

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
CURSO DE CIÊNCIA DA COMPUTAÇÃO

GABRIEL GUOLO LEFFA

Ecolocalização em Ambiente Virtual

Trabalho de Graduação.

Prof. Dr. Anderson Maciel

Porto Alegre, julho de 2013.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos Alexandre Netto

Vice-Reitor: Prof. Rui Vicente Oppermann

Pró-Reitor de Graduação: Prof. Sérgio Roberto Kieling Franco

Diretor do Instituto de Informática: Prof. Luís da Cunha Lamb

Coordenador do Curso de Ciência da Computação: Prof. Raul Fernando Weber

Bibliotecária-Chefe do Instituto de Informática: Beatriz Regina Bastos Haro

AGRADECIMENTOS

Agradeço primeiramente ao meu orientador, Professor Anderson Maciel, pela paciência, ensinamentos e motivação para realizar esse trabalho.

Também agradeço a todas as pessoas que emprestaram seu tempo para realizar os exercícios propostos nos testes .

Agradeço também a minha família, em especial à minha mãe, Vera Regina Guolo Leffa, e ao meu pai, Márcio Joel Leffa, por todo apoio e suporte durante essa jornada.

Por último, agradeço à UFRGS e ao Instituto de informática pela estrutura disponibilizada.

SUMÁRIO

LISTA DE FIGURAS.....	5
LISTA DE TABELAS	7
RESUMO.....	8
ABSTRACT	9
1 INTRODUÇÃO	10
1.1 Estrutura do Trabalho	10
2 CONCEITOS E TRABALHOS RELACIONADOS.....	12
2.1 Som.....	12
2.1.1 Eco e Reverberação	12
2.2 Áudio Digital	12
2.2.1 Taxa de Amostragem.....	13
2.2.2 Estéreo	13
2.2.3 Arquivos Wave	13
2.3 Modelo de Reflexão de Phong.....	14
2.4 Modelos de Rendering de Som	16
2.4.1 Ray Tracing	16
2.4.2 Fonte Imagem.....	17
2.4.3 Radiosidade Com Trocas Variando no Tempo.....	18
2.5 Orientação e Locomoção por Áudio	18
2.5.1 Ray Tracing	19
3 MODELAGEM E IMPLEMENTAÇÃO	20
3.1 Modelagem	20
3.1.1 Shading	22
3.2 Informações de Entrada.....	23
3.2.1 Fonte de Som	23
3.2.2 Receptores	23
3.2.3 Arquivo de Áudio	23
3.2.4 Ambiente 3D.....	24
3.2.5 Volume	24
3.3 Processamento.....	24
3.3.1 Interpolação da Posição dos Vértices	25
3.3.2 Redução da Energia e Atraso.....	26
3.3.3 Distribuição de Energia Entre os Caminhos.....	27
3.3.4 Caminhos de Nível 2	28
3.3.5 Caminhos de Nível 1	28
3.3.6 Caminhos de Nível 0	29
3.4 Saída.....	29
3.4.1 Subrotina Saída.....	29
3.4.2 Escrita no Buffer ds Saída	30
4 AVALIAÇÃO.....	32

4.1	Avaliação de Performance	32
4.2	Avaliação com Usuários	33
4.2.1	Primeiro Teste.....	33
4.2.2	Segundo Teste.....	35
4.2.3	Terceiro Teste.....	36
4.2.4	Quarto Teste	38
4.2.5	Questionários	38
4.2.6	Hipóteses	39
5	RESULTADOS E DISCUSSÕES	40
5.1	Avaliando as Capacidades dos Usuários	40
5.2	Discussão Sobre a Percepção de Objetos Próximos	40
5.3	Discussão Sobre a Orientação e Locomoção	44
5.4	Resposta Sonora.....	44
5.5	Revisando as Hipóteses	45
6	CONCLUSÃO.....	47
	REFRÊNCIAS	48

LISTA DE FIGURAS

Figura 2.1: Comparação entre uma onda contínua e sua versão digital.	13
Figura 2.2: Equação do modelo de Phong.	14
Figura 2.3: Vetores utilizados no modelo de Phong.	15
Figura 2.4: Resultados para alfa igual à 1 (direita) e 20 (esquerda).	15
Figura 2.5: Da esquerda para direita: Flat, Gouraud e Phong Shading.	16
Figura 2.6: Reflexões com diferentes coeficientes de difusão. $a=0$, $b=0.2$, $c=1$. (RINDEL, 2000).	17
Figura 2.7: Formação de novas fontes de som através do espelhamento da original (GIBBS; VIVEIROS, 2000).	17
Figura 2.8: Visualização da propagação da onda, cada imagem representa um instante de tempo. (TSINGOS; GASCUEL, 1998).	18
Figura 2.9: Composição do sistema de treinamento. (a) Diagrama de blocos. (b) CG. (c) Fotografia do sistema. (SEKI; SATO, 2011).	19
Figura 3.1: Equação de Phong adaptada para uso com áudio.	20
Figura 3.2: Representação visual dos vetores da equação da Figura 3.1.	20
Figura 3.3: Equação da Figura 3.1 estendida para ser usada com dois pontos em sequência.	21
Figura 3.4: Representação visual dos vetores utilizados na equação da Figura 3.3.	22
Figura 3.5: Equação para o cálculo do som que parte direto a fonte para o receptor.	22
Figura 3.6: Subdivisão de um triângulo de acordo com fatores 1, 2, 3 e 4.	23
Figura 3.7: Execução do algoritmo encontrando novos pontos em um triângulo arbitrário. DIV igual à 2.	25
Figura 3.8: Representação visual dos elementos utilizados no algoritmo 3.	26
Figura 3.9: As duas situações validadas pelos testes das linhas 8 e 10, o primeira situação é evitada caso a face não seja definida como tendo dois lados, a segunda situação é sempre evitada.	27
Figura 3.10: Processo de escrita no buffer de saída. Da direita para a esquerda, as tabelas significam: buffer de entrada, buffer intermediário, buffer de saída antes das escritas e buffer de saída após as escritas.	31
Figura 4.1: Representação visual de uma cena no teste. Na imagem da esquerda, as esferas vermelhas representam os receptores, enquanto a branca, é a representação visual da fonte. A imagem da direita é a visualização disponível ao usuário, os receptores virtuais ficam nos limites laterais externos dessa imagem.	34
Figura 4.2: Representação visual do segundo teste com uma parede à direita. As esferas douradas representam os receptores, enquanto a branca representa a fonte de som.	35
Figura 4.3: Medidas utilizadas no posicionamento dos receptores e da fonte de som nesse teste.	36

Figura 4.4: Na imagem à esquerda, as esferas douradas representam os receptores, e a esfera branca representa a fonte de som. Na visualização disponível ao usuário (direita), apenas a representação da fonte é preservada, indicando o limite frontal do conjunto controlável.....	37
Figura 4.5: Corredores gerados aleatoriamente para o terceiro teste, a esfera vermelha indica a posição inicial do conjunto controlável pelo usuário.....	38
Figura 5.1: Respostas dos usuários ao questionário Q2. O eixo horizontal indica o nível de dificuldade que os usuários tiveram. O eixo vertical indica quantas vezes cada dificuldade foi escolhida. As dificuldades estão agrupadas aos pares para melhor visualização do resultado nos gráficos. A lista a direita de cada gráfico também indica quantas vezes cada resposta foi escolhida, mas não agrupando as respostas.....	42
Figura 5.2: Respostas do questionário Q1. O eixo horizontal indica as dificuldades escolhidas pelos usuários agrupadas aos pares. O eixo vertical indica o número de vezes que cada resposta foi escolhida. A lista a direita de cada gráfico também indica quantas vezes cada nível de dificuldade foi escolhido, mas não agrupando as respostas.....	43
Figura 5.3: Distribuição dos erros nas seções 1 e 3.....	43
Figura 5.4: Gráficos com tempos de resposta para a situação ilustrada à esquerda. O gráfico superior representa o canal da esquerda, enquanto o inferior representa o da direita. O eixo horizontal representa o tempo em que uma “partícula de som” chega ao receptor virtual, enquanto o eixo vertical representa a energia com que essa partícula chegou.....	45

LISTA DE TABELAS

Tabela 4.1: Medição de performance para os níveis 0 e 1.	32
Tabela 4.2: Medição da performance para o nível 2..	32
Tabela 4.3: Ordem das cenas em cada seção, considerando a distância entre a fonte de som e sua representação visual.	34
Tabela 5.1: Respostas dadas pelos usuários ao primeiro teste	40
Tabela 5.2: Resultados do primeiro teste. Os usuários estão na mesma ordem da Tabela 5.2. Os tempos estão medidos em segundos.	41
Tabela 5.2: Número de vezes em que cada som foi reproduzido por cada usuário nas seções 3 e 4.	41

RESUMO

Morcegos e golfinhos utilizam uma técnica chamada ecolocalização para perceber o mundo a sua volta. Esta técnica é pouco utilizada por seres humanos, que baseiam quase toda sua orientação em informações visuais. Em humanos, a aplicação de ecolocalização pode ter grande utilidade para pessoas com deficiência visual tanto na interação com o mundo real quanto com ambientes virtuais.

Neste trabalho propõe-se um método para treinar, desenvolver e avaliar essa capacidade em pessoas leigas nesta técnica. Para isso foi desenvolvido um modelo de rendering de som inspirado no modelo de iluminação de Phong. A técnica de Phong foi escolhida devido ao fato de ser fisicamente coerente e eficiente para poder ser utilizada em tempo real.

O modelo foi implementado e avaliado com usuários em um ambiente virtual simulado. Foram projetados experimentos para atestar a capacidade do modelo em prover informação de orientação mesmo para pessoas que não receberam treinamento específico. Os testes envolveram a realização de tarefas simples, como indicar a posição de uma fonte de som e indicar em que direção se encontra uma parede próxima.

Os resultados mostram que a maioria das pessoas na maioria das vezes consegue identificar a posição de fontes de som diretas e refletidas por objetos, confirmando as principais hipóteses levantadas. Em alguns casos, no entanto, percebe-se que uma fase de aprendizado é necessária para uma orientação mais natural. Testes mais completos e com mais usuários seriam necessários para confirmar hipóteses que permaneceram inconclusivas a partir dos dados coletados nas condições atuais.

Palavras-chave: Ecolocalização, Phong, Som 3D, Acessibilidade.

Echolocation in Virtual Environments

ABSTRACT

Bats and dolphins use the echolocation technique to perceive and interact with their living environments. Human beings, who rely essentially on visual information for orientation, do not explore such technique to the extent we believe they could. In humans it could be very useful for people with visual impairments to orient both in the real and in virtual environments.

In this work we propose a method to train, develop and evaluate the ability for echolocation in people not familiar with such technique. We then developed an acoustic model inspired by Phong's lighting model. Phong's technique was chosen for being physically coherent and efficient to be used in real time.

We implemented and evaluated the model with users in a simulated virtual environment. Experiments have been designed to assess the model's ability to provide orientation information even for people that have not received any specific training. The tests involved simple tasks, such as to indicate the position of a sound source or tell the location of the closest wall.

Results show that most people are able to identify, most of the times, the position of direct and reflected sound sources. They are then able to locate objects, confirming most of our hypotheses. In some situations, however, a learning step is still required for a natural orientation. More complete studies with a larger number of users would be necessary to confirm the hypotheses that remained inconclusive in the analysis of the data collected with the current conditions.

Keywords: Echolocation, Phong, 3D Sound, Accessibility.

1 INTRODUÇÃO

No mundo animal, existem muitas espécies com capacidades visuais extremamente limitadas ou inexistentes. Em geral, a evolução encontrou outro caminho para que tais criaturas percebam o mundo a sua volta. Esses animais emitem sons variados, e através de aparelhos auditivos extremamente sensíveis, percebem no retorno desses sons características físicas do ambiente que os cerca.

Surpreendentemente, essa capacidade existe também em seres humanos. Mesmo que mais limitada em relação a mamíferos como morcegos e golfinhos, algumas pessoas com deficiências visuais severas desenvolveram essa técnica. No cérebro dessas pessoas, a parte que seria destinada a processar a visão fica ativa enquanto elas percebem o ambiente pelo som (THALER; ARNOTT; GOODALE, 2011). Entretanto, hoje se sabe que tal técnica não é apenas um dom natural de algumas pessoas cegas, ela também pode ser praticada e aperfeiçoada, até mesmo por pessoas sem nenhuma deficiência visual.

Uma das pessoas mais conhecidas pelo uso da ecolocalização é Daniel Kish. Kish ficou cego ainda aos 13 meses de idade e ainda na infância aprendeu a utilizar a língua para emitir estalos e perceber o ambiente através do eco desses sons, no que ele chama de "Flash Sonar". Ele também mantém uma organização sem fins lucrativos, que se dedica a ensinar essa técnica e promover atividades envolvendo pessoas cegas.

Porém, na área da computação existem poucos trabalhos relacionados à ecolocalização humana. Esses que poderiam ajudar a desenvolver mecanismos que facilitem o uso dessa técnica e provar hipóteses que seriam impossíveis de serem verificadas em ambiente real.

O objetivo desse trabalho é prover um mecanismo virtual para que se possa ajudar a treinar, desenvolver e analisar a capacidade dos seres humanos de utilizar a técnica de ecolocalização. Para isso será proposto um modelo de rendering de som, e um conjunto de testes simples que visam avaliar a capacidade dos usuários e validar hipóteses de uso do sistema para facilitar o uso dessa capacidade.

1.1 Estrutura do Trabalho

No capítulo seguinte serão apresentados alguns conceitos necessários para a compreensão desse trabalho. Também serão apresentados alguns trabalhos relacionados a rendering de som e orientação por áudio

No terceiro capítulo será definido o modelo de rendering de som utilizado por esse trabalho para a implementação dos casos de teste.

No quarto capítulo serão descritos todos os experimentos desenvolvidos, alguns dos quais realizados. Também nesse capítulo, são apresentadas as hipóteses que esse trabalho tenta validar.

Por fim, a apresentação dos resultados dos testes com usuários e discussões desses resultados (capítulo 5), e então a conclusão (capítulo 6).

2 CONCEITOS E TRABALHOS RELACIONADOS

2.1 Som

O som é uma onda mecânica que se propaga em meio material. É perceptível ao ser humano através do sentido da audição, no intervalo de frequências de aproximadamente 20Hz à 20.000Hz.

2.1.1 Eco e Reverberação

Eco e reverberação são reflexos de som que chegam a um ouvinte. Se o ouvinte é capaz de distinguir o som original de seu reflexo, esse reflexo ganha o nome de eco. Caso contrário, os reflexos ganham o nome de reverberação.

Para um ser humano, é perceptível a diferença entre um som original e seu reflexo se ambos estiverem separados por mais de 0,1 segundos.

2.2 Áudio Digital

Como o som na natureza é uma onda contínua, para representar essa onda em meio digital é necessário que ela seja discretizada. O som digital é representado por uma sequência numérica, onde cada valor dessa sequência é uma amostra da amplitude da onda em um instante de tempo. Como é possível observar na Figura 2.1, tanto a informação de tempo (eixo horizontal), quanto a informação de amplitude (eixo vertical) foram discretizadas.

Essa sequência sozinha, entretanto, é insuficiente para reproduzir o áudio. Também são necessárias informações como o número de amostras que ocorrem em um determinado instante de tempo, o tamanho de cada amostra em bits e número de canais no buffer.

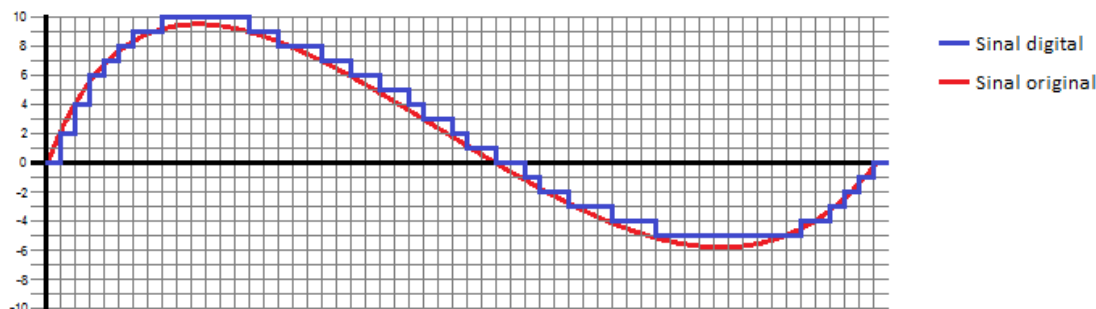


Figura 2.1: Comparação entre uma onda contínua e sua versão digital.

2.2.1 Taxa de Amostragem

Uma das informações citadas anteriormente é o número de amostras que ocorrem em um determinado instante de tempo. A essa informação é dado o nome de taxa de amostragem. É medida em Hertz.

Segundo o Teorema de Amostragem presente em (SHANNON, 1949), para se garantir a recuperação de toda a informação de frequência de um sinal analógico original, a taxa de amostragem na versão digital deve ser o dobro da frequência máxima encontrada no sinal original.

Por isso, normalmente sons digitais são amostrados a taxas superiores a 40.000Hz, visto que o ouvido humano não é capaz de registrar frequências superiores a 20.000Hz.

2.2.2 Estéreo

Um som estéreo é aquele que possui dois canais de áudio que são reproduzidos de maneira independente em um par de fones de ouvido ou auto falantes. O objetivo de se utilizar dois canais de som é dar ao ouvinte a impressão de que os sons ouvidos estão ocorrendo em posições diferentes. Isso é possível devido ao fato do cérebro humano, para localizar uma fonte de som, utilizar, dentre outros fatores, a diferença de atraso e volume em que um som chega a cada ouvido.

Para gravar som em dois canais, dois microfones são posicionados a uma distância similar a distância entre os ouvidos humanos. O áudio gravado em cada microfone é então atribuído a cada um dos canais de direita ou esquerda. A maneira mais comum de armazenar essa informação é separar as amostras de cada canal aos pares. Portanto em um buffer de áudio estéreo, as amostras ficam posicionadas da seguinte forma: primeira amostra da direita; primeira amostra da esquerda; segunda amostra da direita; segunda amostra da esquerda; etc.

2.2.2 Arquivos Wave

Um arquivo wave é uma das maneiras mais simples de armazenamento de áudio em disco. Seu formato "é um subconjunto da especificação RIFF da Microsoft para armazenamento de arquivos multimídia" (WILSON, 2003), consistindo basicamente de um cabeçalho com as informações básicas para interpretar o buffer que segue. Tipicamente esse buffer é armazenado sem nenhum tipo de compressão.

2.3 Modelo de Reflexão de Phong

O modelo de reflexão de Phong é um modelo de iluminação usado em computação gráfica para calcular a iluminação em um ponto. É um modelo local, isso significa que a iluminação calculada em um ponto não sofre influência de outros pontos na cena, apenas da posição da(s) fonte(s) de luz e da câmera virtual.

$$I_p = k_a i_a + \sum_{m \in \text{lights}} (k_d (\hat{L}_m \cdot \hat{N}) i_{m,d} + k_s (\hat{R}_m \cdot \hat{V})^\alpha i_{m,s}).$$

Figura 2.2: Equação do modelo de Phong.

É importante salientar que na equação da Figura 2.2, a contribuição de Bui Tuong Phong refere-se apenas a parte especular do modelo. Porém, assim como em grande parte da literatura sobre o assunto, sempre que este trabalho fizer referência ao Modelo de Phong, estará tratando de toda a equação.

Os significados dos elementos citados na Figura 2.2 estão listados abaixo. Na Figura 2.3 se encontra uma representação visual de alguns desses elementos.

- i e k : quantidade de luz incidente e refletida respectivamente.
- Sufixos a , d , s e m : luz ambiente, luz difusa, luz especular e índice da fonte de luz. Respectivamente.
- L : vetor que aponta para a fonte de luz partindo da superfície.
- N : vetor normal à face
- R : vetor L refletido sobre a normal
- V : vetor que aponta para a câmera virtual
- O acento circunflexo indica que o vetor está normalizado
- Alfa: Controla o quão específico é o reflexo especular do objeto, quanto maior, mais específico (mais próximo de um espelho), a Figura 2.4 ilustra o comportamento dessa constante.

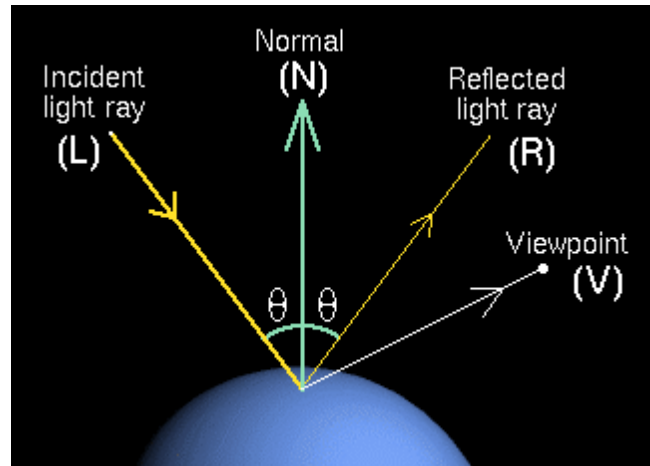


Figura 2.3: Vetores utilizados no modelo de Phong.

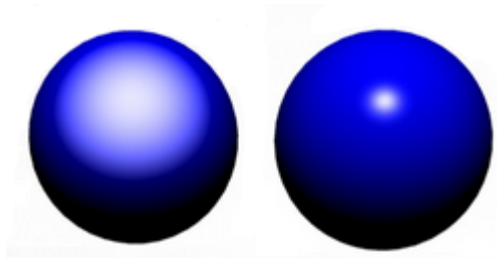


Figura 2.4: Resultados para alfa igual à 1 (esquerda) e 20 (direita).

O modelo de Phong é utilizado apenas para calcular a iluminação de um ponto. Aplicar os resultados obtidos nas faces depende do método de shading escolhido. Estão listados abaixo, os métodos de shading mais comuns. Na Figura 2.5 se encontra um exemplo do resultado de cada método.

- Flat Shading: É calculado um ponto por face, o resultado é utilizado por todos os pontos visíveis da mesma.
- Gouraud Shading: São calculados nos vértices das faces, cada um com uma normal. Os valores nos demais pontos da face são uma interpolação linear desses resultados.
- Phong Shading: É calculado para cada ponto visível na face, as normais são uma interpolação das normais dos vértices.

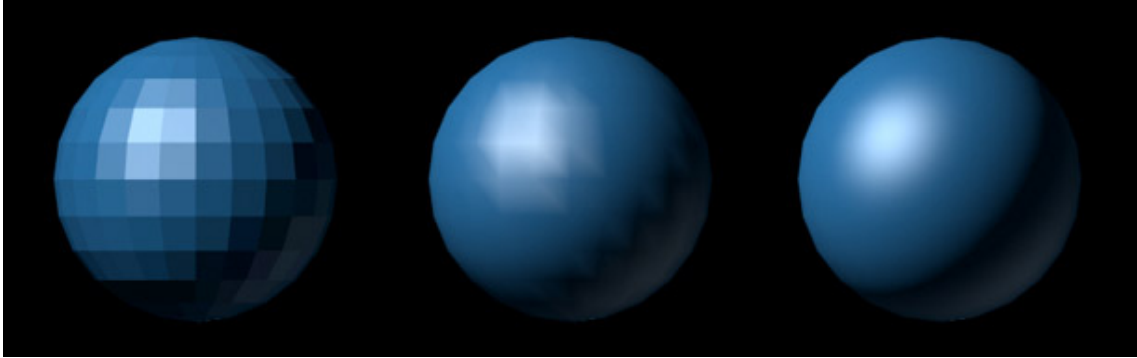


Figura 2.5: Da esquerda para direita: Flat, Gouraud e Phong Shading.

Neste trabalho, foi feita a suposição de que o mesmo princípio pode ser aplicado à um modelo sonoro, que será detalhado no capítulo 3.

2.4 Modelos de Rendering de Som

A grande maioria dos modelos de rendering de som são extremamente similares a modelos e técnicas de computação gráfica visual. As maiores diferenças estão no fato de existir a necessidade de calcular o tempo de propagação nos modelos de áudio, enquanto em imagens esse tempo quase sempre é considerado insignificante. Também no áudio, tipicamente a parcela de reflexão especular é mais significativa (TSINGOS; GASCUEL, 1998).

Outra diferença é o fato das imagens serem ricas no espaço e pobres no tempo, enquanto com o som acontece o inverso. Ou seja, para imagens, a cada instante de tempo é necessário obter vários pontos (pixels), mas 24 imagens por segundo são suficientes para uma animação com fluidez. Já para o som, são necessários apenas dois valores por instante de tempo, um para cada ouvido, entretanto, esse instante de tempo é menor que um milésimo de segundo, visto que, tipicamente, para uma reprodução de áudio com qualidade são necessárias mais de 40.000 amostras por segundo.

2.4.1 Ray Tracing

O método consiste em projetar diversas partículas em diferentes direções partindo da fonte de som. Cada vez que uma partícula encontra uma face, a mesma é refletida. Nesse método, cada face tem um coeficiente de absorção e um coeficiente de difusão. O coeficiente de absorção indica a quantidade de energia que a partícula perderá ao refletir na face. O coeficiente de difusão indica o quão simétricos são os ângulos dos raios refletidos. O comportamento do coeficiente de difusão é ilustrado na Figura 2.6. A partícula também perderá energia proporcionalmente ao quadrado da distância percorrida

No Ray Tracing, para cada receptor, é estabelecida uma área ou volume no entorno de sua posição. Todas as partículas que entrarem nesse volume serão consideradas no resultado final. Abordagens alternativas utilizam cones ou pirâmides para traçar o caminho das partículas. Nesses casos, não é necessário o volume entorno do ouvinte.

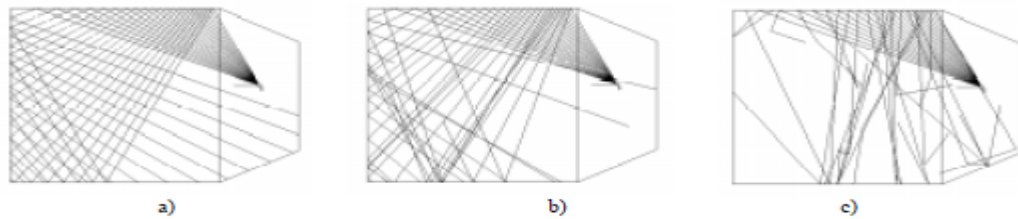


Figura 2.6: Reflexões com diferentes coeficientes de difusão. $a=0$, $b=0.2$, $c=1$. (RINDEL, 2000).

2.4.2 Fonte Imagem

Este método encara cada superfície como um espelho para a fonte de som. Ao invés de se ser definido o caminho do som através de um ambiente, a fonte de som é espelhada em cada superfície, criando novas fontes, que são as imagens de primeiro grau. Essas podem ser espelhadas novamente, criando as imagens de segundo grau, e assim sucessivamente.

Assim, como pode ser visto na Figura 2.7, o número de fontes imagens cresce exponencialmente conforme o número do grau de reflexões. Ainda existem outras desvantagens. Em uma sala mais complexa, existirão muitas imagens de segunda ordem inválidas, portanto, existe a necessidade de validar as imagens nesse caso. Também neste método, só é possível simular superfícies com coeficiente de difusão igual a zero.

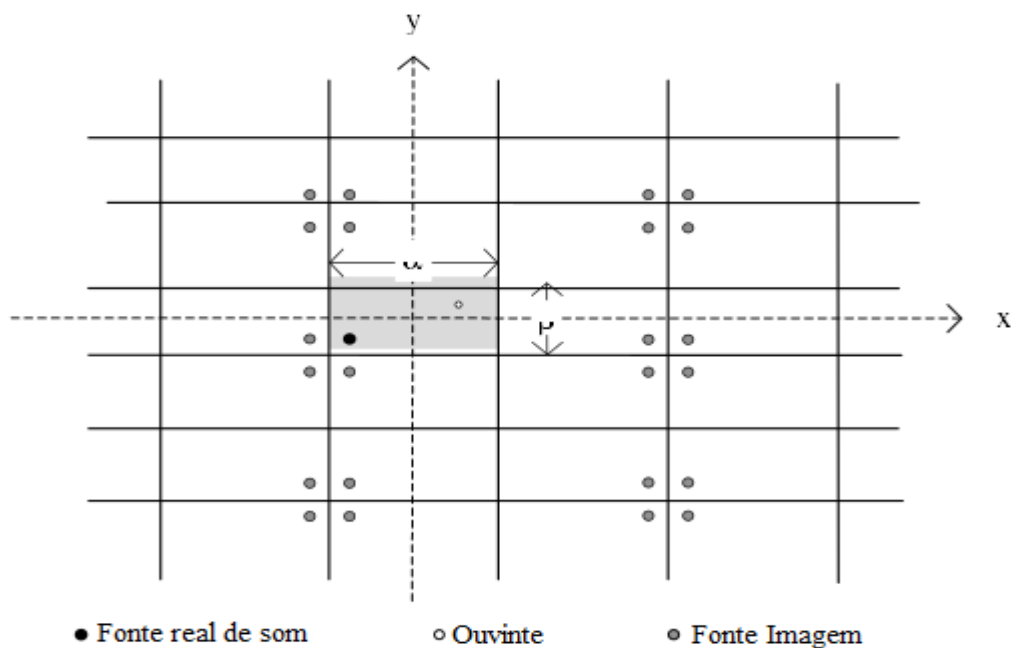


Figura 2.7: Formação de novas fontes de som através do espelhamento da original (GIBBS; VIVEIROS, 2000).

2.4.3 Radiosidade Com Trocas Variando no Tempo

No modelo de radiosidade, cada superfície é dividida em partes menores, e para cada par dessas divisões é calculado um coeficiente de visualização, que indica, o quão visível é uma parte em relação a outra. Feito isso, em cada divisão são adicionadas as informações das outras divisões conforme o coeficiente calculado. Esse processo é repetido diversas vezes até a cena resultante convergir para um resultado ou um valor de iterações limites for atingido.

Esse método costuma ser utilizado para superfícies difusas, o que não é o mais indicado para um modelo sonoro. Porém, em (TSINGOS; GASCUEL, 1998) os autores apresentam o modelo de radiosidade para ser utilizado com áudio. Conforme desenvolvem o modelo, os autores mantêm um paralelo com o modelo visual de radiosidade, levando em conta que as diferenças citadas no início desse capítulo. Na Figura 2.8 é exibido um exemplo desse modelo em funcionamento. Da segunda para a terceira imagem, é possível notar a onda refletindo na superfície inferior.

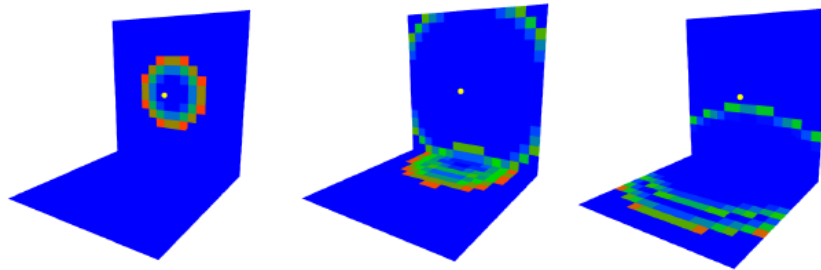


Figura 2.8: Visualização da propagação da onda, cada imagem representa um instante de tempo. (TSINGOS; GASCUEL, 1998).

2.5 Orientação e Locomoção por Áudio

Treinamentos para aperfeiçoar as capacidades de orientação e locomoção são comuns para pessoas cegas. Um desses treinamentos consiste em aperfeiçoar a parte auditiva para a questão da orientação, através de sons ambientes, como um veículo passando ou portas se abrindo ou fechando. Esse treinamento tipicamente é feito em um ambiente real, o que pode ser perigoso e estressante para o aluno.

Em (SEKI; SATO, 2011) foi desenvolvido um sistema de treinamento usando realidade virtual para simular um ambiente facilmente programável pelo orientador do treinamento. Como é possível ver na Figura 2.9, esse sistema é composto por 10 fontes de som, 10 processadores 3D, mixer, fones de ouvido e um computador.

O objetivo desse trabalho é prover um método de treino sem riscos e menos estressante para o usuário, que pode ser utilizado como etapa anterior ao treinamento em ambiente real. Uma desvantagem desse método de treinamento são os sensores utilizados, eles não têm longo alcance, por isso, a área que o aluno pode se deslocar é bastante limitada em torno dos equipamentos utilizados.

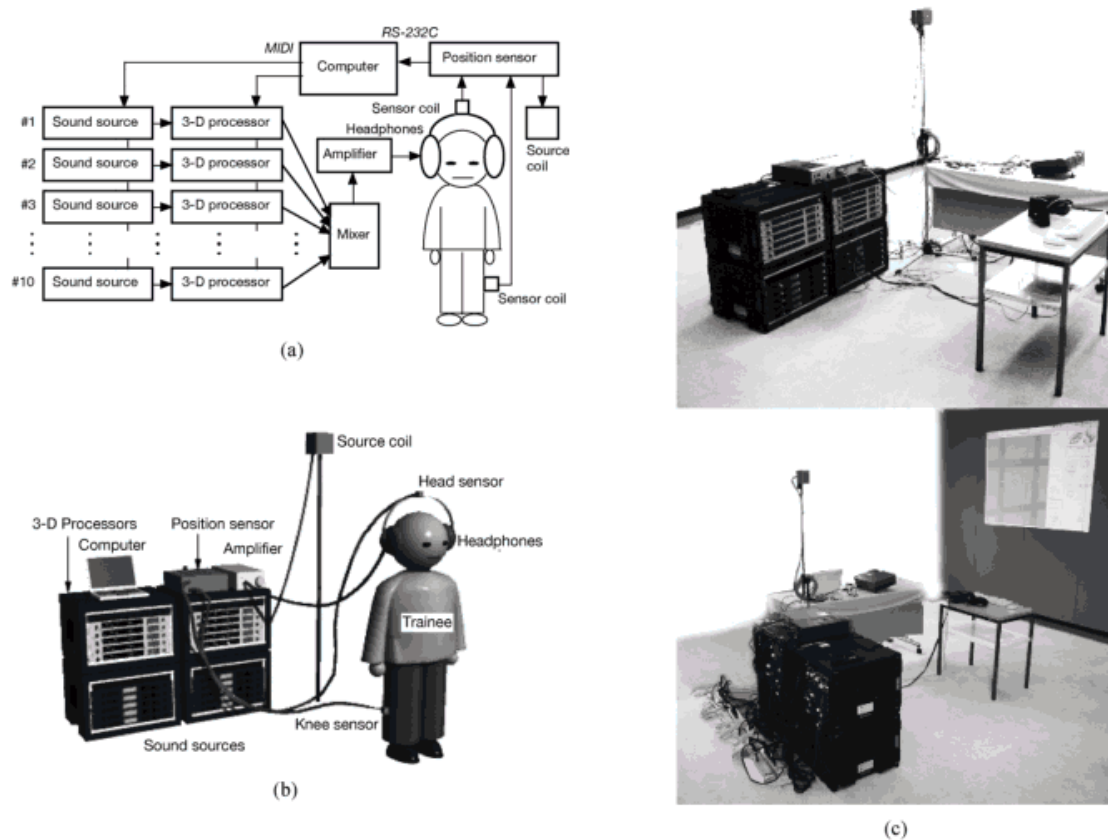


Figura 2.9: Composição do sistema de treinamento. (a) Diagrama de blocos. (b) CG. (c) Fotografia do sistema (SEKI; SATO, 2011).

2.5.1 Função de Transferência da Cabeça

Em (SEKI; SATO, 2011), também é feito uso de um conceito importante na questão da localização de sons no espaço pelo homem, a função de transferência da cabeça.

Existem dois tipos de pistas que ajudam o ser humano a localizar a posição de uma fonte de som no espaço. A primeira são as diferenças (intensidade e tempo) do som entre um ouvido e outro. A segunda é resultado das alterações sofridas pelo som ao entrar em contato com partes da anatomia humana, como orelhas e cabeça. A função de transferência da cabeça é uma função que traduz um som partindo de uma determinada posição no som que seria ouvido por uma pessoa após passar pelas alterações referentes à anatomia humana.

3 MODELAGEM E IMPLEMENTAÇÃO

O objetivo do modelo é prover um método que possa ser usado para simular um ambiente acústico fisicamente coerente com a realidade e em tempo real. A abordagem proposta é que ele rode dentro de um laço infinito, a cada volta do laço, é processado um trecho do que será ouvido pelo usuário do programa.

Para a implementação deste trabalho, não foi utilizada nenhuma biblioteca de áudio em alto nível, apenas o arquivo de cabeçalho do Windows mmsystem, que prove a comunicação do software com o driver da placa de som. Essa foi feita porque para a implementação apresentada aqui, apenas funções básicas de manipulação de buffers de áudio são necessárias.

3.1 Modelagem

O modelo descrito a seguir é basicamente o modelo de Phong adaptado para ser usado no processamento de áudio. Portanto, é preciso levar em conta as diferenças citadas na seção 2.3. A Figura 3.1 ilustra uma das equações utilizadas por esse modelo, enquanto a Figura 3.2 exibe uma representação visual para os vetores usados nessa equação.

$$R \left[t + \frac{(\|L\| + \|V\|)}{sp} \right] = \frac{en(d(\hat{L} \cdot \hat{N}) + s(\hat{R} \cdot \hat{V})^\alpha) F[t]}{(1 + \|L\|^2)(1 + \|V\|^2)}$$

Figura 3.1: Equação de Phong adaptada para uso com áudio.

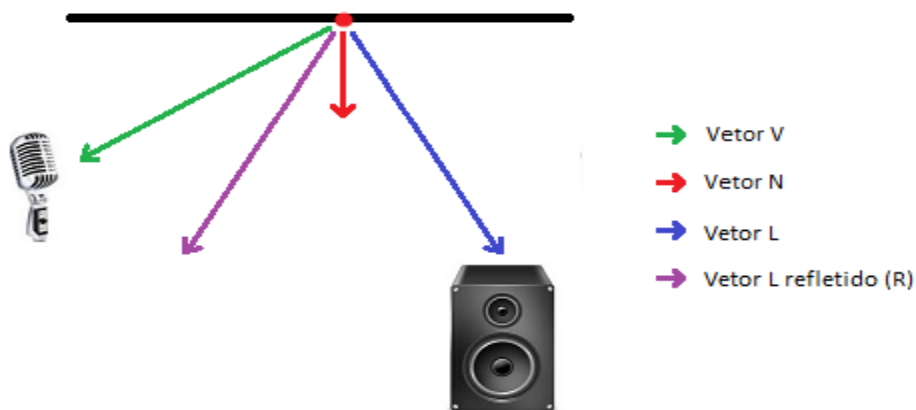


Figura 3.2: Representação visual dos vetores da equação da Figura 3.1.

Em relação à equação da Figura 2.2, os novos elementos dessa equação são os seguintes:

- R: Resultado do processamento. O valor nos colchetes é o tempo transcorrido desde o início da reprodução.
- t: Tempo transcorrido desde o início da reprodução na fonte de som.
- sp: Velocidade do Som.
- en: Energia com que esta sendo reproduzido o áudio na fonte de som.
- d: Coeficiente de difusão.
- s: Coeficiente de especularidade.
- F: Fonte de som.

Além de ter sido incluída a informação de tempo, em relação a equação original também foi incluído a perda de energia em devido à distância percorrida, como pode ser visto no denominador do lado direito da equação. Outras diferenças, são a ausência de parcela ambiente e dos coeficientes de difusão e especularidade para a energia incidente.

Uma característica possível de se perceber por essa equação é o fato desse modelo ser focado no tempo, e não na frequência. Ou seja, o momento em que uma amostra de som é reproduzida na fonte pode ser diferente do momento da reprodução da amostra equivalente nos receptores. Já as frequências da fonte só vão diferir da dos receptores se alguma alteração ocorrer nas outras variáveis da equação durante a reprodução.

Outra característica citada na seção 2.3 é o fato da parcela especular no som ser mais significativa em relação à luz. Com isso, utilizar apenas coeficientes de especularidade mais elevados pode não ser suficiente para uma boa simulação. Então, foi incluída a possibilidade de se calcular o atraso e a energia resultantes do som refletido em uma sequência de dois pontos antes de chegar ao receptor.

$$R \left[t + \frac{(\|L_1\| + \|V_1\| + \|V_2\|)}{sp} \right] = \frac{en(d_1(\hat{L}_1 \cdot \hat{N}_1) + s_1(\hat{R}_1 \cdot \hat{V}_1)^{\alpha_1}) (d_2(\hat{L}_2 \cdot \hat{N}_2) + s_2(\hat{R}_2 \cdot \hat{V}_2)^{\alpha_2}) F[t]}{(1 + \|L_1\|^2)(1 + \|V_1\|^2)(1 + \|V_2\|^2)}$$

Figura 3.3: Equação da Figura 3.1 estendida para ser usada com dois pontos em sequência.

Na equação da Figura 3.3, a maioria dos elementos possui um sufixo 1 ou 2, que indica a qual ponto está ligada cada variável. Uma representação visual dos vetores dessa equação é exibida na Figura 3.4. Quase todos os elementos são os mesmos da equação da Figura 3.1, com exceção dos seguintes:

- V_1 : vetor que aponta para o segundo ponto, partindo do primeiro.

- L_2 : vetor que aponta para o primeiro ponto, partindo do segundo.

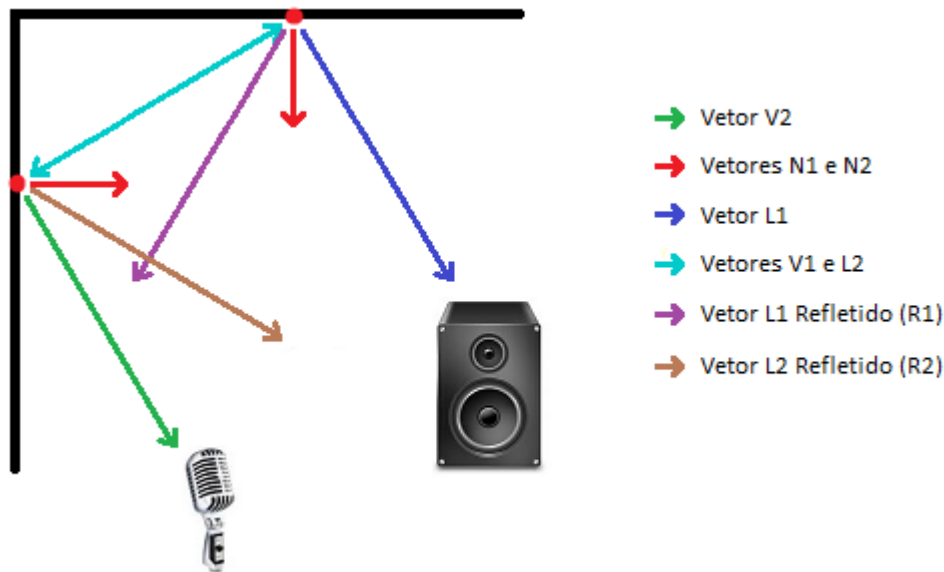


Figura 3.4: Representação visual dos vetores utilizados na equação da Figura 3.3.

Ainda é necessário calcular o som que chega direto aos receptores, não refletindo em nenhum ponto no seu caminho. Para isso é utilizada a equação presente na Figura 3.5. O vetor "V" representa o que aponta para o receptor, partindo da fonte.

$$R \left[t + \frac{\|V\|}{sp} \right] = \frac{F[t]en}{(1 + \|V\|^2)}$$

Figura 3.5: Equação para o cálculo do som que parte direto da fonte para o receptor.

3.1.1 Shading

O modelo descrito até aqui, assim como sua contraparte visual, também precisa estar associado a uma técnica de shading para ser aplicado. Como a técnica apresentada até aqui trabalha apenas no sistema de coordenadas do universo, não existe processo de rasterização, então os pontos nas faces são obtidos através de sua subdivisão. O número de subdivisões da face é arbitrário, a Figura 3.6 ilustra um triângulo dividido através dos fatores 1, 2, 3 e 4.

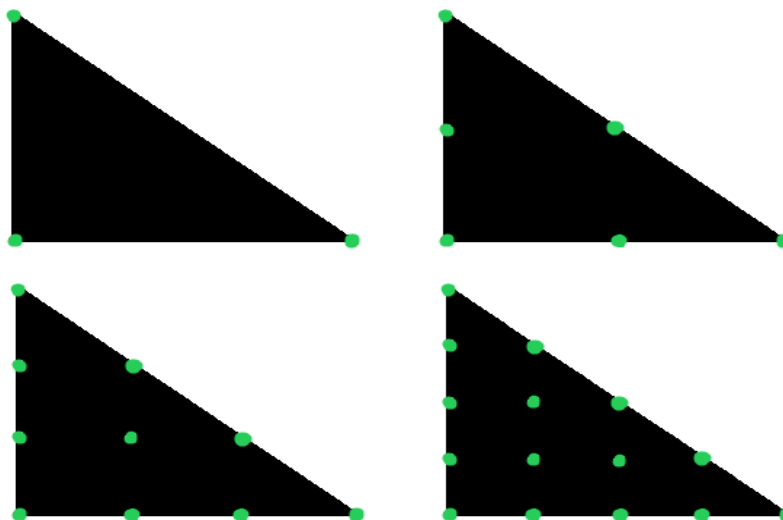


Figura 3.6: Subdivisão de um triângulo de acordo com fatores 1, 2, 3 e 4.

Para cada ponto encontrado, é aplicada a equação presente na Figura 3.1, utilizando a normal da face, enquanto que. Para cada par de pontos em faces diferentes, é aplicada a equação demonstrada pela Figura 3.3. As normais são as mesmas para cada face.

3.2 Informações de Entrada

3.2.1 Fonte de Som

A fonte de som é única. É definida por sua posição no espaço, através das coordenadas x , y e z , todas fornecidas em ponto flutuante. Uma vez que uma posição é definida, ela será usada para todos os cálculos até que seja informada uma nova posição.

3.2.2 Receptores

Para a recepção são utilizados dois receptores virtuais, um para cada canal de som estéreo. Esses receptores simulam microfones ou ouvidos humanos. Cada receptor é definido individualmente por sua posição no espaço e pelo canal a que essa posição corresponde (direita ou esquerda). Assim como para a fonte, as últimas posições fornecidas serão usadas para os cálculos.

3.2.3 Arquivo de Áudio

O arquivo fornecido deve ser do tipo wave descomprimido, com 16 bits por amostra e um ou dois canais. Em caso de um canal (mono), a parte dados do arquivo é copiada diretamente para um buffer. Em caso de dois canais, os dados devem ser convertidos para um canal, para isso, é feita a média aritmética entre os canais.

De todas as informações de entrada, esta é a única que não deve ser fornecida durante a execução do laço principal da simulação, uma vez que copiar buffers inteiros é uma operação muito custosa salvo em arquivos muito pequenos. Portanto, o arquivo é fornecido antes do início da simulação. O que se fornece durante a simulação é o comando para reproduzir o arquivo, que pode ser reproduzido mais de uma vez mas não se sobrepondo, uma reprodução só pode começar assim que a anterior estiver terminada.

3.2.4 Ambiente 3D

Formado por faces triangulares. Para definir uma face o programador da simulação deve informar:

- 3 vértices, cada um representado pelas coordenadas x, y e z em ponto flutuante
- Uma normal, também possuindo 3 coordenadas. A normal fornecida deve, necessariamente, ser um vetor unitário.
- Um coeficiente de absorção, utilizado para indicar quanto da energia incidente na superfície é absorvida ou perdida. Um indica perda total e zero indica ausência de perda.
- Um coeficiente de especularidade.
- Um valor indicando se a face é válida nos dois lados, ou apenas na direção apontada pela normal.

Cada face definida é adicionada a uma lista, portanto, a cada volta no laço principal a lista é esvaziada.

3.2.5 Volume

O volume é a energia inicial com que será dividida pelos diversos caminhos de som que partem da fonte. É um valor em ponto flutuante e pode ser definido a qualquer momento da simulação. Valores maiores que 1 nessa variável podem causar overflow no momento em que os resultados do processamento forem multiplicados pelas amostras do buffer de áudio de entrada. Detalhes dessa multiplicação serão apresentados na seção 3.4.

3.3 Processamento

Quase todo o processamento sobre as informações listadas até então corresponde à etapa a ser feita em cada volta do laço principal. Esse processamento consiste em calcular o atraso e a energia que chegam aos receptores após uma série de caminhos que “partículas de som” percorrem. Os caminhos são os seguintes:

- Caminhos de nível 2: São os caminhos que partem da fonte, passam por um ponto, por outro ponto em uma face diferente e chegam aos receptores
- Caminhos de nível 1: São os caminhos que partem da fonte, passam por um ponto em uma face e chegam aos receptores
- Caminhos de nível 0: São os caminhos que partem da fonte e chegam diretamente aos receptores.

Os caminhos estão listados acima na ordem em que são calculados, Isso é feito porque os resultados dos cálculos dessa etapa são somados à um buffer de energias e os caminhos de nível 2 têm mais chances de resultarem em valores menores. Como a soma é feita sobre números de ponto flutuante, somar valores menores primeiro é a melhor prática para reduzir as chances de perdas por erros de arredondamento.

Antes de calcular os caminhos, entretanto, a definição de uma face já contempla uma fase importante de processamento que deve ser detalhada.

3.3.1 Interpolação da Posição dos Vértices

Como o algoritmo opera sobre pontos em uma face, é preciso mais pontos do que apenas os vértices. Para tanto, cada triângulo é dividido em partes menores de acordo com um parâmetro inteiro positivo DIV. Esse valor será usado para dividir o triângulo em partes menores, quanto maior o valor, maior o número de divisões. O algoritmo 1 é executado após a definição de cada face e mostra como é obtida a localização dos pontos, enquanto a Figura 3.7 ilustra a execução do algoritmo em um triângulo arbitrário.

algoritmo 1 Interpolar posição dos vértices

passo1 = (vertice2 - vertice1)/DIV

passo2 = (vertice3 - vertice1)/DIV

para i = 0 enquanto i <= DIV

 para j = 0 enquanto j <= i

 ponto.posição = vertice1 + passo1 * DIV * i + passo2 * DIV * j

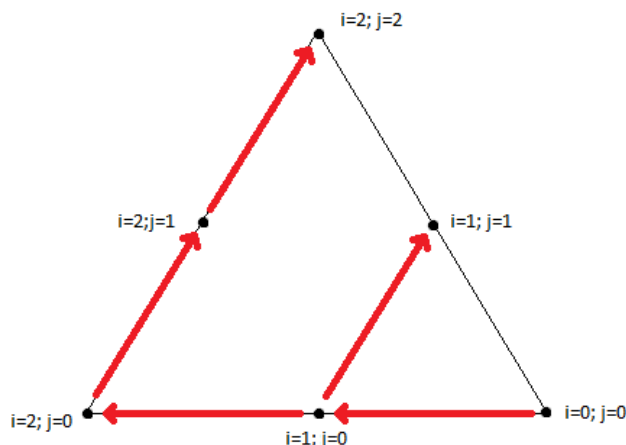


Figura 3.7: Execução do algoritmo encontrando novos pontos em um triângulo arbitrário. DIV igual à 2.

Ainda nessa etapa é necessário obter o número de vezes que o mesmo ponto se repete em outras faces. Para isso é utilizado o algoritmo 2:

algoritmo 2 Encontrando pontos repetidos

ponto.repetidos = 1

para i=0 enquanto i < faces.tamanho

 para j = 0 enquanto j < faces[i].pontos.tamanho

 se ponto.posição = faces[i].pontos[j].posição

 ponto.repetidos++

 faces[i].pontos[j]++

Manter o número de repetições de cada localização é importante para distribuir melhor a energia da fonte de som, mais detalhes serão apresentados a seguir. Após esse processamento, o ponto encontrado é armazenado em uma lista na sua respectiva face.

3.3.2 Redução de Energia e Atraso

Quando percorre um determinado caminho, o som perde sua energia de acordo com dois fatores, o quadrado da distância percorrida e a perda inerente à reflexão em uma face. Já o atraso é proporcional apenas à distância. O algoritmo 3 codifica a rotina "calcularCaminho", que é usada para calcular os caminhos nível 2 e 1. Essa rotina recebe como parâmetros a energia inicial, a posição de origem, uma face, um ponto nessa face e a posição de destino. Todos esses elementos estão representados visualmente na Figura 3.8. O retorno é uma dupla contendo a energia resultante e o atraso.

algoritmo 3 Subrotina calcularCaminho

```

01: calcularCaminho(energia, origem, f, p, dest)
02: pOrigem = origem - p.posicao
03: pDest = dest - p.posicao
04: atraso = (modulo(pDest)+modulo(pOrigem))/velocidadeDoSom
05: energia = (energia/(1+modulo(pOrigem)2))/(1+modulo(pDest)2)
06: pOrigem = normalizar(pOrigem)
07: pDest = normalizar(pDest)
08: se !f.doisLados e dotProduct(f.normal, pOrigem) < 0
09:   retornar(0,0)
10: se sinal(dotProduct(f.normal, pOrigem)) != sinal(dotProduct(f.normal, pDest))
11:   retornar(0,0)
12: energia = (energia*(1 - f. absorção))/p.repetidos
13: refOrigem = 2*dotProduct(rOrigem, f.normal)*normal - pOrigem
14: energia = energia*((1-f.especular)*dotProduct(f.normal, pOrigem) +
f.especular*dotProduct(pDest, refOrigem))
15: retornar(energia, atraso);

```

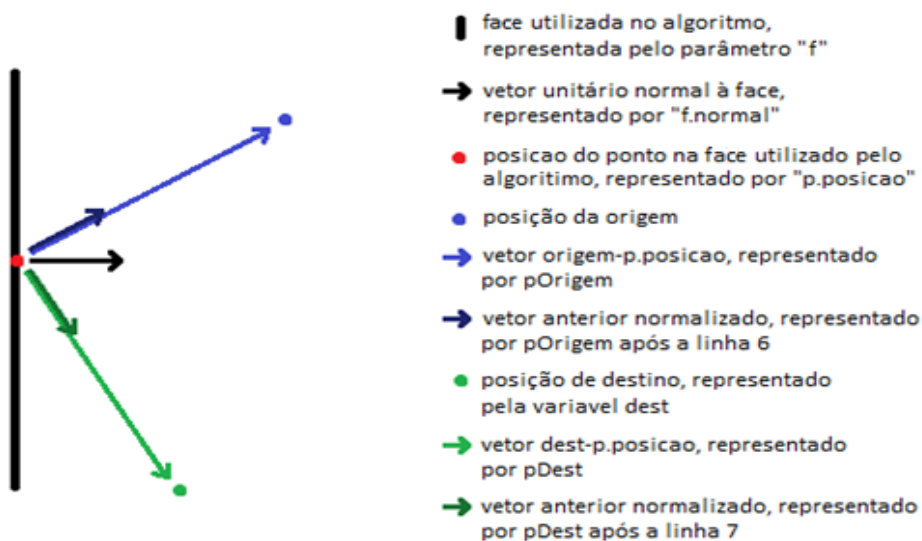


Figura 3.8: Representação visual dos elementos utilizados no algoritmo 3.

- Linhas 2 e 3: Obtenção de dois vetores que, partindo de p , apontam para origem e destino, respectivamente.
- Linhas 4 e 5: Cálculo do atraso e do decaimento da energia provocados na distância percorrida. A velocidade do som utilizada é medida em metros por tempo de amostra, um tempo de amostra é o tempo médio gasto para reproduzir uma única amostra de som.
- Linhas 6 e 7: Normalização dos vetores utilizados anteriormente.
- Linhas 8 e 11: Validações. A primeira verifica se a normal da face aponta para o lado oposto ao que se encontra o ponto de origem, nesse caso, o caminho calculado deve ser invalidado se a face for indicada como tendo um lado. A segunda verifica se o ponto de origem e o ponto de destino estão no mesmo lado em relação à face. A Figura 3.9 ilustra as situações evitadas por essas validações.
- Linha 12: Atualização da energia de acordo com o coeficiente de reflexão e o número de pontos na mesma posição de p . O número de pontos com a mesma localização deve ser levado em conta porque todos eles terão caminhos passando por si, entretanto, do ponto de vista geométrico, dois pontos com a mesma localização são na verdade o mesmo ponto, assim como os caminhos que passam por eles.
- Linhas 13 e 14: Atualização da energia de acordo com a equação do modelo de reflexão de Phong.

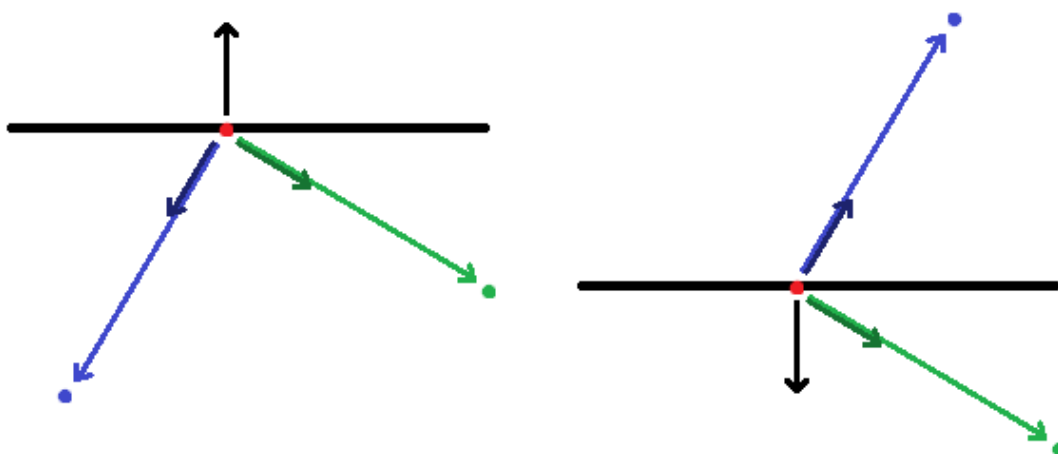


Figura 3.9: As duas situações validadas pelos testes das linhas 8 e 10, a primeira situação é evitada caso a face não seja definida como tendo dois lados, a segunda situação é sempre evitada.

3.3.3 Distribuição de Energia Entre os Caminhos

A distribuição de energia entre os caminhos depende dos níveis habilitados. Tendo apenas o nível 0, a energia será dividida apenas entre os receptores, caso o nível 1 também esteja habilitado, além de dividir a energia entre os caminhos que vão para cada canal, também é necessário dividir entre todos os pontos por onde passam os caminhos de nível 1. O mesmo raciocínio serve entre os níveis 1 e 2: com o nível 1 habilitado e o 2 desligado, toda a energia que chega a um ponto em uma face será dividida entre os 2

canais. Ligando o nível 2, também é necessário dividir essa energia entre os pontos nas outras faces.

algoritmo 4 Calculo do divisor da energia para cada nível

divisor (face)

retornar nívelLigado[0]*2 +

nívelLigado[1]*(numPontos-numPontosRepetidos) +

nívelLigado[2]*(face!= null)*

(numPontos - numRepetidos)*

((numPontos - numRepetidos) - (face.numPontos - face.numRepetidos))

3.3.4 Caminhos de Nível 2

Um caminho de nível 2 começa na fonte de som, passa por um ponto em uma face, após isso, passa por outro ponto em outra face, para finalmente chegar a um microfone virtual. São calculados para todos os pares de pontos, bem como para os dois microfones virtuais. O algoritmo a seguir mostra como são calculados os caminhos de nível 2.

algoritmo 5 Caminhos de nível 2

1: para cada par s1, s2 em faces

2: para cada par p1, p2 com p1 em s1.pontos e p2 em s2.pontos

3: parte1 = calcularCaminho(volume/divisor(s1), fonte, s1, p1, p2)

4: distCentral = módulo(p1.posição - p2.posição)

5: parte1.energia = parte1.energia*(1+distCentral²)

6: parte1.atraso = parte1.atraso - distCentral/velocidadeDoSom

7: para i = 0 enquanto i<2

8: parte2 = calculaCaminho(parte1.energia, p1, s2, p2, receptor[i])

9: saída(i, parte2.energia, parte1.atraso+parte2.atraso)

Calcular um caminho de nível 2 consiste em calcular um caminho saindo da fonte de som, passando pelo primeiro ponto e chegando em um segundo ponto, e em seguida, calcular um caminho saindo do primeiro ponto, passando pelo segundo e chegando nos receptores. Com isso, o atraso e a energia perdida entre o primeiro e o segundo ponto, são levados em conta duas vezes no cálculo. Então, entre as linhas 4 e 6, esses valores são devolvidos ao retorno da primeira chamada de calcularCaminho. A função “saída” será melhor detalhada a partir da seção 3.3.

3.3.5 Caminhos de Nível 1

Um caminho de nível 1 começa na fonte de som, passa por um ponto e termina em um dos dois receptores virtuais. Abaixo, o algoritmo utilizado para o cálculo.

algoritmo 6 Caminhos de nível 1

```

1: para cada s em faces
2:   para cada p em s.pontos
3:     para i = 0 enquanto i < 2
4:       caminho = calcularCaminho(volume/divisor(s), fonte, s, p, receptor[i])
5:       saída(i, caminho.energia, caminho.atraso)

```

3.3.6 Caminhos de nível 0

Para calcular os caminhos de nível 0 é utilizado o algoritmo 7.

algoritmo 7 Caminhos de nível 0

```

1: para i = 0 enquanto i < 2
2:   distancia = modulo(fonte - receptor[i])
3:   energia = 1/(1+ distancia2)
4:   atraso = volume/divisor(null)/velocidadeDoSom
5:   saída(i, caminho.energia, caminho.atraso)

```

3.4 Saída
3.4.1 Subrotina Saída

A função saída utilizada no fim do cálculo dos caminhos serve para gravar o resultado obtido em um vetor que em cada posição, armazena a energia e a informação de atraso será convertida no índice relativo à posição atual do ponteiro de som.

algoritmo 8 Subrotina saída

```

saída(canal, energia, atraso)
índice = inteiro(atraso*2) + canal;
para i = 0 enquanto i < intBuffer.tamanho
  se intBuffer[i].índice = índice
    intBuffer[i].energia = intBuffer[i].energia + energia
  retornar
intBuffer.adicionar(índice, energia)

```

Como o atraso informado é medido em tempo de amostra, é necessário apenas multiplicá-lo por 2 visto que cada amostra possui dois canais, a soma do número do canal serve para diferenciar a amostra entre direita e esquerda.

Em uma segunda etapa dessa rotina, é verificado no buffer intermediário (intBuffer) se já existe uma entrada com o índice recém calculado. Caso exista, a energia informada é apenas acumulada a essa entrada, e nenhuma entrada adicional é criada. Existem duas razões para isso, a primeira é o fato dos valores de energia estarem em ponto flutuante que serão multiplicados por inteiros de 16 bits, ou seja, são multiplicações que potencialmente gerarão muita perda, quanto mais se estiverem fragmentadas em vários valores. A segunda razão é que todas as entradas em intBuffer serão usadas na próxima etapa. Portanto, quanto menos entradas, mais eficiente será o processamento seguinte.

3.4.2 Escrita no Buffer de Saída

O buffer de saída é o que está constantemente sendo reproduzido, o algoritmo tem que ser rápido o bastante para escrever nesse buffer à frente do ponteiro de reprodução.

algoritmo 9 Escrevendo no buffer de saída

```
ultAmostra = amstAtual
amstAtual = getAmostraAtual()
para i = amstAtual+ATRASO enquanto i < amstAtual*2-ultAmostra+ATRASO
  para j = 0 enquanto j < intBuffer.tamanho
    outBuffer[i*2 + intBuffer[j].indice] += inBuffer[i]* intBuffer[j].energia
```

O método definido no algoritmo 9 transforma o som da fonte no som ouvido pelo usuário. Para isso é obtida a posição da amostra sendo reproduzida no exato momento da execução do método, através da rotina getAmostraAtual. Esse valor é acrescido da constante ATRASO que será detalhada a seguir, desse modo é obtida a posição da amostra no buffer de entrada. Os valores presentes no buffer intermediário então, são aplicados à posição obtida de acordo com o procedimento a seguir, a Figura 3.10 ilustra esse procedimento.

A parte “índice” das entradas do buffer intermediário é somada à posição da amostra no buffer de entrada multiplicada por dois, essa multiplicação é feita para converter uma posição de buffer com um canal de som, para uma posição de buffer de som com dois canais. A soma é feita para gerar o atraso referente a propagação do som, no índice somado já está presente a informação de qual canal está sendo trabalhado (direita ou esquerda). O valor encontrado é a posição da amostra no buffer de saída

Com as posições de amostras nos buffers de entrada e saída, são obtidos os valores referentes a cada posição nos seus respectivos buffers. Com isso o valor de entrada é multiplicado pela parte “energia” das entradas do buffer intermediário e somada ao valor encontrado no buffer de saída. Esse resultado é devolvido ao buffer de saída na mesma posição.

O procedimento descrito acima é aplicado para todas as entradas no buffer intermediário e para um bloco de amostras no buffer de entrada imediatamente posterior a amostra atual. O tamanho desse bloco é uma estimativa do número de amostras que se transcorrem entre cada vez que o algoritmo 9 é executado. A estimativa é a diferença entre a posição da amostra atual e a posição da amostra obtida na execução anterior do método.

A constante ATRASO foi adicionada pelo fato de uma amostra ser reproduzida muito rápido. No momento em que se obtém a amostra presente, já existe outra amostra sendo reproduzida. Esse atraso não deve ser maior que 200 milésimos, pois a partir desse ponto já seria possível percebê-lo.

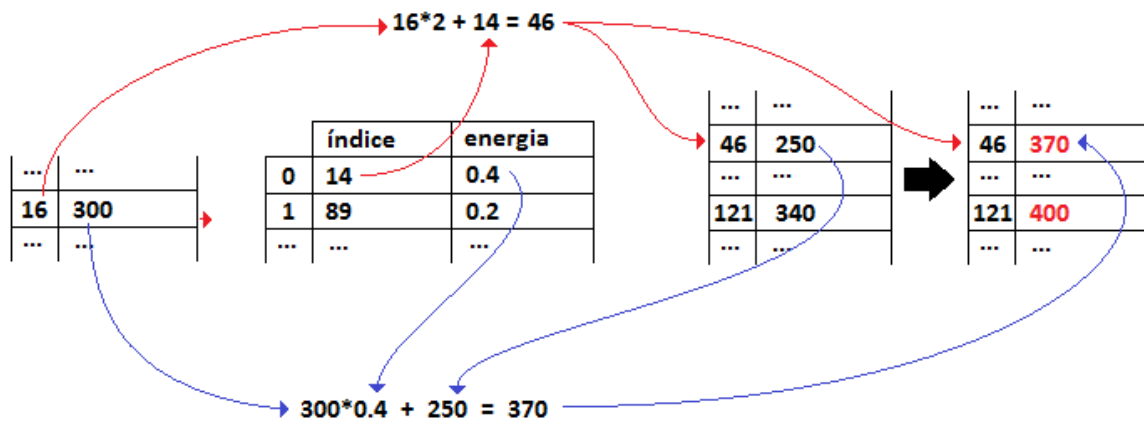


Figura 3.10: Processo de escrita no buffer de saída. Da direita para a esquerda, as tabelas significam: buffer de entrada, buffer intermediário, buffer de saída antes das escritas e buffer de saída após as escritas.

4 AVALIAÇÃO

A avaliação foi dividida em duas etapas, avaliação de performance e avaliação com usuário.

Para todos os testes foi utilizado um computador com processador Intel Core i7 3630QM 2.40GHz, 12 GB de memória RAM, sistema operacional Windows 7 64 bits e placa de som onboard Intel Display Audio. Os fones de ouvido utilizados foram fones comuns encontrados no mercado em uma faixa de preço de R\$ 70,00.

4.1 Avaliação de Performance

Como pode ter sido observado, no modelo descrito anteriormente existem dois pontos de processamento mais críticos, a definição de novas faces e o cálculo dos caminhos até a saída. A complexidade da definição de cada face é $O(n)$ onde n é o número de faces, já a complexidade do cálculo de caminhos é $O(n^i)$, onde i é o número de níveis.

Para avaliar os tempos necessários na execução do algoritmo, foi feito um teste renderizando quantidades diferentes de faces, nas Tabelas 4.1 e 4.2 estão apresentados os resultados. Todos os tempos são medidos em milissegundos.

Tabela 4.1: Medição de performance para os níveis 0 e 1.

Número de Faces	1000	2000	3000	4000
Caminhos de Nível 0	1ms	1 ms	1 ms	1 ms
Caminhos de Nível 0 e 1	2 ms	7 ms	21 ms	21 ms

Tabela 4.2: Medição da performance para o nível 2.

Número de Faces	100	200	300	400
Caminhos de Nível 2	90 ms	911 ms	5970 ms	31100 ms

Com esse dados, é possível perceber que apenas a complexidade dos algoritmos não explica a diferença de performance dos níveis 0 e 1 para o nível 2. O problema no nível 2 é o fato do número de entradas no buffer intermediário(explicado em 3.4) ser potencialmente muito grande. Portanto, o primeiro intervalo estimado será muito pequeno comparado ao tempo necessário para preencher o buffer de saída. Com isso, o próximo intervalo estimado será consideravelmente maior que o anterior, o que resultará

em um tempo ainda maior para preencher o buffer de saída. Logo, nos caminhos de nível 2, o intervalo estimado demora muito para convergir. Então, esse nível não foi utilizado em testes posteriores.

4.2 Avaliação com Usuários

Para avaliação com o usuário foram propostos quatro testes, em cada um o usuário deveria realizar tarefas simples utilizando-se da informação auditiva e, por vezes, visual fornecida pela aplicação.

Todos os testes foram divididos em seções, onde a tarefa a ser realizada é a mesma, porém com algumas condições diferentes.

Em todos os testes foi utilizada uma música quando necessário um som ininterrupto. Essa opção foi feita visto que a música escolhida possui muitas variações de frequência, voz e seria mais confortável ao usuário realizar o teste com uma música do que com uma sequência de sons aleatórios.

Para os testes 2 e 3, ainda foram utilizadas três batidas de instrumentos de bateria, essa escolha foi feita com base no fato da forma de onda de uma batida de bateria ser similar a de um estalo emitido por usuários da técnica de ecolocalização, possuindo grande variação de frequências

4.2.1 Primeiro Teste

O teste é dividido em duas seções, cada uma composta por três cenas. Conforme a Figura 4.1, uma cena é composta por:

- Quatro paredes de 4,5 metros cada, posicionadas de modo a formar um quadrado. Cada parede é formada por seis faces, em todas elas o coeficiente de especularidade é igual a 0,9 o de absorção é igual a 0,4 e a constante alfa é igual a 1. Todas foram divididas em 21 pontos.
- Uma fonte de som, que estará reproduzindo uma música.
- Um par de receptores virtuais, posicionados a 20 centímetros de distância entre si e controláveis pelo usuário.
- Uma representação visual da fonte de som, em forma de uma esfera branca.

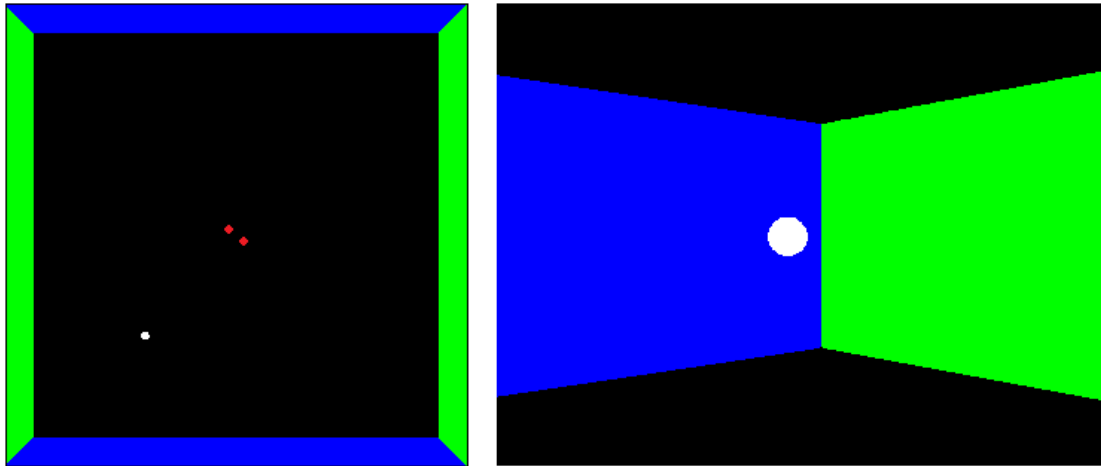


Figura 4.1: Representação visual de uma cena no teste. Na imagem da esquerda, as esferas vermelhas representam os receptores, enquanto a branca, é a representação visual da fonte. A imagem da direita é a visualização disponível ao usuário, os receptores virtuais ficam nos limites laterais externos dessa imagem.

Antes de começar o teste, todos os usuários tiveram a chance de realizar uma versão muda do teste para se adaptar aos controles.

Em cada cena o usuário deve utilizar as setas do teclado para se deslocar em primeira pessoa durante 15 segundos, tendo mais 25 segundos para continuar escutando a música da posição em que parou.

Após percorrer três cenas, o usuário deve informar em qual das três a representação visual da fonte de som e a própria fonte de som estão na mesma posição.

As variáveis independentes do teste são as seguintes:

- Distância entre a fonte de som e sua representação visual: Em apenas uma das cenas essa distância será zero. Em outra a distância será de aproximadamente 50 centímetros. Na cena restante será maior que dois metros. A Tabela 4.3 mostra a ordem em que essas distâncias ocorrem.
- Número dos caminhos de som: Na primeira seção, o som ouvido pelo usuário tem apenas os caminhos de nível 0. Na segunda são os caminhos de nível 0 e 1.

Tabela 4.3: Ordem das cenas em cada seção, considerando a distância entre a fonte de som e sua representação visual.

Distância(m)	Seção 1	Seção 2
0	3	1
Aprox. 0.5	2	3
Maior que 2	1	2

Como variável dependente nesse teste, existe apenas a resposta do usuário para cada seção. A resposta pode estar:

- Correta: o usuário indicou corretamente a cena em que a distância entre fonte e representação visual é igual à zero.
- Parcialmente incorreta: o usuário indicou como resposta a cena em que a representação visual e a fonte de som estão a aproximadamente 50 centímetros
- Totalmente incorreta: a resposta do usuário foi a cena em que a distância entre a fonte e a representação visual é maior que dois metros

4.2.2 Segundo Teste

No segundo teste a fonte de som e os receptores ficam em uma posição fixa de modo a aproximar a posição da boca e dos ouvidos de uma pessoa, conforme o desenho da Figura 4.3. À direita ou a esquerda desse conjunto fica posicionada uma parede, o objetivo do usuário é, através do som refletido pela parede, descobrir em que lado esta se encontra.

O teste é dividido em quatro seções, cada seção é composta por uma sequência de dez paredes posicionadas aleatoriamente, cada parede é formada por dois polígonos, com as mesmas propriedades do teste anterior. Não serão posicionadas mais de três paredes seguidamente no mesmo lado. Na Figura 4.2, é mostrada uma representação visual desse teste, a parede à direita.

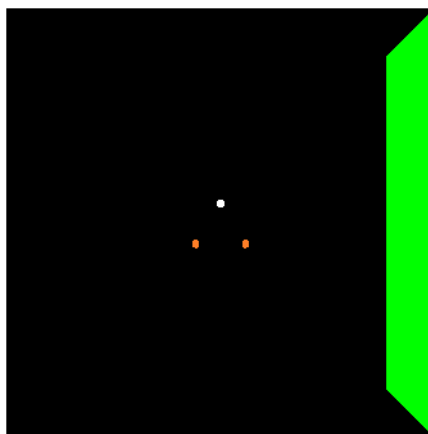


Figura 4.2: Representação visual do segundo teste com uma parede à direita. As esferas douradas representam os receptores, enquanto a branca representa a fonte de som.

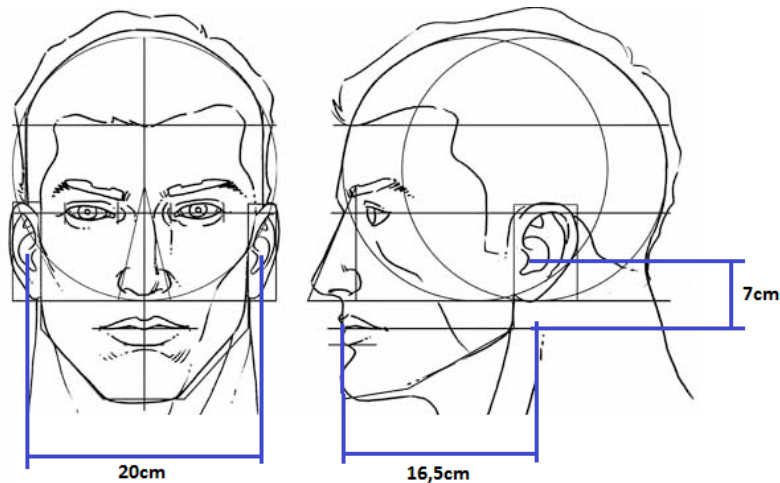


Figura 4.3: Medidas utilizadas no posicionamento dos receptores e da fonte de som nesse teste.

Esse teste conta com as seguintes variáveis independentes:

- Níveis dos caminhos do som: Na primeira e terceira seção estão presentes tanto os caminhos de nível 0, quanto os de nível 1. Na segunda e quarta, apenas o nível 1.
- Forma da reprodução do som: Na primeira e segunda seção é reproduzida uma música durante toda a seção. Na terceira e quarta, o usuário pode reproduzir pequenos sons através das teclas 1, 2, 3 e 4. Cada um deles com no máximo 250 milissegundos e em intervalos de pelo menos 750 milissegundos. Sendo o primeiro um som agudo e os três seguintes, batidas de instrumentos de bateria.
- Posição das paredes: Uma parede pode ficar a direita ou à esquerda da fonte e dos receptores e sempre à mesma distância (75 cm).

As variáveis dependentes desse teste são:

- O número de acertos que o usuário obteve em cada seção.
- Quantidade de cada som utilizado nas seções 3 e 4.
- O tempo gasto para resolver cada seção.

4.2.3 Terceiro Teste

Assim como o segundo teste, o terceiro também é dividido em quatro seções. Em cada uma delas, o usuário, encontrando-se em uma cena similar a do primeiro teste, controla de maneira livre um conjunto formado por um par de receptores e uma fonte de som, posicionadas nas mesmas proporções e medidas utilizadas no segundo teste, como pode ser visto na Figura 4.4.

Entretanto, cada parede nesse teste tem três portas, resultando 12. Aos olhos do usuário todas são iguais, mas aos ouvidos, uma se encontra aberta (não reflete som). O objetivo do usuário é encontrar a porta aberta.

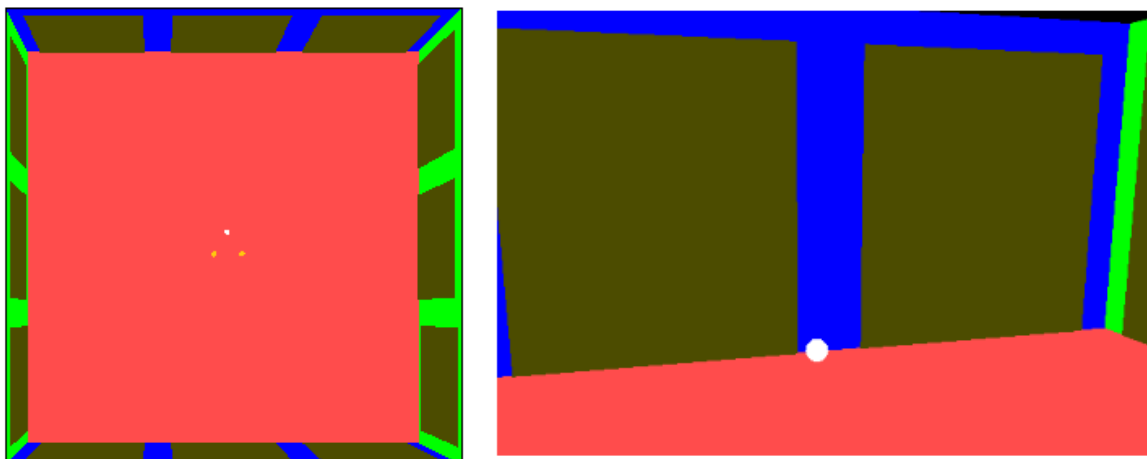


Figura 4.4: Na imagem à esquerda, as esferas douradas representam os receptores, e a esfera branca representa a fonte de som. Na visualização disponível ao usuário (direita), apenas a representação da fonte é preservada, indicando o limite frontal do conjunto controlável.

As paredes têm as mesmas medidas e propriedades do primeiro teste, com exceção daquela em que se encontra a porta aberta, que nesse caso, é formada por apenas quatro polígonos. Nesse teste foi adicionado o chão formado por dois polígonos com coeficiente de absorção igual a 0,7. As demais propriedades são as mesmas das paredes.

Assim como no primeiro teste, foi possível ao usuário utilizar uma versão muda do aplicativo para ambientação antes do teste começar. Em cada cena o usuário tem dois minutos para escolher uma porta.

O usuário pode se locomover livremente, inclusive colidindo nos limites da sala. Quando julgar ter a resposta, ele deve ir de encontro a porta desejada e pressionar a tecla "enter".

Para esse teste, foram definidas as seguintes variáveis independentes:

- Níveis dos caminhos do som: Na primeira e terceira seção estão presentes tanto os caminhos de nível 0, quanto os de nível 1. Na segunda e quarta, apenas o nível 1.
- Forma da reprodução do som: Na primeira e segunda seção é reproduzida uma música durante toda a seção. Na terceira e quarta, o usuário pode reproduzir livremente pequenos sons (no máximo 250ms) através das teclas 1, 2, 3 e 4. Sendo o primeiro um som agudo e os três seguintes, batidas de instrumentos de bateria.
- Posição da porta aberta: Definida aleatoriamente para cada seção, pode ser qualquer uma das 12.

As variáveis dependentes são as seguintes:

- Trajetória do usuário pela sala em cada seção.

- Tempo gasto pelo usuário para finalizar cada seção.
- Quantidade de cada som utilizado nas seções 3 e 4.
- Porta escolhida pelo o usuário, caso ele tenha feito uma escolha antes dos dois minutos.

4.2.4 Quarto Teste

Ele seria composto por um corredor de comprimento fixo com curvas de noventa graus. Assim como os testes anteriores, ele seria dividido em quatro seções.

O objetivo do teste seria controlar um conjunto formado por um par de receptores e uma fonte de modo a chegar ao fim do corredor, sendo que os controles não seriam livres, o usuário só poderia girar de 90 em 90 graus em torno do eixo do conjunto, assim como os avanços e retrocessos seriam executados em uma distância definida. Alguns corredores podem ser vistos na Figura 4.5. O usuário percorreria o corredor sem nenhum auxílio visual.

Por motivos de performance, esse teste não foi aplicado. Ocorreram muitas variações de velocidade durante a execução do teste, o que não só prejudicaria muito a noção de posicionamento do usuário, como também prejudicou a execução dos sons, deixando-os muito distorcidos e com um ruído muito elevado.

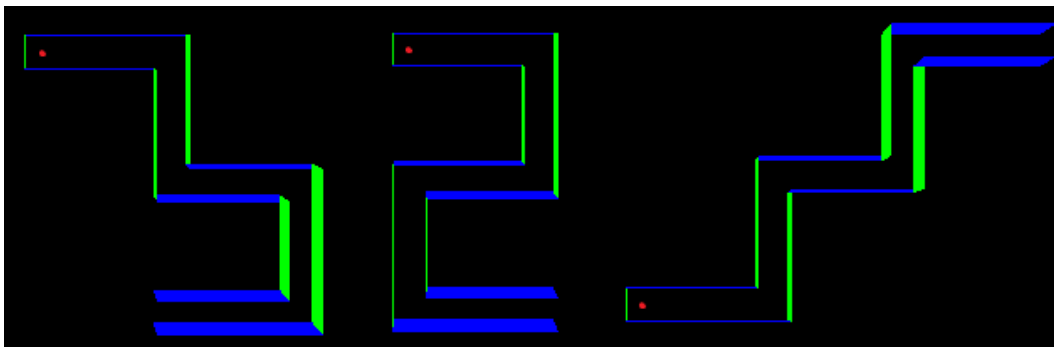


Figura 4.5: Corredores gerados aleatoriamente para o quarto teste. A esfera vermelha indica a posição inicial do conjunto controlável pelo usuário.

4.2.5 Questionários

Os questionários são realizados após determinadas etapas dos testes. Com exceção do questionário Q5, todos eles pedem ao usuário para avaliar, em uma escala de 1 a 10, a dificuldade que o mesmo encontrou em realizar tarefas específicas. Foi esclarecido a todos os usuários que essa avaliação deveria ser subjetiva:

- *Q1*: Realizado depois da primeira seção do segundo teste. É pedido ao usuário para avaliar a dificuldade da primeira e a última parede da primeira seção.
- *Q2*: Realizado após o segundo teste, é pedida a avaliação da dificuldade de cada seção desse teste.
- *Q3*: Igual ao questionário anterior, realizado após o terceiro teste.
- *Q4*: É aplicado logo depois do questionário anterior. Pede a avaliação da dificuldade do segundo e terceiro teste.
- *Q5*: Utilizado no primeiro teste, pergunta ao usuário em uma escala de 1 a 10 a experiência que o mesmo julga ter com vídeo games.

4.2.6 Hipóteses

Todos os testes e questionários foram desenvolvidos para avaliar as seguintes hipóteses:

- *H1*: É possível identificar a presença de objetos próximos através do som refletido neles.
- *H2*: A ausência do som de nível 0 facilita a validação da hipótese *H1*.
- *H3*: Os ouvidos são sensíveis a mudanças bruscas na posição de objetos próximos, o que pode facilitar na sua localização.
- *H4*: É possível se localizar e se locomover em um ambiente utilizando apenas o som refletido nele.
- *H5*: A ausência de som de nível 0 facilita a validação de *H4*.

Essas hipóteses serão discutidas no próximo capítulo, juntamente com os resultados obtidos nos testes e questionários.

5 RESULTADOS E DISCUSSÕES

5.1 Avaliando as Capacidades dos Usuários

O objetivo original do primeiro teste era avaliar a capacidade auditiva de cada usuário, entretanto, conforme o teste foi sendo conduzido, ele mostrou que poderia ser utilizado para avaliar também a capacidade que o usuário tem de controle livre de um objeto virtual. Em geral, usuários com pouca ou nenhuma experiência com jogos eletrônicos tiveram mais dificuldades, muitos não utilizavam mais de uma seta ao mesmo tempo ou pressionavam os botões com pequenos cliques, ao invés de pressionar cada botão continuamente. A Tabela 5.1 mostra as respostas dadas no primeiro teste

Tabela 5.1: Respostas dadas pelos usuários ao primeiro teste

	Experiência com Jogos	Resposta seção 1	Resposta seção 2
Usuário 1	9	Correta	Parcialmente Incorreta
Usuário 2	10	Correta	Parcialmente Incorreta
Usuário 3	3	Correta	Parcialmente Incorreta
Usuário 4	9	Correta	Totalmente Incorreta
Usuário 5	8	Correta	Correta
Usuário 6	1	Parcialmente Incorreta	Totalmente Incorreta
Usuário 7	1	Totalmente Incorreta	Correta
Usuário 8	9	Correta	Correta

Outro detalhe que foi possível observar, foi que a maioria dos usuários se mostrou em dúvida nas duas seções entre a resposta correta e a parcialmente incorreta, porém, quase todos acertaram a resposta na primeira seção, enquanto menos da metade acertou na segunda.

5.2 Discussão Sobre a Percepção de Objetos Próximos

O segundo teste é o primeiro teste que avalia a capacidade dos usuários em utilizar ecolocalização, é um teste extremamente simples, entretanto, como nenhum usuário tem deficiências visuais ou tinha usado ecolocalização, ele é adequado para ser aplicado. Como é possível ver na Tabela 5.2, os usuários em geral tiveram poucas dificuldades para resolver esse teste.

Tabela 5.2: Resultados do primeiro teste. Os usuários estão na mesma ordem da Tabela 5.1. Os tempos estão medidos em segundos.

	Seção 1		Seção 2		Seção 3		Seção 4	
	Acertos	Tempo Gasto	Acertos	Tempo Gasto	Acertos	Tempo Gasto	Acertos	Tempo Gasto
Usuário 1	8	37	10	14	10	71	10	33
Usuário 2	10	34	10	13	7	150	10	22
Usuário 3	10	37	10	20	10	66	10	44
Usuário 4	10	32	10	10	10	49	10	22
Usuário 5	10	52	10	11	8	53	10	29
Usuário 6	7	40	10	19	10	83	10	52
Usuário 7	10	38	10	21	8	131	10	25
Usuário 8	9	40	10	7	10	47	10	31
Média	9,25	38,75	10	14,375	9,125	81,25	10	32,25

Pela Tabela 5.2 é possível perceber que todos os usuários tiveram mais facilidade resolvendo as seções sem som de nível 0. Em todas as seções o número de acertos foi alto, mas a diferença de tempo entre as seções com e sem nível 0 é muito significativa. Em alguns casos, foi possível perceber que alguns usuários estavam resolvendo a seção 2 com tanta facilidade, que pareciam mais preocupados com o tempo do que com a resposta em si, pois estavam sempre tentando dar uma resposta assim que permitido pela interface.

Na seção 4 os usuários levaram mais tempo em relação a seção 2 porque ela exigia emitir sons antes de dar uma resposta, alguns usuários fizeram questão de ouvir todos os sons pelo menos uma vez à cada nova parede, outros deram preferência para um ou dois sons, mas acharam melhor ouvi-los mais de uma vez antes de responder. Isso é reforçado pela Tabela 5.3, onde é possível visualizar que quatro dos oito usuários deixaram de utilizar pelo menos dois sons disponíveis em algumas paredes. Essa tabela também reforça a diferença de dificuldade entre seções com e sem caminhos de nível 0, já que na seção 4, os usuários precisaram reproduzir menos sons para encontrar uma resposta.

Tabela 5.4: Número de vezes em que cada som foi reproduzido por cada usuário nas seções 3 e 4.

	Seção 3				Seção 4			
	som 1	som 2	som 3	som 4	som1	som 2	som 3	som 4
Usuário 1	38	30	22	23	18	15	16	20
Usuário 2	59	32	62	41	12	12	6	2
Usuário 3	13	15	16	16	11	10	10	14
Usuário 4	18	12	25	5	0	0	21	0
Usuário 5	23	18	14	15	12	11	10	10
Usuário 6	13	15	35	29	14	14	15	10
Usuário 7	51	50	42	41	16	10	2	0
Usuário 8	6	42	15	4	3	34	0	0
Média	27,625	26,75	28,875	21,75	10,75	13,25	10	7

A diferença de dificuldade entre seções com e sem nível 0 também foi percebida pelos usuários. Apenas o Usuário 5 julgou ter tido mais dificuldade na segunda seção em relação a primeira. Entre as seções 3 e 4, todos julgaram a quarta como a mais fácil. A Figura 5.1 mostra o resumo das respostas do questionário Q2.

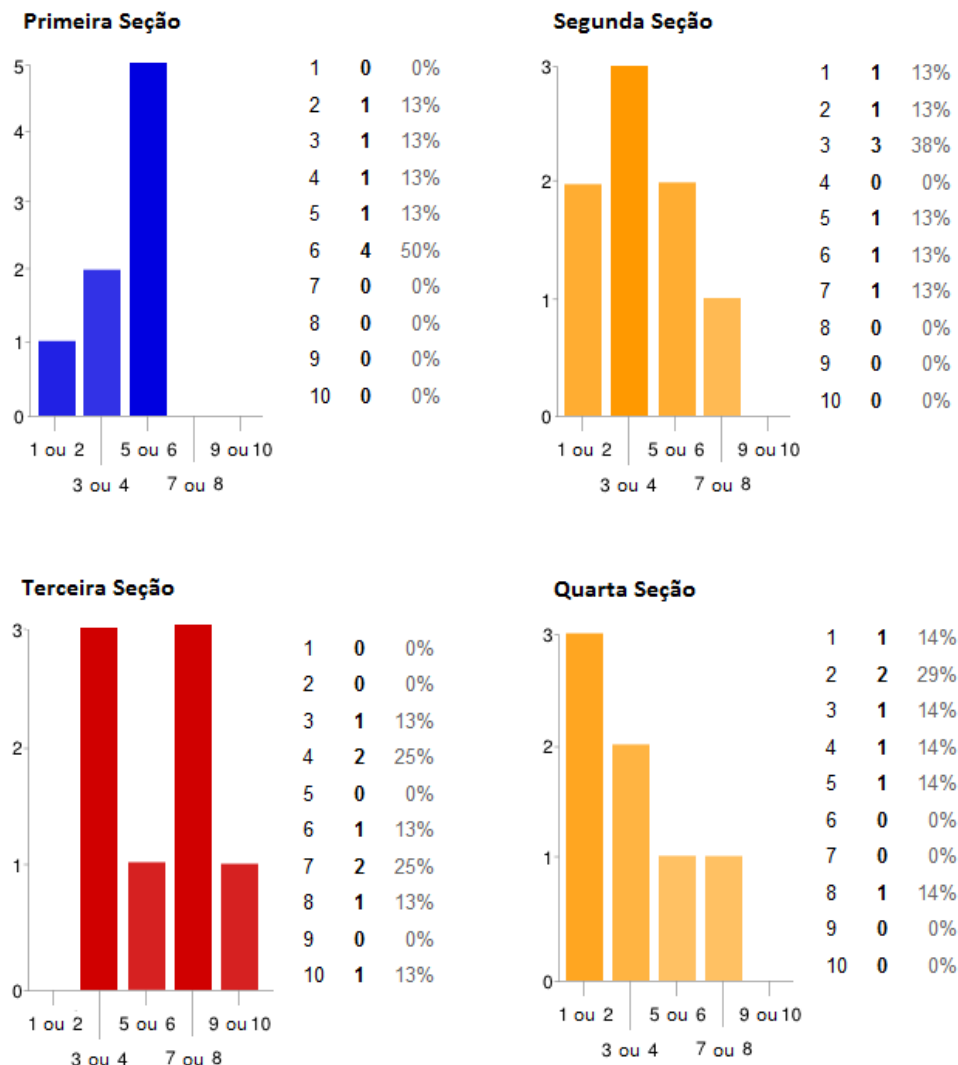


Figura 5.1: Respostas dos usuários ao questionário Q2. O eixo horizontal indica o nível de dificuldade que os usuários tiveram. O eixo vertical indica quantas vezes cada dificuldade foi escolhida. As dificuldades estão agrupadas aos pares para melhor visualização do resultado nos gráficos. A lista a direita de cada gráfico também indica quantas vezes cada resposta foi escolhida, mas não agrupando as respostas.

Os resultados das seções 1 e 3 também foram positivos, apesar de afirmarem que tiveram maior dificuldade, todos os usuários acertaram 70% ou mais nessas seções. Algo que foi observado durante a aplicação da seção 1, foi que alguns usuários pareceram ter mais dificuldade ao resolver as primeiras superfícies, inclusive fazendo queixas em relação à dificuldade do teste, mas após algumas respostas, todos começaram a responder mais rápido. Como é possível ver na Figura 5.2, As respostas do questionário Q1 indicam isso que os usuários acharam que o teste estava significativamente mais fácil no final em relação ao início.

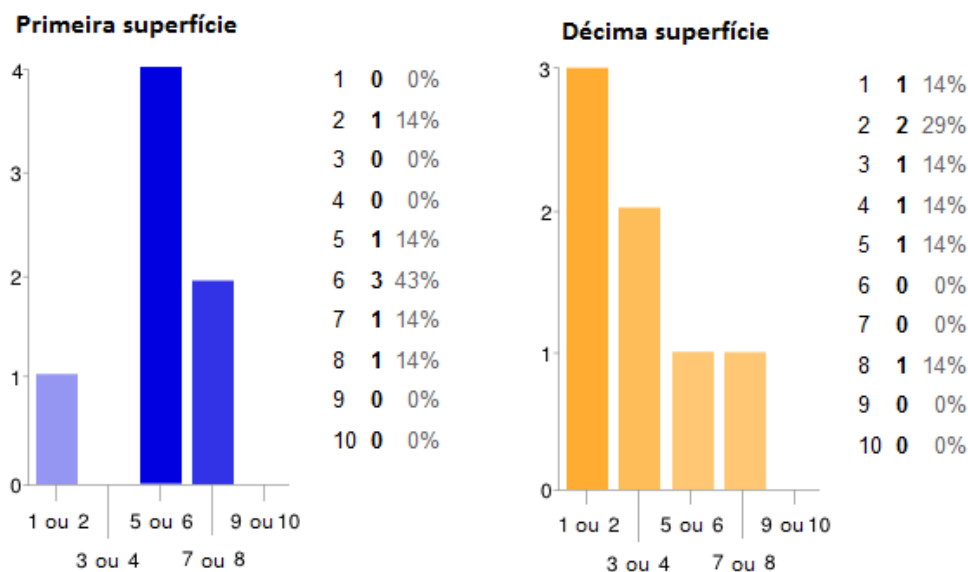


Figura 5.2: Respostas do questionário Q1. O eixo horizontal indica as dificuldades escolhidas pelos usuários agrupadas aos pares. O eixo vertical indica o número de vezes que cada resposta foi escolhida. A lista a direita de cada gráfico também indica quantas vezes cada nível de dificuldade foi escolhido, mas não agrupando as respostas.

O mesmo não pode ser dito em relação a seção 3, apesar do número de erros ser similar, o momento em que eles ocorrem mostra que os usuários não tiveram mais facilidade ao final da seção. Como é possível ver na Figura 5.3, ao contrario da primeira seção onde o número de erros diminui conforme a seção chega ao fim, na terceira seção os erros foram distribuídos de maneira mais constante.

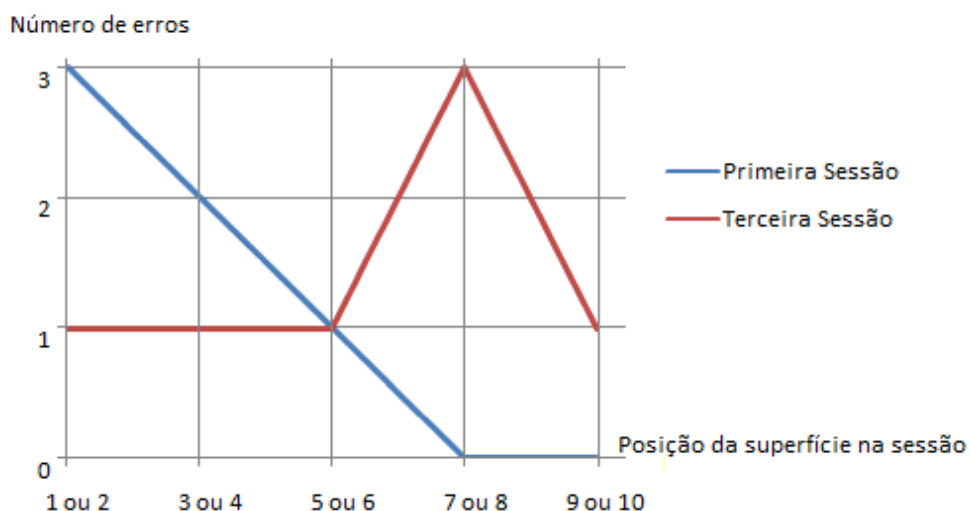


Figura 5.3: Distribuição dos erros nas seções 1 e 3.

5.3 Discussão Sobre a Orientação e Locomoção

Esse teste tinha o objetivo de analisar a capacidade de localização e orientação dos usuários. Entre todos os usuários, apenas o último escolheu a porta certa em duas ocasiões, justamente as duas sem som de nível 0. Todos os demais erraram todas as respostas.

Uma característica muito comum à maioria dos usuários foi a tentativa de encontrar a saída da sala se aproximando das portas pela frente, quando o teste anterior indicava que uma boa abordagem seria ficar de lado para as portas.

Tendo em vista os resultados dos testes até o sétimo usuário, ao oitavo testador foram dadas as seguintes dicas:

- Procurar se movimentar sempre de lado para as portas, visto que o método de rendering utilizado nesse trabalho é melhor adaptado para a localização de sons laterais, já que existem duas informações que ajudam nessa localização: a diferença de atraso e de energia em cada ouvido. Enquanto que para sons frontais e traseiros existe apenas uma: a diferença de energia, e ainda assim, é necessário se locomover para perceber essa diferença.
- Não se aproximar tanto das portas. Essa dica foi dada tentando manter o usuário a uma distância das portas similar à distância em que o usuário ficou das paredes no segundo teste.

Esse usuário não só respondeu corretamente duas seções, mas também afirmou ter certeza absoluta de sua resposta na quarta seção.

Isso pode indicar que a tarefa exige um treinamento muito maior em relação ao segundo teste. Entretanto nada pode ser concluído, visto que apenas um usuário conseguiu respostas corretas.

5.4 Resposta Sonora

O segundo teste também serviu para avaliar o comportamento do modelo descrito no capítulo 3, para isso foram gerados dois gráficos, um para cada canal, contendo os tempos de atraso e energias com que cada “partícula de som” chega aos receptores virtuais. Os gráficos estão na Figura 5.4

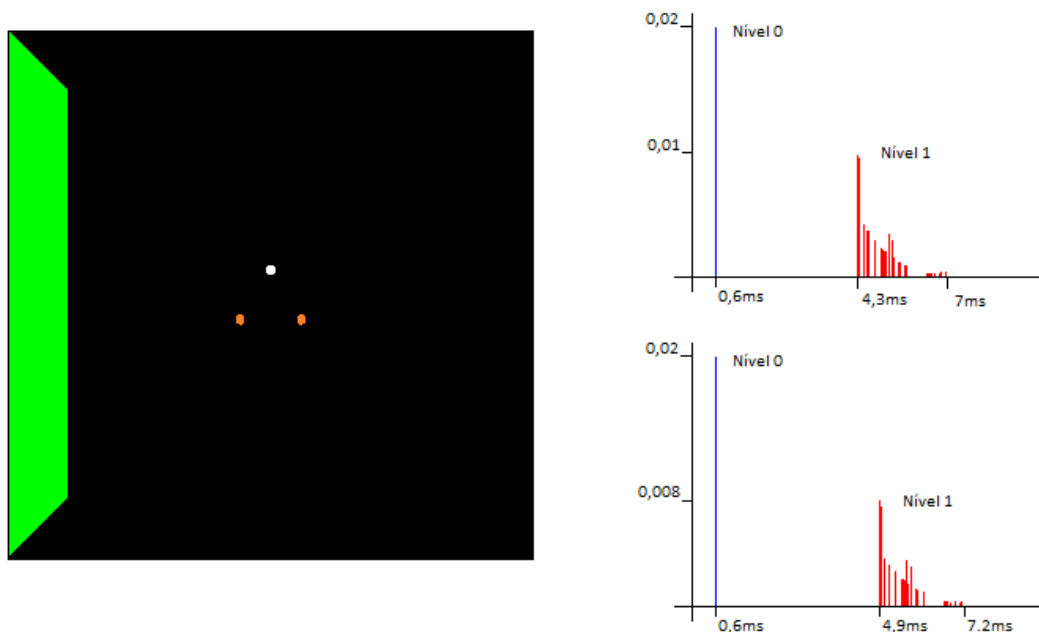


Figura 5.4: Gráficos com tempos de resposta para a situação ilustrada à esquerda. O gráfico superior representa o canal da esquerda, enquanto o inferior representa o da direita. O eixo horizontal representa o tempo em que uma “partícula de som” chega ao receptor virtual, enquanto o eixo vertical representa a energia com que essa partícula chegou.

Como é mostrado na Figura 5.4, em ambos os canais o som de nível 0 chega ao mesmo tempo e com a mesma força em cada receptor, já o som de nível 1 tem mais energia e chega antes ao receptor da esquerda em relação ao da direita.

5.5 Revisando as Hipóteses

As três primeiras hipóteses foram avaliadas através dos resultados do segundo teste e todas foram confirmadas. A Hipótese H1 diz que "é possível identificar a presença de objetos próximos através do som refletido neles". Como é possível observar na tabela 5.3, a média de acertos no teste foi superior a 9, e nenhum usuário precisou de mais de 3 minutos para responder cada seção.

A Hipótese H2 diz que "a ausência do som de nível 0 facilita a validação da hipótese H1". Mesmo com resultados muito bons nas seções 1 e 3 (com som de nível 0), nas seções 2 e 4 (sem som direto) os resultados foram ainda melhores. Com 100% de acerto, um tempo de resposta inferior à metade das seções com som de nível 0 e com a grande maioria dos usuários considerando essas seções as mais fáceis.

A Hipótese H3 diz que "os ouvidos são sensíveis a mudanças bruscas na posição de objetos próximos, o que pode facilitar na sua localização". Perceber mudanças bruscas na posição dos objetos só seria possível nas duas primeiras seções do segundo teste, como nas seções sem nível 0 a pontuação foi a mesma, é preciso comparar as seções 1 e 3. Pelo gráfico na Figura 5.4 é possível perceber que enquanto na primeira seção os usuários foram errando menos conforme a seção chegava ao fim, na terceira seção os erros não diminuíram conforme a seção transcorria, pelo contrário, ouve um número de

erros maior no fim da seção. Além disso, todos os usuários consideraram que a seção 1 estava muito mais fácil no final em relação ao início.

O terceiro teste serviu para verificar as hipóteses H4 e H5, e ambas não puderam ser confirmadas.

Para H4, "É possível se localizar e se locomover em um ambiente utilizando apenas o som refletido nele", de um total 32 tentativas (8 usuários, 4 seções), apenas 2 tiveram um resultado positivo. Já para a Hipótese H5, "a ausência do som de nível 0 facilita a validação da hipótese H4", o oitavo usuário respondeu corretamente as duas seções sem som de nível 0, porém é muito pouco para confirmar a hipótese, principalmente porque as condições do terceiro teste para esse usuário foram diferentes das condições dos demais.

6 CONCLUSÃO

Nesse trabalho foi proposta uma técnica de rendering de som com o objetivo específico de ajudar a treinar, desenvolver e analisar a capacidade de seres humanos utilizarem a técnica de ecolocalização.

Analisados os resultados dos exercícios propostos a oito usuários não cegos e que nunca utilizaram a ecolocalização, mostrou-se que essa técnica pode ser utilizada em ambiente virtual com bastante facilidade, obtendo um alto nível de sucesso em um dos testes e confirmando três de cinco hipóteses.

Considerando o teste em que as hipóteses não foram confirmadas, ainda sim é possível analisar essa situação de maneira positiva, visto que um usuário, com as dicas certas, foi capaz de resolver o teste na condição anteriormente provada ser a mais fácil. Obviamente, esse teste ainda carece de mais resultados, mas existe um bom indício de que, com o treinamento adequado ele também possa ser resolvido e suas hipóteses serem provadas.

Sobre a técnica de ecolocalização para seres humanos e o envolvimento da ciência da computação na área, certamente ainda existe um longo caminho a ser percorrido. Uma grande parte das pessoas com deficiências visuais sequer sabe que seres humanos podem se localizar através do eco de sons emitidos. E sobre a participação da ciência da computação na área, basta perceber que apenas um trabalho referenciado aqui trata desse tema.

Em trabalhos futuros a técnica de rendering de som apresentada aqui deve ser melhorada em alguns aspectos. Atualmente ela não é capaz de simular todos os tipos de ambientes. Falta o que seria um equivalente ao zbuffer de técnicas visuais. Ainda é preciso melhorar a performance do nível 2. Existe uma boa chance de conseguir um grande salto de performance através de um processamento mais paralelizado, utilizando programação de placas gráficas. Também é necessário identificar a causa e corrigir o ruído presente em todos os testes realizados aqui.

Além do modelo de rendering utilizado, próximas pesquisas sobre o tema podem envolver encontrar tipos de som que mais ajudam um ser humano a se ecolocalizar, algoritmos que amplifiquem a informação de localização presente no som, e um estudo comparando a capacidade do ser humano de produzir sons, com a capacidade de ouvi-los e compreendê-los.

REFERÊNCIAS

FOLEY, J, D.; VAN DAM, A.; FEINER, S, K.; HUGHES, J, F. **Computer Graphics: Principles and Practice**. Addison-Wesley, Reading, 1990.

GIBBS, B, M.; VIVEIROS, E,B. **An Image Source Method of Predicting Acoustic Louvre Performance in Enclosures**. Liverpool: [s.n.], 2000.

RINDEL, J, H. The Use of Computer Modeling in Room Acoustics. **Journal Of Vibroengineering**, [s.l.], Index 41-72, 2000.

SEKI, Y.; SATO, T. A Training System of Orientation and Mobility for Blind People Using Acoustic Virtual Reality. **IEEE Transactions on Neural Systems and Rehabilitation Engineering**, Tsukuba, v. 19, n. 1, p. 95-104, Fev. 2011

SHANNON, C, E. Communication in the Presence of Noise. **Proc. Institute of Radio Engineers**, v. 37, n. 1, p. 10-21, Jan. 1949

THALER, L.; ARNOTT, S, R.; GOODALE, M, A. Neural Correlates of Natural Human Echolocation in Early and Late Blind Echolocation Experts. **PLoS ONE**, 2011, [s.l.], Disponível em : <
<http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0020162>>.
 Acesso em: jul. 2013.

TSINGOS N.; GASCUEL, J, D. Acoustic simulation using hierarchical time-varying radiant exchanges. **CiteSeerX**. Pennsylvania, 1998, Disponível em: <
citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.59.439&rep=rep1&type=pdf>.

WILSON S. Wave PCM Soundfile Format. **Center for Computer Research in Music and Acoustics**, [S.l.], Janeiro 2003. Disponível em:
 <<https://ccrma.stanford.edu/courses/422/projects/WaveFormat/>>.