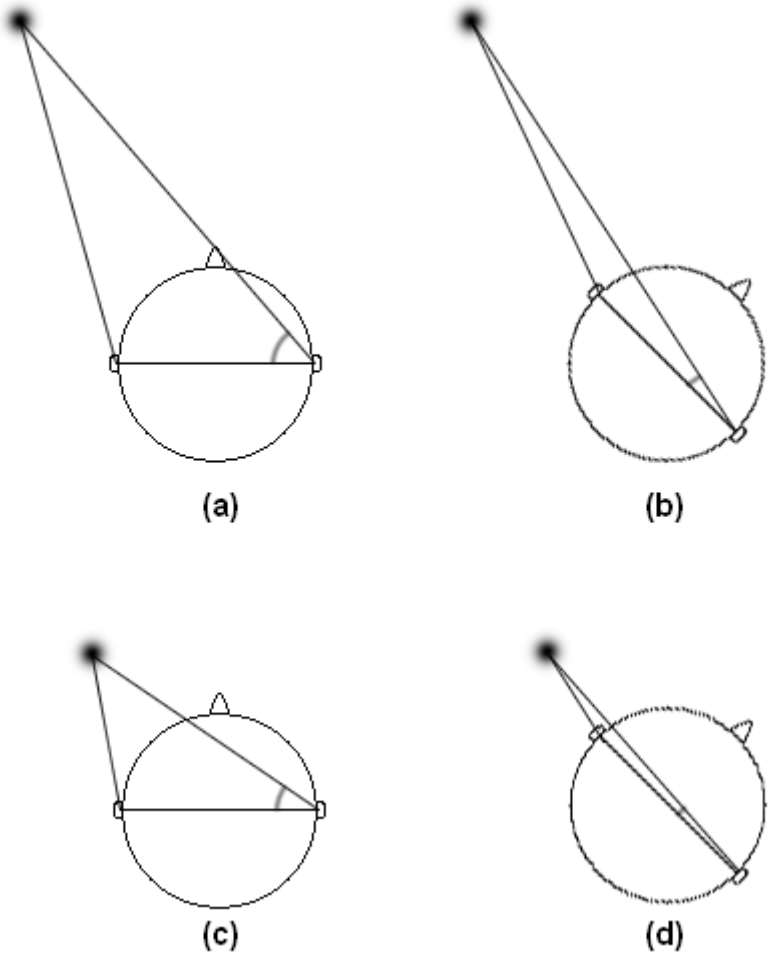

Figura 13 - Exemplo de paralaxe de movimento

Para a detecção da distância, o cérebro também utiliza da reverberação do ambiente como auxílio. Se a fonte sonora está muito próxima do ouvinte, o som reverberado percorrerá uma distância maior até refletir e ser captado pelo ouvinte. Porém se a fonte sonora está distante, o som reverberado percorrerá uma distância não muito diferente da fonte sonora,

resultando em uma percepção melhor da reverberação (ver Figura 14). Este comparativo de reverberação também auxilia na detecção da distância (GARDNER, 1999).

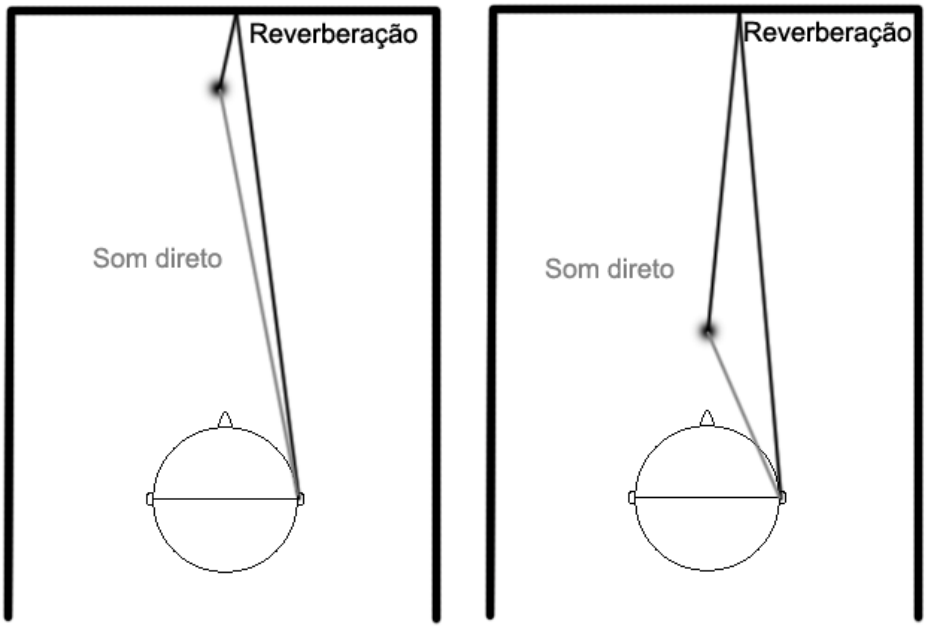

Figura 14 - Captação da reverberação

2.4 A Produção do áudio binaural

Considerando as características que o cérebro utiliza para discernir a posição e distância de uma fonte sonora discutidas anteriormente, pode-se definir como áudio binaural um sinal de áudio, transmitido em dois canais (esquerdo e direito), que permite ao ser humano determinar a posição da fonte sonora no espaço tridimensional.

2.4.1 Uso de Microfone binaural

Para se obter o áudio binaural podem ser usados dispositivos especiais de gravação denominados Manequim para Pesquisa Acústica da Knowles Eletronics (*Knowles Electronics Manikin for Acoustic Research* – KEMAR). Esse dispositivo consiste em um manequim antropomórfico da cabeça humana e parte do torso, com medidas baseadas em um exemplar humano. Esse manequim não necessariamente precisa ser uma representação fiel do corpo humano, no entanto, as orelhas devem ser perfeitamente reproduzidas, pois são essenciais à captação de áudio binaural. Dentro de cada ouvido do manequim há um microfone que é usado para captar o som depois de ter passado pela orelha. Como é necessário que o microfone inserido no canal de um KEMAR seja pequeno e de boa qualidade, recomenda-se

o uso de microfones do tipo *Core Sound Binaural*. (ANDERSON 2010). O KEMAR com microfones pode ser usado para capturar e gravar os sons de diferentes posições de fontes sonoras, resultando em excelente qualidade binaural, uma vez que a réplica da cabeça humana reproduz as suas características físicas, resultando em alterações nas características do som, que permitem a análise binaural do cérebro.

O problema da produção do som binaural, através da gravação usando o KEMAR, é a necessidade do uso do manequim para a gravação a cada distância e posição desejada. No caso de uma videoconferência, isso implicaria no uso de um manequim toda vez que for realizada uma gravação, e de conhecimento de processamento de sinais para operar o manequim, assim como o espaço necessário para o manequim, este muito maior que um microfone convencional (ANDERSON 2010).

2.4.2 Uso de Banco de dados

Por outro lado, através de um KEMAR um banco de dados pode ser obtido e utilizado posteriormente para gerar o áudio binaural digital. Esse banco de dados é gerado usando um KEMAR dentro de uma sala anecóica, onde são posicionados alto-falantes em posições que mantém sempre a mesma distância em relação ao centro do KEMAR. Um sinal de áudio específico, geralmente semelhante a uma função impulso, é gerado nos alto-falantes, e captado pelos dois microfones no KEMAR, desta forma é obtido o áudio binaural relativo ao áudio gerado em todas as posições escolhidas. As medidas para os canais de áudio direito (R) e esquerdo (L), são obtidas de forma independente, usando um microfone para capturar o áudio de um canal. Estes dados são armazenados como uma função no tempo na forma de resposta ao impulso relativo à cabeça (*Head Related Impulse Response* - HRIR) ou no domínio da frequência como a função de transferência relativa à cabeça (*Head Related Transfer Function* - HRTF) (CIPIC 2010).

O sinal de áudio específico produzido pela fonte sonora para obtenção de um banco de dados não necessita ser uma função impulso. A vantagem de se utilizar uma função similar ao impulso é que se obtém diretamente a resposta ao impulso ou a função de transferência. A desvantagem é que a função impulso é um sinal de pouca energia, podendo gerar efeitos não lineares nos alto-falantes ou microfones, ou seja, o som gerado pelos alto-falantes pode não ser mais uma função impulso pois esta foi modificada pelas características físicas dos alto-falantes tais como imperfeições, assim como as características físicas dos microfones podem alterar o som captado. Todavia, qualquer outro sinal utilizado que não seja uma função impulso já adiciona certa complexidade na obtenção do HRIR ou HRTF, pois necessita de

um cálculo adicional para a conversão do sinal utilizado para uma função impulso (CHENG, 2001).

Obtido o banco de dados, um áudio não binaural de apenas um canal pode ser transformado em áudio binaural através do processamento do áudio original com o banco de dados. Se o banco de dados estiver na forma de HRIR, é realizada a convolução do sinal de áudio monofônico $x(n)$ com a resposta ao impulso relativo à cabeça para obtenção do sinal de áudio no fone direito $y_{\mathrm{R}}(n)$ e no fone esquerdo $y_{\mathrm{L}}(n)$.

$$y_{\mathrm{R}}(n) = hrir_{\mathrm{R}}(n) * x(n) \tag{3}$$

$$y_{\mathrm{L}}(n) = hrir_{\mathrm{L}}(n) * x(n) \tag{4}$$

Se o banco de dados estiver na forma de HRTF, o sinal de áudio $x(n)$ é transformado para o domínio da frequência através de uma Transformada Rápida de Fourier (FFT):

$$X(\Omega) = \sum_{n=-\infty}^{\infty} x(n) e^{-j\Omega n} . \tag{5}$$

Obtido o sinal de áudio no domínio da frequência $X(\Omega)$, é realizada a multiplicação pela HRTF respectiva de cada canal para a obtenção dos sinais $Y(\Omega)_L$ e $Y(\Omega)_R$.

$$Y(\Omega)_L = HRTF(\Omega)_L . X(\Omega) \tag{6}$$

$$Y(\Omega)_R = HRTF(\Omega)_R . X(\Omega) \tag{7}$$

Porém como são necessários os sinais de áudio no domínio do tempo para serem executados nos fones de ouvido, é necessário converter os sinais $Y(\Omega)_L$ e $Y(\Omega)_R$ para o domínio do tempo (LATHI, 2007).

$$y[n]_L = \frac{1}{2\pi} \int_{2\pi} Y(\Omega)_L e^{jn\Omega} d\Omega \tag{8}$$

$$y[n]_R = \frac{1}{2\pi} \int_{2\pi} Y(\Omega)_R e^{jn\Omega} d\Omega \tag{9}$$

Também pode ser usado o método para HRTF com um banco de dados na forma de HRIR, através da transformação do sinal do banco de dados para o domínio da frequência:

$$HRTF(\Omega)_L = \sum_{n=-\infty}^{\infty} hrir(n)_L e^{-j\Omega n} \tag{10}$$

$$HRTF(\Omega)_R = \sum_{n=-\infty}^{\infty} hrir(n)_R e^{-j\Omega n} . \tag{11}$$

Desta forma é possível dispensar o uso de convoluções para a obtenção do áudio

binaural utilizando um banco de dados no domínio do tempo.

Para a realização da FFT, o uso da função de janelamento no domínio do tempo do sinal original é necessário, portanto também é necessário escolher o tipo de janela apropriado para realizar a FFT (SHENOI, 2006).

O procedimento de janelamento consiste em obter trechos de mesmo tamanho ao longo de um vetor de dados.

Neste estudo foram analisados dois bancos de dados disponíveis na internet, o banco de dados gerado pelo CIPIC/IDAV *Interface Laboratory - University of California* (BD-CIPIC) realizado pelo prof. V. Ralph Algazi (CIPIC, 2010) e o banco de dados gerado pelo MIT Media Lab (BD-MIT) realizado por Bill Gardner e Keith Martin (GARDNER, 1994).

2.5 Tecnologias semelhantes

Existem outras tecnologias semelhantes que já foram desenvolvidas com o objetivo de obter uma localização espacial através de aparelhos eletrônicos ou softwares, para execução em dois ou mais canais de áudio, porém de outras formas que não necessariamente através do uso do áudio binaural.

2.5.1 Roomsim

O Roomsim é um software criado por Douglas R. Campbell e Kalle J. Palomaki (CAMPBELL, 2010) que consiste em simular uma sala virtual e realizar diversos efeitos sonoros com base em propriedades físicas (temperatura , superfície das paredes, teto e chão, pressão atmosférica, etc.) da sala virtual gerada. Há também a opção de selecionar o tipo de receptor, incluindo a simulação de uma pessoa como receptor. O software permite utilizar os bancos de dados do CIPIC e BD-MIT.

Foram realizados alguns testes com esse software, mas verificou-se que os sinais binaurais gerados cujas distâncias do receptor variassem de um metro resultavam em variações imperceptíveis de distância.

2.5.2 Áudio 3D

O Áudio 3D consiste em técnicas usadas para obter sinais de áudio além da base do estéreo. A técnica consiste em alterar a fase do sinal dos canais de áudio direito e esquerdo. A técnica de Alargamento do Estéreo utiliza a manipulação das fases do sinal lateral (*side* - S) e do sinal central (C), obtidos a partir dos canais esquerdo (*left* - *L*) e direito (*right – R*)

$$C = \frac{L+R}{2}; S = \frac{L-R}{2} \tag{12}$$

Desta forma, a parte positiva do sinal lateral S é somada ao sinal do canal esquerdo e a mesma parte com fase invertida é somada ao canal direito (KIRKEBY, 2005).

3 Sistema Binaural Virtual

O Sistema Binaural Virtual (SBV) consiste em gerar o áudio binaural a partir de um banco de dados, desta forma, sendo feito apenas via software. Neste capítulo será apresentado de que forma o SBV foi realizado, os recursos utilizados, os testes de validação realizados e as conclusões diante do uso a videoconferências.

O SBV apresentado é o resultado do trabalho de projeto final realizado, consistindo de um sistema para localização do som em um plano tridimensional. O SBV pode ser considerado de baixa carga computacional e não necessita de dispositivos adicionais para o uso, exceto um fone de ouvido para cada ouvinte.

3.2 Material utilizado

Com a possibilidade de utilizar um banco de dados ao invés de ter de criar um manequim para obtenção do áudio binaural, optou-se pela criação de sistema por software que gerasse virtualmente o áudio binaural, desta forma, dispensando o uso do manequim.

O SBV foi inteiramente realizado na plataforma de software Matlab®. Para a realização do projeto e testes de validação foi utilizado um computador pessoal comum. Para fins de testes do áudio binaural foram usados fones de ouvido de uso convencional.

3.3 Interface do SBV

Para fins de estudos, testes e compreensão do áudio binaural, foi elaborada uma interface no Matlab a fim de manipular de forma mais fácil os dados obtidos e gerados pelo SBV. Esta interface possui as seguintes rotinas de execução: Primeiramente o usuário escolhe para ser processado um arquivo de áudio no formato wave. Posteriormente, a lista de opções principal é apresentada (Figura 15) na qual podem ser escolhidas as opções de gerar uma posição isolada ou um círculo de posições, e também alterar o arquivo de áudio a ser processado. Caso a opção escolhida seja um círculo de posições, uma nova lista de opções é apresentada

para a escolha do plano em que o círculo será gerado, e depois um painel (Figura 16) é mostrado no qual o usuário deve configurar as posições do círculo. Por outro lado, se apenas uma posição é escolhida, o mesmo painel é mostrado, porém sem o campo “intervalo”. Neste ultimo caso a opção de exportar o áudio binaural para um arquivo em formato wave é apresentada.

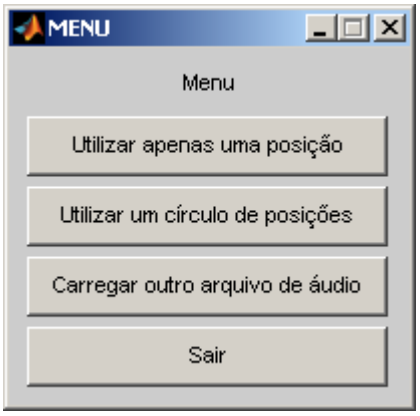

Figura 15 - Menu principal

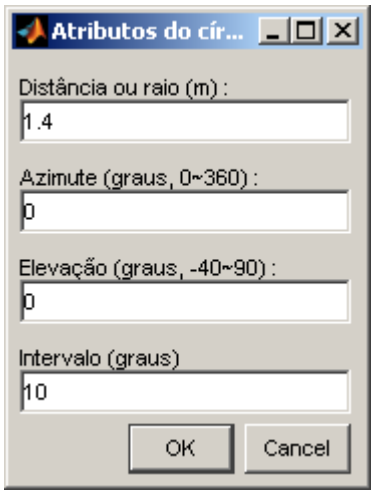

Figura 16 - Menu de configuração das posições no círculo

Especificada as posições, a interface do SBV permite através de uma lista de opções escolher o tipo de função a ser utilizada no procedimento de janelamento, conforme ilustra a Figura 17.

Figura 17 - Menu para escolha do tipo de janela

Depois dessas seleções o SBV processa o sinal de áudio de acordo com a configuração escolhida. Em seguida, a interface entra em uma segunda rotina de execução, quando é apresentada uma lista de opções na qual é possível executar o áudio escolhido antes e depois do processamento do binaural (Figura 18, neste exemplo um círculo de posições foi optado para ser gerado), assim como executar um círculo de posições de forma sequencial, executando o áudio binaural conforme as posições geradas no círculo.

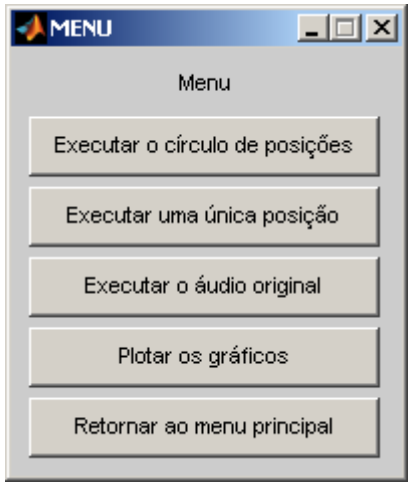

Figura 18 - Menu do círculo de posições

Neste menu há a opção de plotar os gráficos, que resulta na exibição de outro menu (Figura 19) que permite gerar gráficos dos áudios tanto no domínio do tempo (Figura 20) como no da frequência (Figura 22), podendo comparar os gráficos gerados tanto dentre o áudio original e o binaural como dois áudios binaurais de duas posições de um círculo. Para os casos de comparar o áudio binaural com o original, o gráfico é gerado ilustrando o áudio original com uma linha vermelha e o áudio binaural com duas linhas azul e verde, sendo estas para os canais esquerdo e direito respectivamente.

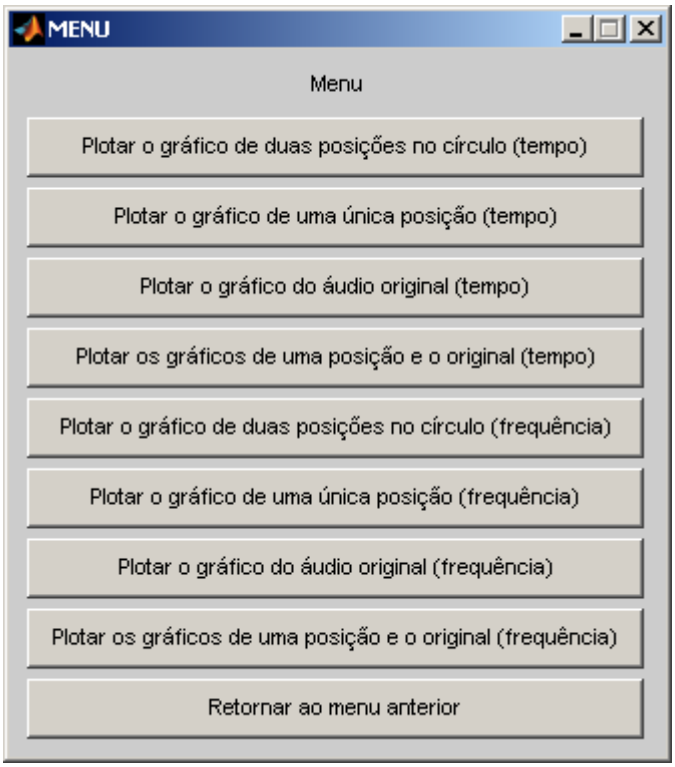

Figura 19 - Menu para plotar os gráficos

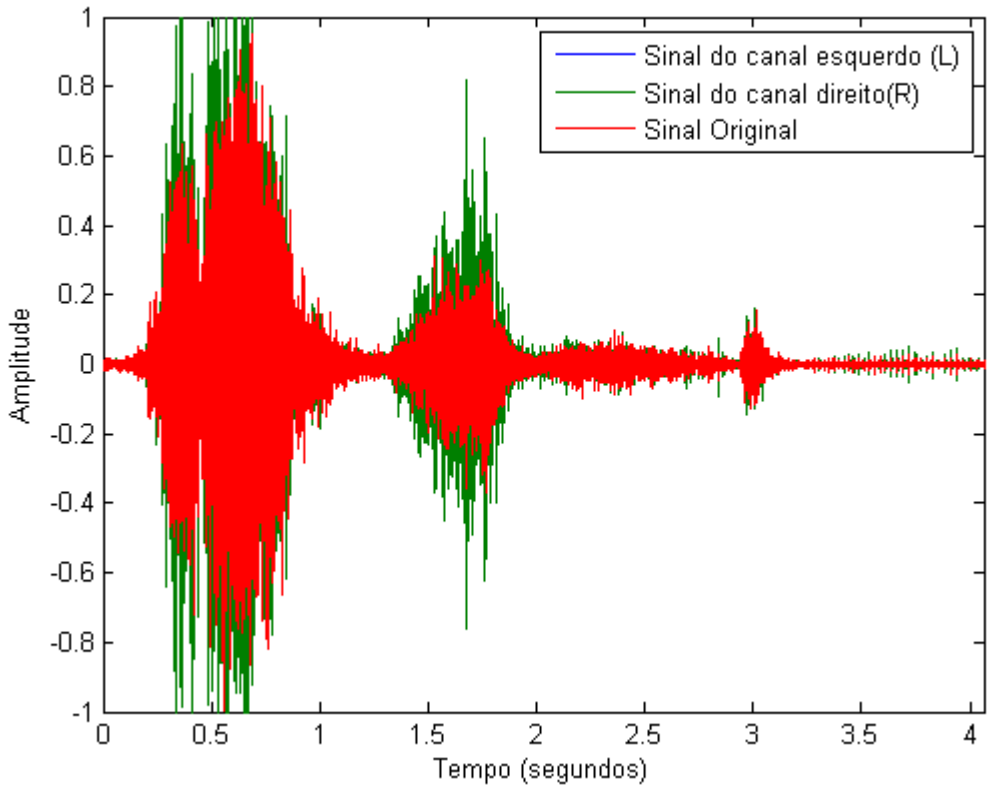

Figura 20(a) - Exemplo de um trecho de 4 segundos de sinal de audio no domínio do tempo. Note que nessa escala de tempo os sinais do canal esquerdo e direito aparecem sobrepostos.

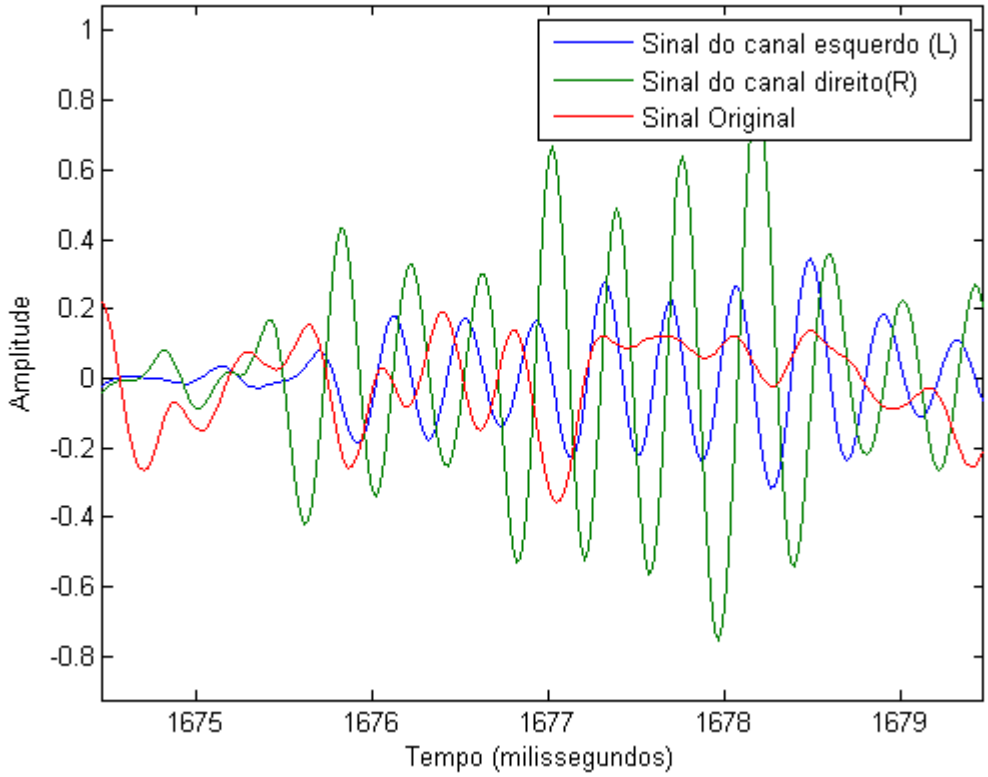

Figura 21(b) - Exemplo de um trecho de 4 milisegundos de sinal de audio no domínio do tempo. Note que nessa escala de tempo os sinais do canal esquerdo e direito são diferentes.

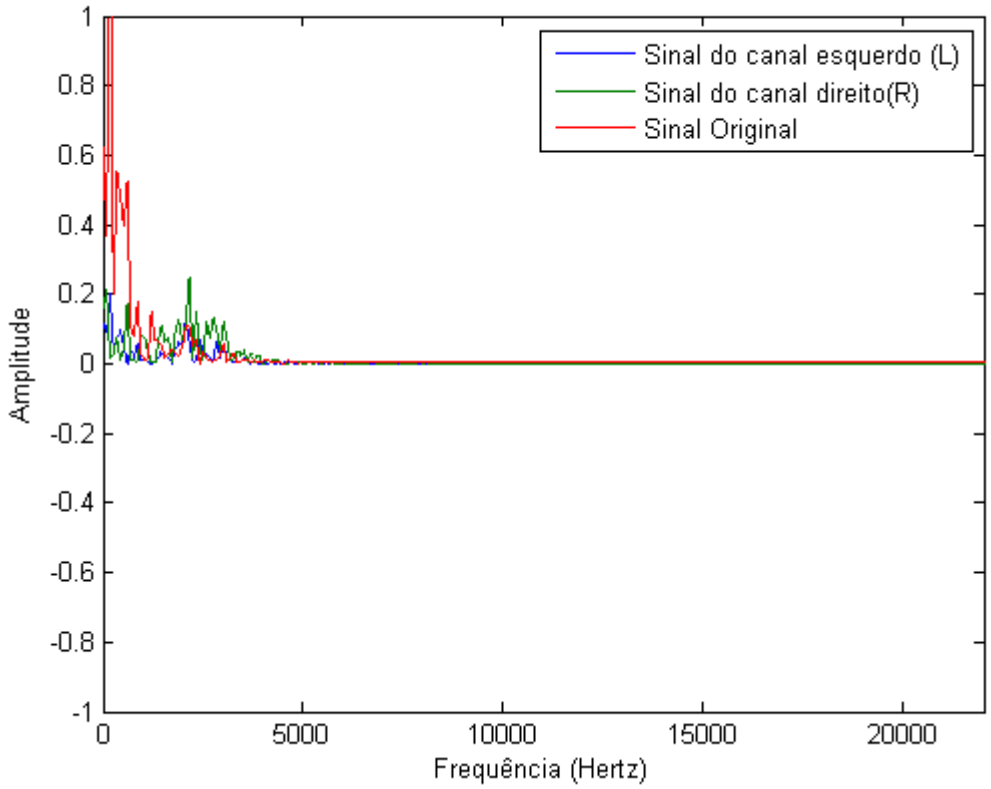

Figura 22 - Exemplo do espectro de freqüência do sinal de áudio.

3.4 Método básico de geração do sinal binaural a partir do BD-MIT

O SBV foi projetado com base no BD-MIT. Este banco de dados foi escolhido pois possui os dados em arquivos de som separados no formato PCM de 8 bits e com os canais separados em arquivos distintos, possibilitando uma manipulação mais simples desses dados. O BD-MIT está no formato HRIR, sendo que cada arquivo possui a HRIR de uma posição em coordenadas esféricas e cada canal em um arquivo diferente, resultando em dois arquivos por posição. Portanto, o processamento realizado no SBV consiste em converter o sinal de áudio para o domínio da frequência [conforme equação (5)], converter a HRIR da posição desejada para o domínio da frequência, ou seja, transformar a HRIR em uma HRTF através de uma FFT [conforme equações (10) e (11)], e realizar a multiplicação do sinal de áudio com a HRTF [conforme equações (6) e (7)]. Depois o sinal binaural resultante é convertido para o domínio do tempo [conforme equações (8) e (9)]. Desta forma é possível reduzir drasticamente a carga computacional em relação ao cálculo direto no domínio de tempo, no

qual é necessário realizar a convolução de *hrir(t)* com *x(t)* [conforme equações (3) e (4)] (CAMPBELL, 2010).

Como a característica ITD já está presente nas HRIR do BD-MIT, a convolução de um sinal sonoro mono com a HRIR de cada canal (direito e esquerdo) resulta em um sinal binaural que pode ser percebido pelo ser humano em diferentes posições espaciais tridimensionais, não necessitando portanto de gerar o ITD manualmente.

3.5 Procedimento de Janelamento

Para o processamento binaural dos sinais originais de áudio monofônicos é necessário realizar o corte deste sinal em amostras que sejam do mesmo tamanho (512 amostras) da função HRTF do BD-MIT, ou seja, realizar o janelamento do áudio original.

Foi necessário estudar os vários tipos de janelas, e analisar a variação dentre elas, com o intuito de se obter um áudio fiel ao original e sem ruídos. Os estudos realizados consistiram em comparar diversas janelas conhecidas e analisar o resultado obtido em experimentos subjetivos com avaliadores humanos. Dentre as janelas disponíveis, foi utilizada a janela de Blackman

$$\omega(n) = 0,42 - 0,5\cos\left(2\pi\frac{n}{N}\right) + 0,08\cos\left(4\pi\frac{n}{N}\right) \quad 0 \leq n \leq N \tag{13}$$

sendo *n* a amostra e *N* o tamanho da janela.

Como as HRIR do BD-MIT possuem 512 amostras, o áudio a ser processado foi dividido em trechos iguais de 512 amostras, sendo que nenhuma amostra pertence à dois ou mais trechos. Cada trecho então é processado independentemente, e feito o mesmo em todos os trechos obtidos do áudio, o SBV une de forma contínua todos os trechos processados de forma sequencial, sendo que nenhuma amostra é sobreposta dentre os trechos. Para realizar este processo é primeiramente determinado quantos trechos serão obtidos, através da equação

$$N = \frac{T_A}{512} \tag{14}$$

sendo N o número total de trechos a serem processados e T_A o tamanho do áudio. O número obtido necessita ser um inteiro, portanto o valor obtido é arredondado para cima.

O processo de janelamento do áudio em janelas de 512 amostras é feita através da equação

$$y[1,512] = w[1,512].x[k+1, k+512] \text{ para } k = 512.P \tag{15}$$

sendo y o trecho obtido, w a janela utilizada, x o trecho do áudio a ser processado e P o número do trecho sendo processado, tendo o trecho inicial como valor 0 e então somado 1 a cada trecho a ser obtido, sendo o último o valor calculado na equação (14).

3.6 Método de interpolação para posições não existentes no BD-MIT

O BD-MIT possui suas posições dispostas em intervalos uniformes, sendo para a elevação intervalos de 10° a partir de -40° até 90°, e para o azimute intervalos uniformes que variam de 5° a 30°, sendo maiores os intervalos de azimute conforme a proximidade das elevações de 90° e -40°, e na elevação de 90° há somente uma posição. A Figura 23 ilustra como é a distribuição das posições do BD-MIT.

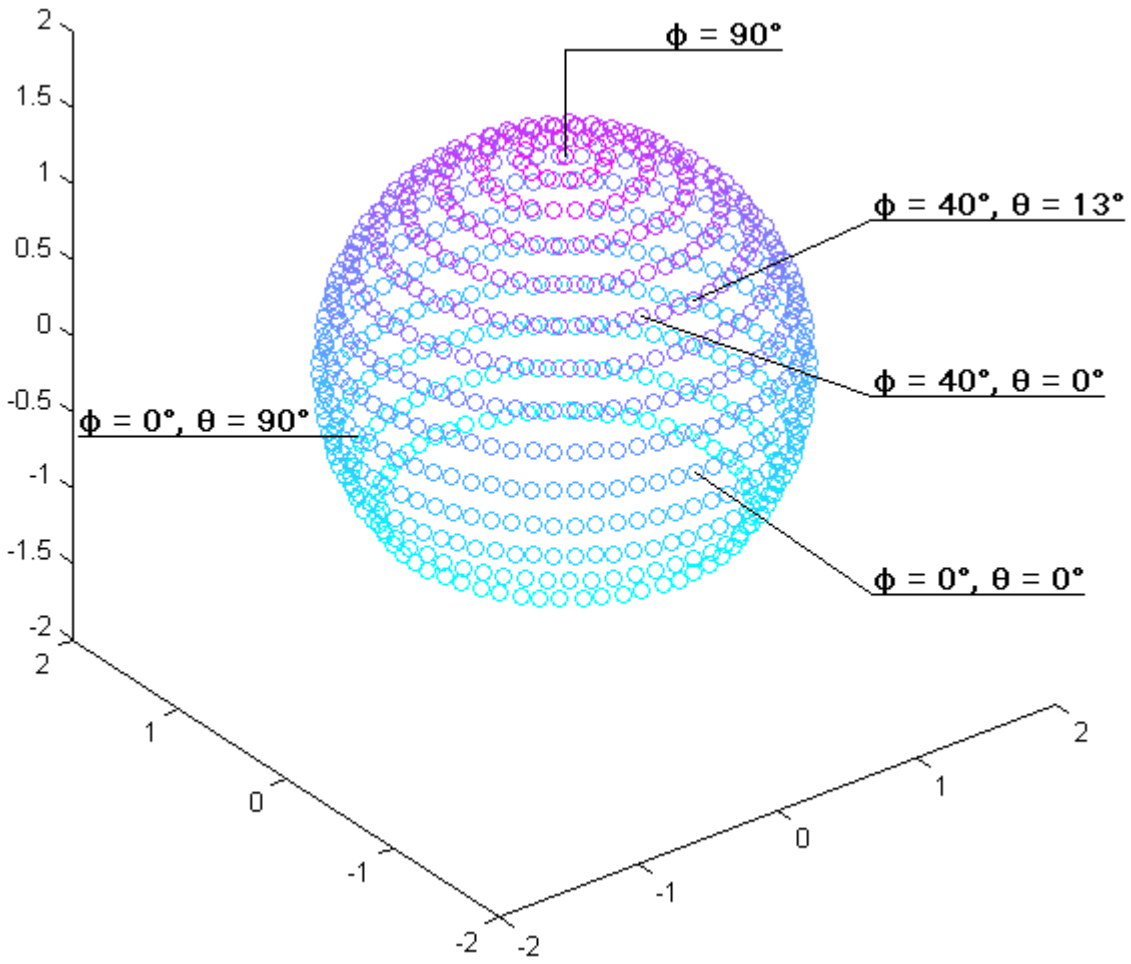

Figura 23 - Figura ilustrativa das posições disponíveis no BD-MIT.

No posicionamento do áudio binaural em uma videoconferência pode ocorrer que

eventualmente seja necessário obter a HRIR para posições que não constam do BD-MIT. Neste caso é possível obter-se uma HRIR aproximada da posição intermediária desejada pela interpolação das posições disponíveis. Para essa interpolação foi utilizada a seguinte técnica: uma vez obtida a posição a ser calculada, é verificado no BD-MIT as posições mais próximas disponíveis de ambos os canais, tanto na elevação quanto no azimute. Como os intervalos dos azimutes variam conforme a elevação é primeiramente obtida as elevações mais próximas, resultando sempre em duas posições, uma elevação acima da desejada, e outra abaixo da desejada. Depois são obtidas as posições de azimute mais próximas, sendo duas para cada elevação, um azimute após a posição e um antes, resultando em quatro azimutes obtidos no total. A Figura 24 abaixo ilustra como é o processo de interpolação. Como o BD-MIT possui as mesmas posições para ambos os canais (L e R), o processo é o mesmo para ambos os canais independentemente, então é atribuído o sinal *S* para a explicação.

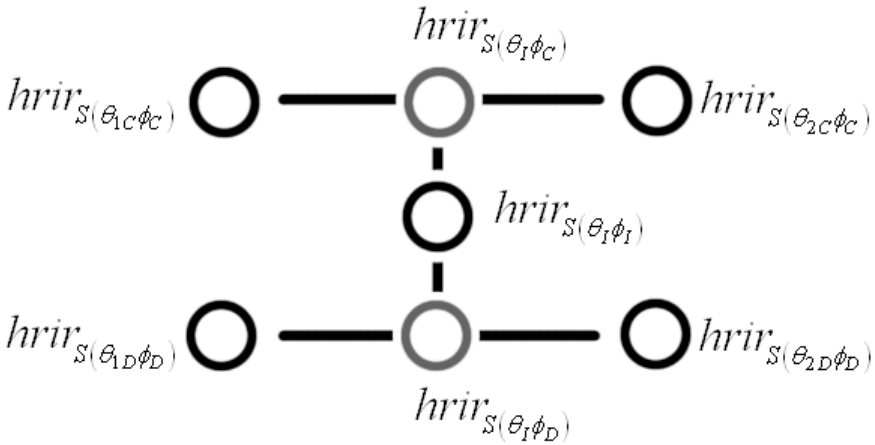

Figura 24 - Figura ilustrativa da interpolação.

Feito isso é realizada a interpolação dos azimutes de cada elevação independentemente, através de

$$hrir_{S(\theta_I\phi_C)} = \frac{(\theta_{1C} - \theta_I).hrir_{S(\theta_{2C}\phi_C)} + (\theta_I - \theta_{2C}).hrir_{S(\theta_{1C}\phi_C)}}{\theta_{1C} - \theta_{2C}} \quad (16)$$

onde $hrir_{S(\theta_I\phi_C)}$ é a resposta impulso superior a ser obtida, θ_I o azimute obtido para a interpolação, θ_{1C} o azimute posterior ao θ_I, θ_{2C} o azimute anterior ao θ_I, $hrir_{S(\theta_{1C}\phi_C)}$ a resposta impulso do azimute θ_{1C}, $hrir_{S(\theta_{2C}\phi_C)}$ a resposta impulso do azimute θ_{2C}. Para as HRIR a serem obtidas utiliza-se a elevação superior.

Para obter a interpolação da elevação inferior, uma fórmula semelhante é utilizada,

$$hrir_{S(\theta_I\phi_D)} = \frac{(\theta_{1D} - \theta_I).hrir_{S(\theta_{2D}\phi_D)} + (\theta_I - \theta_{2D}).hrir_{S(\theta_{1D}\phi_D)}}{\theta_{1D} - \theta_{2D}} \quad (17)$$

onde $hrir_{S(\theta_I\phi_D)}$ é a resposta impulso inferior a ser obtida, e as variáveis com denotação "*D*" do mesmo tipo da fórmula anterior, porém com valores distintos.

Tendo as interpolações dos azimutes, pode ser obtido a interpolação das elevações, com a equação

$$hrir_{S(\theta_I\phi_I)} = \frac{(\phi_C - \phi_I).hrir_{S(\theta_I\phi_D)} + (\phi_I - \phi_D).hrir_{S(\theta_I\phi_C)}}{\phi_C - \phi_D} \quad (18)$$

onde $hrir_{S(\theta_I\phi_I)}$ a resposta impulso da posição desejada, ϕ_C a elevação superior e ϕ_D a elevação inferior.

3.7 Considerações sobre a distância da fonte sonora

Como o BD-MIT corresponde a um banco de dados obtido para uma distância padrão de 1,4 metros, não há necessidade de realizar nenhum ajuste de distância quando a fonte sonora está posicionada a 1,4 metros de distância do ouvinte. Para distâncias diferentes da distância usada na aquisição do banco de dados é necessário realizar uma correção tanto na questão do ângulo de chegada do sinal ao ouvido, como na amplitude, para representar a atenuação sofrida.

Nas figuras 24 a 26, são ilustrados a fonte sonora como a esfera cinza, a casca esférica azul ilustrando as posições do BD-MIT, a cabeça ao centro da casca esférica, a linha preta a posição relativa à cabeça da fonte sonora, as linhas vermelha e azul indicam as posições relativas ao ouvido esquerdo e direito respectivamente, as esferas vermelha e azul representam as posições nas quais as linhas vermelha e azul cruzam a casca esférica, e as linhas roxa e ciano indicam as posições das esferas vermelha e azul em relação à cabeça .

Conforme mostram as figuras 24 a 26, a variação da distância da fonte sonora em relação ao receptor resulta em uma variação do azimute e da elevação da posição original. Isso implica que a variação da distância faz com que as HRIR não correspondam mais à posição correta relativo ao BD-MIT, que é sempre relativo à cabeça. Como se tem apenas as HRIR da casca esférica de posições, cujo raio é a distância padrão utilizada no BD-MIT, a variação nos ângulos de azimute e elevação originais mudam a posição na esfera de dados. Como a linha

preta ilustra, a orientação de posição dos dados do BD-MIT é relativa à posição na casca esférica de posições, mas como a variação de distância resultou em duas posições diferentes na casca esférica, é necessário calcular as posições do BD-MIT referente às novas posições obtidas, ilustradas como as linhas ciano e roxa.

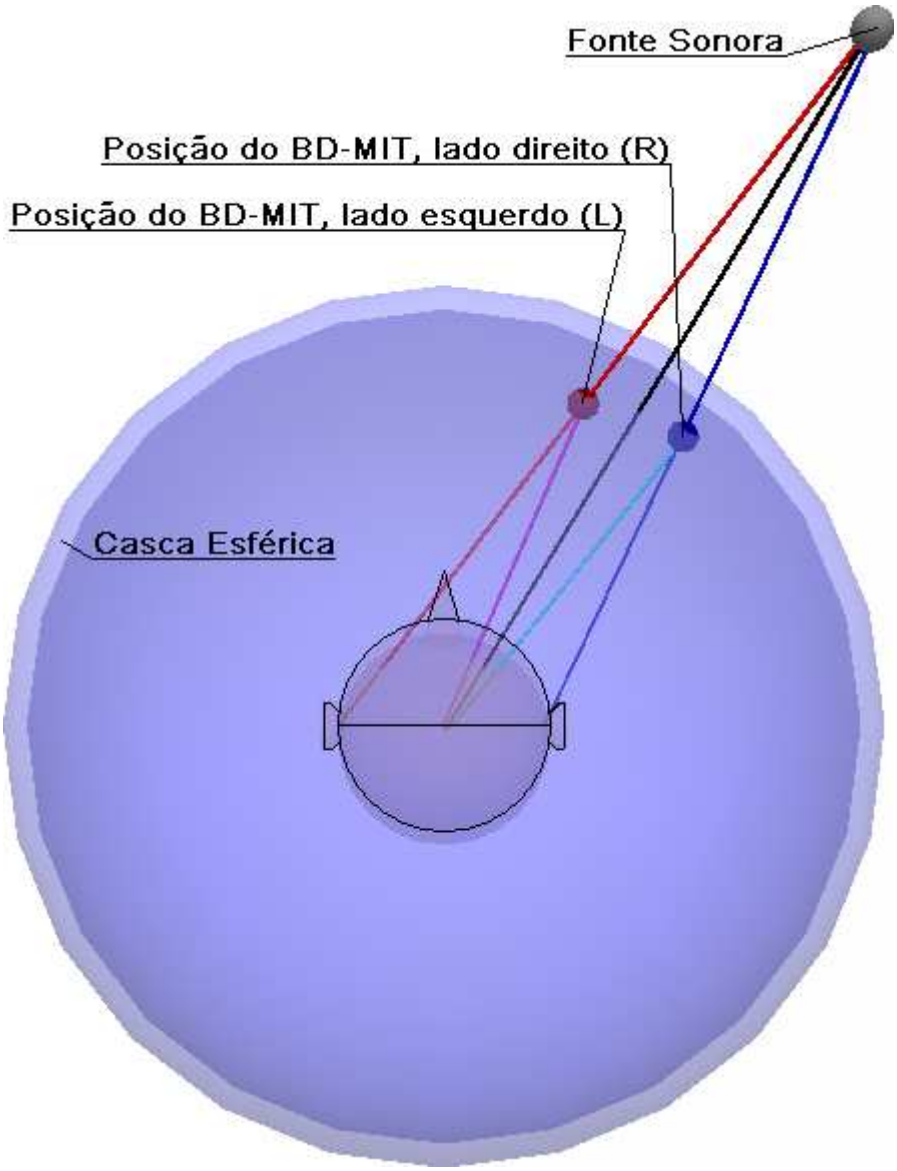

Figura 25 - Variação dos ângulos num plano tridimensional, visão superior (plano XY).

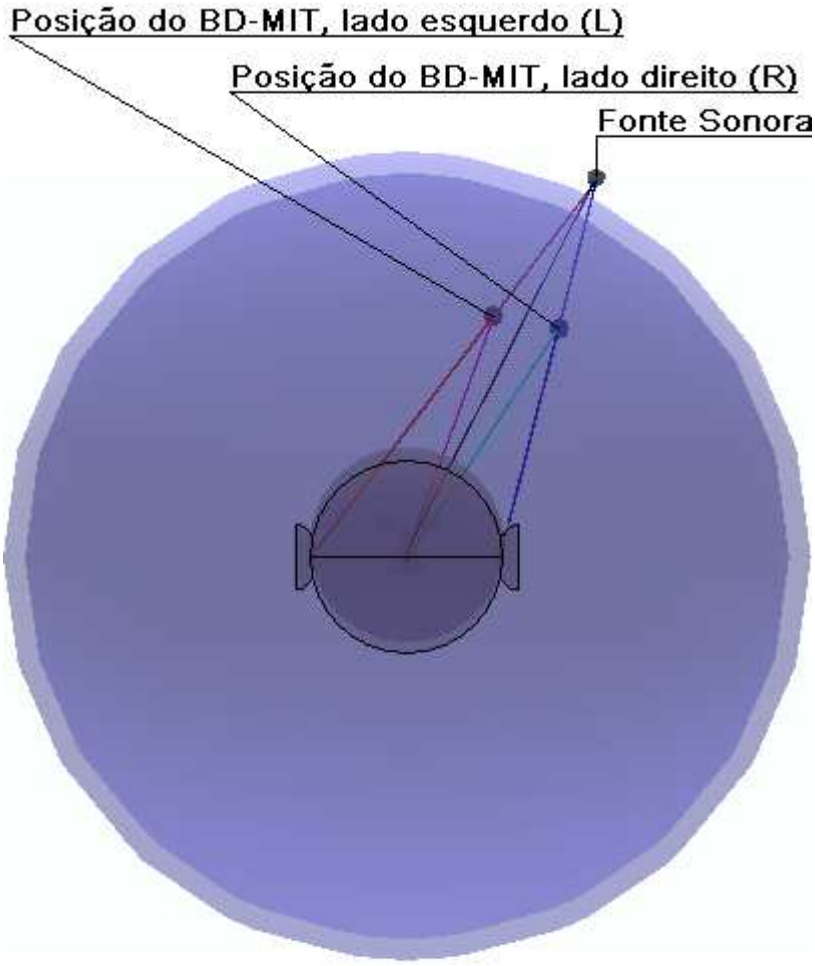

Figura 26 - Variação dos ângulos num plano tridimensional, visão traseira (plano XZ).

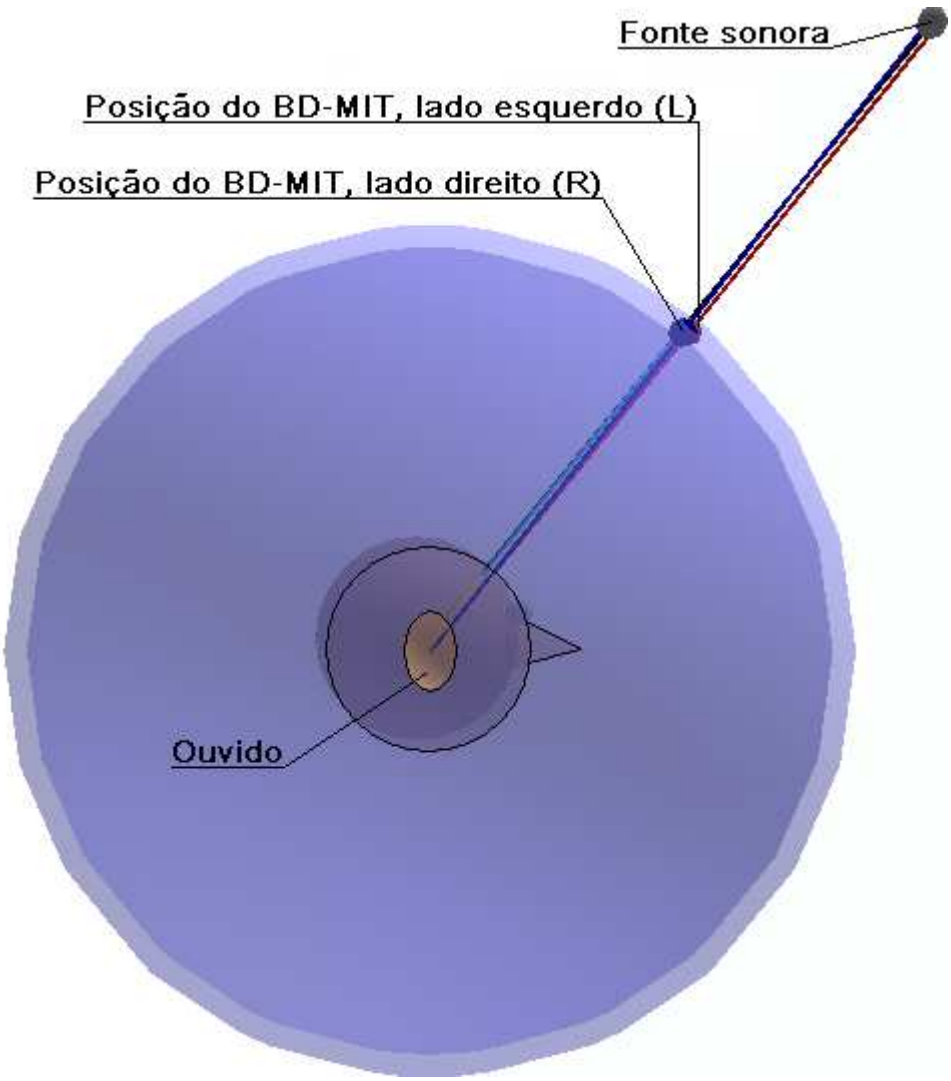

Figura 27 - Variação dos ângulos num plano tridimensional, visão lateral (plano YZ).

Para realizar o cálculo das novas posições é necessário determinar a intersecção da reta do ponto de origem até o centro da cabeça com a esfera de posições do BD-MIT. Desta forma é possível determinar qual são as novas posições na esfera referente ao BD-MIT, e portanto utilizar as HRIR adequadas.

Como o cálculo exige que as posições estejam em coordenadas cartesianas, primeiramente a posição da fonte sonora é convertida conforme a equação (2).

Considerando a função da casca esférica cujo centro é a posição $x = 0$, $y = 0$ e $z = 0$,

$$x^2 + y^2 + z^2 = r^2 \tag{19}$$

e a função da reta na forma paramétrica,

$$\begin{aligned} x_r &= x_R + w(x_F - x_R) \\ y_r &= y_R + w(y_F - y_R) \\ z_r &= z_R + w(z_F - z_R) \end{aligned} \tag{20}$$

sendo *r* relativo à reta, *F* à posição da fonte sonora, e *R* à posição do receptor (um dos ouvidos), é possível obter o ponto de intersecção da reta com a esfera obtendo o ponto no qual os valores x, y, z da reta e esfera são iguais.

$$\begin{aligned} &\left(w\left(x_F - x_R\right)\right)^2 + 2w\left(x_F - x_R\right) + {x_R}^2 + \\ &\left(w\left(y_F - y_R\right)\right)^2 + 2w\left(y_F - y_R\right) + {y_R}^2 + \quad . \\ &\left(w\left(z_F - z_R\right)\right)^2 + 2w\left(z_F - z_R\right) + {z_R}^2 = r^2 \end{aligned} \tag{21}$$

Agrupando os parâmetros com w e sem w, podemos obter a equação de 2º grau

$$aw^2 + bw + c = 0 \tag{22}$$

onde os valores dos coeficientes *a*, *b* e *c* são

$$\begin{aligned} a &= \left(x_F - x_R\right)^2 + \left(y_F - y_R\right)^2 + \left(z_F - z_R\right)^2 \\ b &= 2\left(x_R\left(x_F - x_R\right) + y_R\left(y_F - y_R\right) + z_R\left(z_F - z_R\right)\right). \\ c &= {x_R}^2 + {y_R}^2 + {z_R}^2 - r^2 \end{aligned} \tag{23}$$

As raízes da equação (22) são:

$$w = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \tag{24}$$

Como é de interesse apenas uma das intersecções, a mais próxima da fonte sonora, apenas a raiz positiva de *w* é usada na substituição da função da reta na forma paramétrica, e posteriormente convertida para coordenadas esféricas. Assim o valor positivo de w é substituído na equação (20) para obter os valores *x*, *y* e *z* do ponto de intersecção. Este ponto em seguida pode ser convertido da coordenada cartesiana para vertical-polar através da equação (1).

Porém, como os ouvidos não ficam no centro da cabeça e sim nas laterais, é definida uma posição diferente de zero para as orelhas no eixo *y*, definido pelo BD-CIPIC de aproximadamente 0,072m de distância do centro da cabeça, portanto é alterado o valor de y_R = 0,072 para a orelha direita e y_R = -0,072 para a orelha esquerda.

Também se percebeu que a variação não seria igual para ambos os ouvidos, gerando duas posições de HRIR distintas, uma para cada ouvido, ilustrado nas figuras 24 e 25, demonstrado pelo fato das linhas azul e vermelha estarem em uma posição diferente da linha preta, a referência do BD-MIT. Isso foi resolvido simplesmente carregando as HRIR do BD-MIT de posições distintas, já que estas já estão dispostas em canais separados.

4 Resultados

Neste capítulo são apresentados os resultados obtidos com o SBV, tanto para o áudio binaural analisado de forma isolada, como o seu uso em um ambiente virtual de videoconferência. Como a percepção do áudio binaural depende de um ser humano, os testes foram realizados de forma subjetiva usando três avaliadores humanos (FSB, JDSK, RPO). Nestes testes, cada avaliador deveria indicar aproximadamente de qual direção e distância o som foi percebido. Estes dados subjetivos foram em seguida comparados com as posições nas quais o som foi gerado. O resultado foi considerado positivo quando não houveram erros significativos na localização.

4.1 Método de avaliação

Devido ao resultado final ser subjetivo a detecção por um ser humano, foram necessários avaliadores para opinarem sobre a eficácia e eficiência do SBV. Os testes consistiam basicamente da comparação dentre a posição detectada e a posição gerada, para verificar a precisão, mais especificamente se a posição foi detectada no lado correto (em cima, ao lado, à frente, etc.) e numa região próxima da esperada.

4.2 Testes realizados

Para avaliar os testes na parte inicial foram usados os avaliadores EAS e MM. Os primeiros testes consistiam de verificação de sensação de percepção, com posições aleatórias e mesma distância, e também fazendo comparativo entre o BD-MIT e BD-CIPIC. Posteriormente foram feitos testes com círculos de posições. E por fim testes com variação de distância.

4.2.1 A rotação do som no plano horizontal

Este teste consistiu em analisar a percepção da origem sonora num plano horizontal, sem

qualquer elevação. Um mesmo som era gerado a partir do azimute 0° e então incrementando o ângulo em intervalos fixos de 45°, até chegar aos 360°, gerando um círculo de posições. Para cada posição o mesmo sinal sonoro era usado, de modo que o avaliador deveria perceber esse som nas posições previamente definidas. Conforme a posição é alterada, percebe-se o som se movimentando em círculo ao redor da cabeça. Os resultados deste teste foram positivos, indicando que o SBV posiciona corretamente o som neste plano, ou seja, a detecção da posição do som em relação à frente, lateral, e costas foram bem sucedidas.

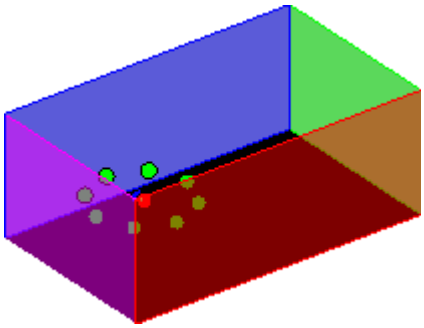

Figura 28 - Figura ilustrativa dos pontos gerados no plano horizontal.

4.2.2 A rotação do som no plano vertical

Para estudo da rotação do som no plano vertical foram utilizados dois testes. Estes testes consistiam em analisar a percepção da origem do som num plano vertical. O primeiro teste foi relativo ao plano XZ (plano ilustrado na Figura 26), portanto a fonte sonora era primeiramente posicionada exatamente à frente (azimute 0° elevação 0°) e então incrementando a elevação em intervalos fixos, até chegar a posição azimute 180° elevação 0°. Isso se deve ao fato de ter sido utilizado as coordenadas esféricas vertical-polar, portanto após a posição de elevação 90° a variação continuou decrementando a elevação na mesma quantidade anteriormente, porém com azimute 180°.

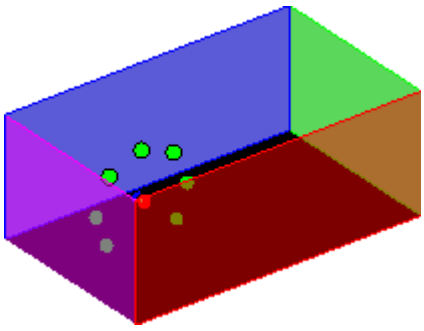

Figura 29 - Figura ilustrativa dos pontos gerados no plano vertical, plano XZ.

O segundo teste foi relativo ao plano YZ (plano ilustrado na Figura 18), portanto a fonte

sonora era primeiramente posicionada à esquerda (azimute 90° elevação 0°) e então a alteração da posição decorreu exatamente como o teste no plano XZ, neste caso após a posição de elevação 90°, alterando o azimute para 270°.

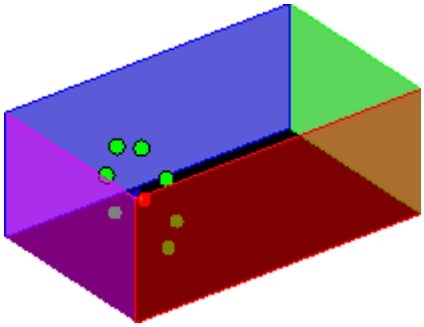

Figura 30 - Figura ilustrativa dos pontos gerados no plano vertical, plano YZ.

4.2.3 O afastamento e aproximação do som

Para o teste de afastamento e aproximação da fonte sonora, buscou-se analisar a percepção da origem do som variando somente a distância. Primeiramente foram utilizadas várias posições com intervalos iguais de distância, sem variação de elevação e azimute. Embora a execução do mesmo som em posições diferentes e sequenciais gerassem a sensação de aproximação ou distanciamento e a sensação de distância, a execução de apenas um som em uma distância maior diferente daquela adotada no BD-MIT não resultava em uma percepção correta da distância. Isso se deve ao fato da função da atenuação não estar precisamente correta.

Foi observada também a questão da percepção da distância relativa a um meio realístico, ponderando que o áudio gerado estivesse de fato com uma atenuação correta, porém como foi gerado apenas um som, este estaria presente em um meio de silêncio absoluto, o que não ocorre na realidade. Portanto foram feitos testes com inserção de ruídos de fundo juntamente com o áudio binaural gerado. Foram utilizados ruídos como som de vento e música, e então ajustadas as amplitudes para não interferir com a percepção do áudio binaural. Contudo, não foi obtido nenhum resultado positivo com estes testes.

Para obter uma atenuação correta para cada distância de fonte sonora, foi utilizado a lei de Stokes

$$\alpha = \frac{2\left(\eta + \eta^{v}\right)\omega^{2}}{3\rho V^{3}} \tag{25}$$

onde α é a atenuação em Neper/m, η é a viscosidade dinâmica, η^v é a viscosidade volumétrica, ω é a frequência, ρ a densidade, e V a velocidade do som.

Como a lei de Stokes é dada em Neper por metro, o valor obtido foi então convertido para decibéis, como foi utilizado em todo o projeto. Porém, a atenuação pela lei de Stokes também não resultou em uma atenuação realística, apesar de ter sido utilizado o máximo de precisão possível.

4.2.4 O teste do auditório

Este teste consistiu em gerar um programa que exibisse em um monitor uma imagem fixa de uma simulação de um auditório, e obter a detecção precisa da posição na tela da fonte sonora gerada, armazenando os resultados obtidos no computador. A imagem gerada consistia do auditório sendo visto do palco, mostrando as cadeiras, semelhante a Figura 31.

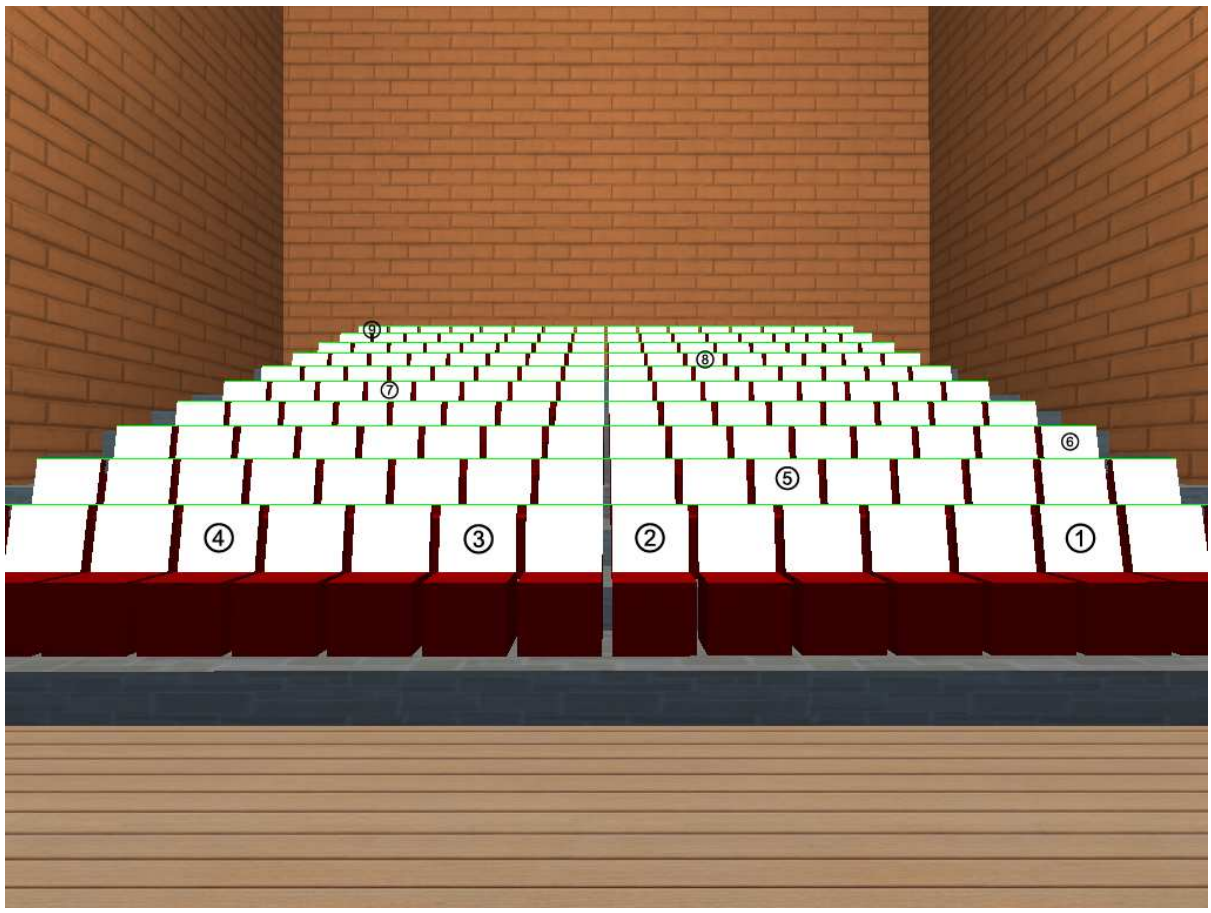

Figura 31 - Auditório gerado com as posições pré-determinadas numeradas.

Posições eram pré-estabelecidas e o avaliador deveria indicar com o mouse a posição que julgasse ser a origem do som conforme este era ouvido. O programa então calculava a distância da fonte sonora gerada com a posição que o avaliador detectou. O cálculo dessa

distância era medida por píxeis na tela, através da seguinte equação

$$Distância = \sqrt{\left(x_2 - x_1\right)^2 + \left(y_2 - y_1\right)^2} \quad (26)$$

sendo x_2 e y_2 as coordenadas da posição selecionada e x_1 e y_1 as coordenadas da posição gerada.

O acerto era avaliado conforme uma distância mínima pré-estabelecida para o erro, em torno de 20 *pixel*, que consistia em um círculo cujo encosto da cadeira na figura coubesse exatamente dentro do círculo. Qualquer posição selecionada fora do círculo era considerada um erro, apesar de que a distância do erro poderia ser relevante para o estudo caso fosse uma distância demasiada. Calculadas todas as distâncias obtidas era gerado um gráfico demonstrando a distância obtida e a margem de erro.

4.2.5 Conclusões dos testes

Os testes iniciais concluíram que apesar da qualidade do BD-CIPIC ser melhor, o uso do BD-MIT seria mais interessante por possuir os dados no formato wave, com canais separados de cada arquivo, tornando a manipulação dos dados mais fácil, e também porque as posições do BD-MIT eram dispostas em variações uniformes. Já o BD-CIPIC não possuía posições além de 80° azimute e as variações de posições não eram uniformes, dificultando o uso e eventuais testes de validação.

Em uma abordagem mais ampla, analisando a percepção de posições mais variadas, como diretamente acima, nas laterais, à frente e atrás, o SBV mostrou ser eficaz em gerar tais posições, exceto exatamente à frente (azimute 0° e elevação 0°), pois o avaliador tinha a sensação de que o som estava imediatamente à frente do ouvinte, sem qualquer percepção de distância. Isso se deve ao fato desta posição específica condizer com um áudio mono. Com os testes dos círculos de posições concluiu-se que o SBV também foi capaz de gerar as posições com eficácia.

Tabela 1 – Avaliação do SBV referente a posições dispostas em círculos.

Avaliador	Plano YZ	Plano XZ	Plano XY
FSB	Percepção como se fosse no plano Z.	Percepção precisa.	Percepção precisa.
JDSK	Nenhuma percepção de posição distinta.	Percepção traseira, mas precisa.	Percepção somente lateral e imprecisa.
RPO	Percepção somente traseira, mas precisa.	Percepção precisa.	Percepção somente traseira, mas precisa.

A Tabela 1 mostra que os resultados foram consistentes por cada avaliador isoladamente, sendo que os testes foram feitos sem que os avaliadores soubessem de onde o som iria aparecer. O avaliador JDSK não obteve um bom resultado, enquanto que o avaliador RPO obteve uma percepção mais traseira, ou seja, as posições tendiam a ser percebidas atrás da cabeça, refletindo assim que a variação do formato da orelha também possui um fator relevante para a detecção.

Em relação à distância da fonte sonora, o uso da lei de Stokes, permitiu uma percepção do aumento da distância, mas essa distância percebida não era condizente com a distância calculada, pois mesmo para distância muito grandes nas quais o som não deveria ser percebido, a implementação realizada não resultou em uma atenuação adequada.

No teste de auditório virtual, concluiu-se que devido à proximidade e tamanho do monitor e a falta de precisão do BD-MIT apesar da interpolação, a avaliação demonstrou uma percepção não exata em relação à posição gerada. Os avaliadores também comentaram que em muitos dos casos apesar da posição gerada ter sido alterada, a percepção era da mesma posição anterior. Estes dados estão dispostos na Figura 32, na qual os valores acima do limiar indicam erro de localização.

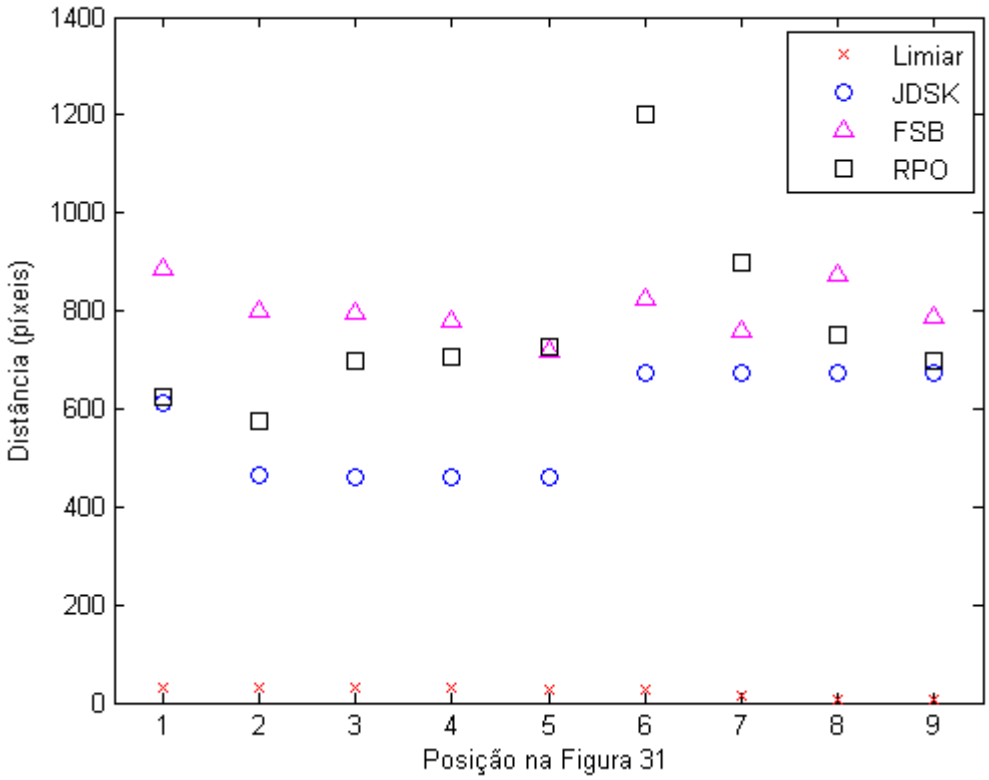

Figura 32 - Teste do auditório.

5 Conclusão e Trabalhos Futuros

Neste trabalho é descrita a implementação de um sistema de áudio binaural que utiliza um banco de dados de HRIR disponibilizado pelo MIT. O sistema desenvolvido tem por objetivo permitir a localização do interlocutor através do áudio em aplicações e videoconferências. O sistema SBV implementado utiliza apenas o banco de dados do MIT contendo HRIR de várias posições específicas, para transformar um sinal sonoro monofônico em um sinal de áudio binaural. Desta forma o sistema evita a necessidade de gravação do som com equipamentos específicos tais como um KEMAR com microfones embutidos. No sistema proposto todo o processo é realizado sem novas aquisições binaurais, usando portanto apenas o processamento do som. Desta forma a implementação do áudio exige apenas um processador no qual o SBV é executado, e um sistema de fones de ouvido para o vídeo conferencista.

Concluiu-se que o SBV depende exclusivamente de um banco de dados preciso e com várias posições disponíveis, e da implementação correta da atenuação para possibilitar a percepção da distância apropriada da fonte sonora. O uso de um sinal de referência auxiliar possibilitou uma melhor percepção da distância.

Não foi realizada uma avaliação de complexidade computacional, mas para se ter uma noção desta complexidade, avaliou-se o tempo de processamento. Para tal o sistema SBV foi executado em um PC com 2GB de memória RAM e processador Intel Pentium Dual Core 1.61GHz. Sob estas condições o tempo de processamento do áudio binaural foi de 1,5 segundos para um sinal de áudio de 60 segundos. Para resultados mais conclusivos sobre a complexidade, é necessário realizar um melhor estudo no qual, o número de operações de soma, multiplicações e armazenamentos precisam ser determinados. Apesar disso, acredita-se que a complexidade é suficientemente baixa, não sendo um empecilho para a aplicação em videoconferências.

Em termos de propostas futuras para a complementação e melhorias do sistema SBV, sugere-se que também sejam investigados os seguintes aspectos:

1. Propor uma função que permita perceber de forma correta a distância que a fonte se

encontra do receptor, uma vez que os testes realizados usando a lei de Strokes não produziram uma atenuação adequada. Sugere-se que se utilize uma combinação de sons próximos e distantes para facilitar a percepção de distância.

2. Geração de um banco de dados com maior precisão, o qual poderia ser gerado usando um KEMAR. Acreditamos que a existência de mais posições disponíveis por azimute e elevação poderia reduzir a imprecisão do processo de interpolação empregado para obter as posições não existentes no banco de dados;
3. Que sejam feitos novos testes de validação com um maior número de avaliadores, para obter uma avaliação estatisticamente mais significativa;
4. Uso do sistema SBV em aplicações de áudio binaural como conferências ao vivo, no qual é usada a tradução simultânea, ou ainda em ambientes maiores nos quais são usados fones de ouvido para a distribuição do som. Um possível exemplo de utilização poderia ser o ambiente usado para as reuniões da ONU, nas quais os conferencistas poderiam receber o áudio traduzido de outros conferencistas, de modo a poder perceber a direção do locutor original;
5. Testes para verificação da melhor separação de áudios no caso de mais de um interlocutor falar ao mesmo tempo;
6. Análise da complexidade computacional do sistema SBV;
7. Integração do SBV em um sistema de videoconferência.

Lista de Abreviaturas e Siglas

BD-CIPIC – Banco de dados do CIPIC/IDAV

BD-MIT – Banco de dados do MIT

FFT – Transformada rápida de fourier (*Fast Fourier Transform*)

HRIR – Resposta ao impulso relativo à cabeça (*Head Related Impulse Response*)

HRTF – Função transferência relativa à cabeça (*Head Related Transfer Function*)

ILD – Diferença de nível interaural (*Interaural level difference*)

ITD – Diferença de tempo interaural (*Interaural time difference*)

KEMAR – Manequim para Pesquisa Acústica da Knowles Eletronics (*Knowles Electronics Manikin for Acoustic Research*)

SBV – Sistema Binaural Virtual

Plano XY – Plano horizontal

Plano XZ – Plano vertical lateral

Plano YZ – Plano vertical frontal

Referências Bibliográficas

ALGAZI, V. Ralph. **CIPIC/IDAV Interface Laboratory - University of California**. Disponível em: <http://interface.cipic.ucdavis.edu> Acessado em: 01/03/2010;

ANDERSON, Jeffrey. **Building a Binaural Dummy-Head**. Disponível em: <http://digdagga.com/dummy/index.html>. Acessado em: 03/03/2010;

CAMPBELL, Douglas R.; PALOMAKI, Kalle J.. **Roomsim, a MATLAB Simulation of "Shoebox" Room Acoustics for use in Teaching and Research**. Disponível em: <http://media.paisley.ac.uk/~campbell/Roomsim/>. Acessado em: 03/03/2010;

CHENG, Corey I.; WAKEFIELD, Gregory H.. Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space. **J Audio Eng Soc**, Vol 49, No 4, p. 231-249, Abril 2001;

GARDNER, Bill; MARTIN, Keith. **HRTF Measurements of a KEMAR Dummy-Head Microphone**, 1994. Disponível em: <http://sound.media.mit.edu/resources/KEMAR.html>. Acessado em: 03/03/2010;

GARDNER, William G.. **3D Audio and Acoustic Environment Modeling**. Wave Arts Inc., Março, 1999;

JANUS, S.. **Audio in the 21st Century**. Intel Press, Maio, 2004;

KIRKEBY, Ole. **Transparent stereo widening algorithm for loudspeakers**, 2005. Disponível em: <http://www.freepatentsonline.com/6928168.html> Acessado em: 09/03/2010;

LATHI, B. P.. **Sinais e Sistemas Lineares**, Editora Bookman, 2007;

SHENOI, B.A.. **Introduction to Digital Processing and Filter Design**, John Wiley & Sons Inc.. Hoboken, New Jersey, 2006. 113 p.