3
Reproduzindo Áudio com Informação Espacial

Neste capítulo fazemos uma breve revisão sobre como seres humanos são capazes de localizar fontes sonoras e também sobre os sistemas de reprodução de áudio que buscam simular a existência de uma fonte sonora no espaço.

3.1
Indicadores de Posição

A Figura 3.1 ilustra uma fonte sonora e os menores caminhos de propagação entre a fonte e os ouvidos de uma pessoa. A distância existente entre os ouvidos ao longo do eixo X faz com que exista uma diferença de comprimento entre esses caminhos (com exceção para fontes localizados no plano YZ). Essa diferença acarreta atenuações e tempos de chegada diferentes para a frente de onda em cada um dos ouvidos. A diferença de amplitude (*interaural level difference* - ILD) e a diferença de tempo (*interaural time difference* - ITD) são dois de três principais fatores fatores importantes para a localização do som no espaço.

Essas duas dicas de localização, no entanto, não são suficientes para permitir uma localização precisa da fonte sonora. Ao longo de qualquer

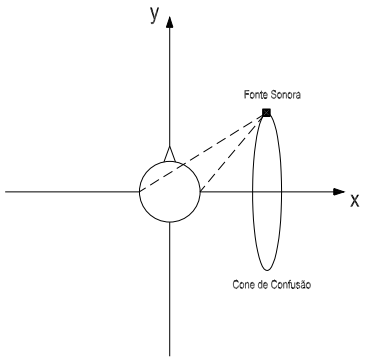

Figura 3.1: Caminhos de Propagação do Som e Cones de Confusão

círculo paralelo ao plano YZ e com centro no eixo X, não existem diferenças de nível e tempo entre as ondas que chegam aos ouvidos. Esses círculos formam os conhecidos cones de confusão, que ilustram por que as ITDs e ILDs não auxiliam na determinação da posição vertical de uma fonte sonora.

A localização da posição vertical de uma fonte sonora é determinada principalmente pela interação da orelha com a onda sonora que alcança o receptor. Esta modifica a onda incidente de forma diferente para cada posição da fonte sonora, permitindo que a localização da fonte seja feita de forma bastante precisa. Essa interação também permite distinguir se a fonte está localizada à frente ou atrás do receptor [34, 57].

3.2 Sistemas de Reprodução de Áudio 3D

Conhecendo os principais fatores que indicam a posição no espaço de uma fonte sonora, podemos analisar rapidamente alguns sistemas de reprodução. De forma geral, qualquer sistema deve utilizar fontes sonoras reais (alto-falantes) para criar fontes virtuais localizadas em posições diversas no espaço.

Podemos dividir os sistemas de reprodução em 3 categorias principais:

- Deslocamento de Amplitude
- Reconstrução de Frente de Onda
- Áudio Binaural

3.2.1 Deslocamento de Amplitude

Esse é o mais simples dos métodos usados para criação de fontes sonoras virtuais, muito utilizado em sistemas *stereo*. De forma geral, esses métodos utilizam apenas a ILD para a geração das fontes virtuais. Um conjunto de fontes reais é selecionado, de acordo com a posição desejada para a fonte virtual, e o sinal emitido pela fonte virtual é dividido entre as fontes reais de acordo com uma determinada fórmula (*panning law* ou *panpot law*). Os sinais enviados para cada uma das fontes reais geralmente diferem apenas de um fator de escala.

Diversos sistemas podem ser montados com o princípio geral descrito acima, diferindo na forma como o sinal original é dividido entre as caixas, no número de caixas e também em suas posições. No caso de um sistema *stereo*

(2 caixas de som), a configuração de um triângulo equilátero é considerada como a melhor. Nesta configuração, o receptor é posicionado em dos vértices do triângulo e as caixas são posicionadas nos vértices restantes, voltadas para o receptor. A equação abaixo mostra como um sinal pode ser dividido entre dois alto-falantes de forma a posicionar uma fonte virtual a θ radianos do eixo Y (eixos definidos como na Figura 3.1). G_1 e G_2 são os coeficientes aplicados ao sinal original antes de enviá-lo a cada um dos alto-falantes e C é uma constante representando a intensidade sonora [30]. Uma formulação semelhante pode ser encontrada em [58].

$$\begin{aligned} G_1^2 + G_2^2 &= C \\ 2 \times \operatorname{sen}\theta &= \frac{G_1 - G_2}{G_1 + G_2} \end{aligned}$$

Para sistemas mais complexos, inclusive sistemas que permitam criar fontes virtuais com diferentes elevações, a formulação de Pullki, VBAP (*vector based amplitude panning*) [30, 59], permite derivar facilmente leis eficientes computacionalmente para arranjos bastante variados de caixas de som. O método consiste em escolher um conjunto de alto-falantes posicionados ao redor da posição da fonte virtual e montar uma base de vetores unitários que têm origem no receptor e apontam para cada alto-falante. A posição da fonte virtual pode então ser escrita como uma combinação linear dos vetores dessa base. Os coeficientes dessa combinação linear são então aplicados ao sinal original como fatores de escala e os sinais resultantes são enviados para seus respectivos alto-falantes. Para posicionar uma fonte no plano, dois alto-falantes são necessários. No espaço, três alto-falantes devem ser selecionados para a criação da base. A Figura 3.2 ilustra esse procedimento. L_1 e L_2 são os eixos da base formada pelos alto-falantes A e B, que foram escolhidos por estarem posicionados ao redor da posição da fonte virtual P.

3.2.2 Reconstrução de Frente de Onda

A reconstrução de frente de onda (*wavefront synthesis*) consiste em reproduzir a onda sonora incidente no receptor com a sobreposição de ondas mais simples, provenientes de uma série de caixas de som. Isso pode ser feito com um número relativamente grande de caixas de som (64 ou mais), o que

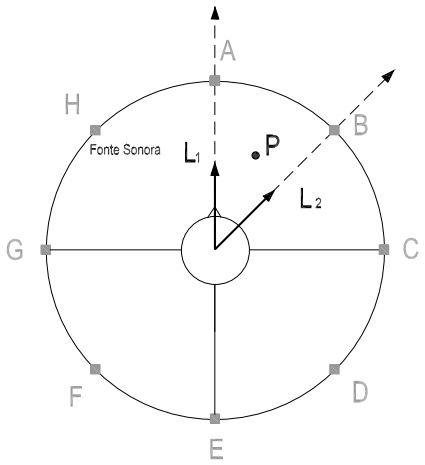

Figura 3.2: VBAP: Selecionando um Conjunto de Alto-falantes

pode acarretar problemas de ordem prática, pois é preciso posicionar de forma precisa e também alimentar cada uma dessas caixas de som.

A determinação dos sinais que devem ser emitidos por cada uma das caixas de som pode ser feita utilizando métodos relativamente complexos que envolvem a discretização do espaço e resolução da equação de onda no domínio do tempo ou das equações de Helmholtz para o problema de análise estacionária no domínio da frequência, que podem ser resolvidas utilizando o método de elementos finitos [50].

Uma alternativa mais simples é a reconstrução da onda sonora através da superposição de ondas planas. Uma onda plana pode ser representada como uma série de cossenos e funções cilíndricas de Bessel [17]. Essa é a abordagem adotada em Ambisonics [24]. Reunindo os primeiros termos dessa série, podemos obter uma série de equações que determinam os sinais que devem ser enviados por cada caixa de som do sistema. Ambisonics fornece uma forma de codificar sinais de áudio 3D de forma independente do sistema de reprodução e também uma forma de particularizar esse sinal codificado para diferentes sistemas de reprodução.

Sistemas que buscam reconstruir uma determinada onda sonora geralmente o conseguem em um determinado ponto do espaço, limitando a movimentação de seus usuários. Mesmo que a posição da cabeça do usuário seja conhecida a cada instante (*head-tracking*), deslocar o ponto de reconstrução ótimo pode não ser simples. A decodificação de sinais Ambisonic não é fácil para arranjos não regulares de caixas de som e a resolução de sistemas derivados do métodos de elementos finitos são computacionalmente caros para aplicações de tempo real.

3.2.3 Áudio Binaural

Uma alternativa bastante interessante para o uso de caixas de som é o uso de fones de ouvido pelos usuários. Não só fica simplificado o problema de posicionar as caixas de som, como também fica simplificado o problema de alimentar as caixas utilizadas no sistema, uma vez que placas de som capazes de alimentar dois canais de áudio diferentes são bastante comuns.

Utilizar os indicadores de posição ILD e ITD com fones de ouvido não apresenta maiores desafios. O maior problema é a falta de interação do som com a orelha do usuário. É essa interação, como vimos anteriormente, a responsável pela informação da elevação da fonte sonora e pela diferenciação de fontes frontais de fontes localizadas atrás do receptor.

Como o som emitido pelo fone de ouvido não sofre a ação da orelha, uma alternativa é fazer com que o som emitido pelos fones de ouvido já contenha as deformações que carregam a informação de posição da fonte sonora. Para gravações, isso pode ser feito com microfones posicionados na entrada do cada canal auditivo. Com isso, pode-se reproduzir com bastante fidelidade os mesmos estímulos de uma experiência real. Para sistemas de realidade virtual, no entanto, geralmente não será possível fazer uma gravação dessa maneira.

O ideal é uma forma de incluir, em um sinal qualquer, as deformações impostas pela interação com a orelha. Isso pode ser feito criando um filtro digital que, quando aplicado a um sinal, cause as mesmas deformações que a interação com a orelha causaria. Esse filtro pode ser criado a partir de uma gravação feita com microfones posicionados na entrada dos canais auditivos de um pulso unitário emitido por uma fonte sonora. A convolução dessa gravação (conhecida como resposta ao impulso) com um sinal causará as mesmas deformações que ocorrem devido à série de reflexões e difrações que ocorrem com o ouvido externo, dando a impressão que o som é proveniente da mesma posição ocupada pela fonte sonora no momento da gravação. Esses filtros são conhecidos como HRTFs (*head-related transfer functions*) [57]. Note que para cada posição do espaço que a fonte virtual pode assumir, devemos ter um par de HRTFs (uma função de transferência para cada ouvido).

Apesar de ser teoricamente simples, a criação desses filtros é bastante trabalhosa e também apresenta dificuldades de ordem prática. Por exemplo: para que seja captada apenas a influência do ouvido externo, da cabeça e do tronco de uma pessoa, a gravação deve ser feita em uma sala que esteja vazia e que apresente o mínimo de reflexões do som emitido pela fonte. Felizmente,

existem bancos de HRTFs disponíveis publicamente [23]. Uma alternativa para a criação de HRTFs é a realização de simulações em computador [50].

Quanto a seu uso, a reprodução de sons através de fones de ouvido para um indivíduo é essencialmente a mesma de uma experiência real quando as HRTFs do mesmo indivíduo são utilizadas [34]. Quando HRTFs não individualizadas são utilizadas, erros de localização vertical e de confusões frente-trás (se a fonte se localiza na frente ou atrás) são agravados. A localização horizontal, no entanto, sofre poucas modificações.

Uma desvantagem do uso de fones de ouvido é que em algumas situações o som pode ser percebido pelo indivíduo como sendo localizado dentro de sua cabeça. Apesar de ser menos comum, o mesmo pode acontecer quando caixas de som são utilizadas [34].

Outro problema associado com o uso de fones de ouvido é a perda do movimento da cabeça em relação às fontes sonoras, uma vez que o fone de ouvido se move junto com a cabeça do indivíduo. Como o movimento da cabeça auxilia bastante na localização de uma fonte sonora, alguns sistemas rastreiam a posição da cabeça do indivíduo, ajustando a posição da fonte virtual de acordo com esse rastreamento.

Áudio binaural também pode ser reproduzido através de um par de caixas de som. Como cada ouvido recebe som de cada uma das caixas, para que a informação de posição da fonte sonora não seja alterada, a superposição das ondas deve ser cancelada. Isso pode ser feito com o uso de filtros conhecidos como *crosstalk-cancellers* [34, 38]. Em [34] é descrita a implementação e os testes de um sistema capaz de reproduzir áudio binaural com caixas de som que se utiliza do rastreamento da posição da cabeça de seu usuário no cálculo dos filtros que cancelam a superposição das ondas.