

Terceiro Trabalho Maior: Simulação/Auralização de Salas por Acústica Geométrica

Data de entrega: 12/12/2014 até 23:55 pelo PACA

Neste trabalho exploraremos uma técnica híbrida de acústica geométrica para simulação de salas, que utiliza fontes virtuais para a construção do início da resposta impulsiva, e preenchimento estocástico para geração da porção final (cauda reverberante da resposta impulsiva). Este trabalho consiste em 2 partes:

- Módulo de simulação de salas para geração de respostas impulsivas que serão armazenadas em arquivos .wav.
- Módulo de auralização de salas a partir de sinais anecóicos e respostas impulsivas.

Estes serão implementados como dois patches Pd, [simula_sala] e [auraliza_sala~]. A especificação destes patches será descrita a seguir.

1 Simulação de salas

O módulo de simulação receberá vários parâmetros para a descrição da geometria da sala, das posições da fonte sonora e do ouvinte, da duração da resposta impulsiva e do arquivo onde a resposta deverá ser gravada. O objeto [simula_sala] deverá receber uma mensagem contendo o nome de um arquivo texto que dará início à simulação. Especificamente, leremos (usando por exemplo o objeto [coll]) os dados a seguir a partir do arquivo texto:

Arquivo de entrada	Descrição
DX DY DZ	Dimensões da sala (m)
FX FY FZ	Posição da fonte (m)
OEX OEY OEZ	Posição do ouvido esquerdo do ouvinte (m)
ODX ODY ODZ	Posição do ouvido direito do ouvinte (m)
α	Coefficiente médio de absorção
T	Tempo total (s) da resposta impulsiva
I	Fração (%) do tempo correspondente à porção inicial
arquivo.wav	arquivo de saída

Esta descrição se refere a uma sala cuboide de dimensões $DX \times DY \times DZ$, onde as posições da fonte e do ouvinte respeitam $FX, OEX, ODX \in [0, DX]$ (e analogamente para os eixos Y e Z). O coeficiente médio de absorção representa a ação das paredes nos raios refletidos, de tal modo que cada reflexão multiplica a amplitude do sinal por $(1 - \alpha)$. A simulação deve produzir uma resposta impulsiva de T segundos, sendo $I \cdot T / 100$ a porção inicial que deve usar o paradigma das fontes virtuais, enquanto o restante da resposta impulsiva será produzido por preenchimento estocástico. O resultado da simulação deve ser armazenado no arquivo wav indicado na última linha da entrada. A resposta impulsiva será produzida em dois canais, sendo um deles referente aos raios recebidos pelo ouvido esquerdo e o outro referente aos raios recebidos pelo ouvido direito.

Este processo poderá demorar um pouco, pelo que implementaremos um sistema simples de notificações para informar o progresso da simulação. Nesta parte inicial seu objeto deverá imprimir na tela principal do Pd (usando o objeto [print]) as informações obtidas do arquivo de entrada. Você pode usar as descrições da tabela acima antes de cada informação obtida. Estas notificações também podem ajudar na depuração do patch.

1.1 Porção inicial da resposta impulsiva: fontes virtuais

A porção inicial da resposta constrói um mapa das fontes virtuais correspondentes às primeiras reflexões. A localização destas fontes possui uma expressão analítica para cada sala de índice (i,j,k), cuja posição inicial corresponde às coordenadas $(i \cdot DX, j \cdot DY, k \cdot DZ)$.

Especificamente, cada reflexão em uma parede faz a fonte passar de uma posição P para a posição (D-P) da sala virtual adjacente, onde D é a medida da sala naquele eixo. Por exemplo, no eixo X as primeiras posições da fonte serão FX (i=0), DX+(DX-FX) (i=1), 2*DX+FX (i=2), e na direção oposta teremos -FX (i=-1), -DX-(DX-FX) (i=-2) e assim por diante. Pode-se verificar então que a sala (i,j,k) terá sua fonte virtual na posição (PX,PY,PZ) onde

$$\begin{aligned} PX &= \lfloor \frac{i+1}{2} \rfloor * 2 * DX + (-1)^i FX \\ PY &= \lfloor \frac{j+1}{2} \rfloor * 2 * DY + (-1)^i FY \\ PZ &= \lfloor \frac{k+1}{2} \rfloor * 2 * DZ + (-1)^i FZ \end{aligned}$$

Os caminhos percorridos pelos raios da sala (i,j,k) até o ouvinte terão portanto comprimentos $c_e = \|(PX, PY, PZ) - (OEX, OEY, OEZ)\|$ (para o ouvido esquerdo) e $c_d = \|(PX, PY, PZ) - (OX, OY, OZ)\|$ (para o ouvido direito) e atravessarão (i.e. serão refletidos por) um total de $|i| + |j| + |k|$ paredes. Considerando que estes raios tinham amplitude unitária na fonte, que a propagação esférica implica em um decaimento de amplitude inversamente proporcional à distância, e que cada reflexão absorve α , teremos que a fonte virtual da sala (i,j,k) contribui na resposta impulsiva com impulso nos instantes $t_e = \frac{c_e}{340}$ segundos (para o ouvido esquerdo) e $t_d = \frac{c_d}{340}$ (para o ouvido direito), com amplitudes dadas por

$$a_e = \frac{(1 - \alpha)^{|i|+|j|+|k|}}{c_e} \quad e \quad a_d = \frac{(1 - \alpha)^{|i|+|j|+|k|}}{c_d}.$$

Impulsos provenientes de salas diferentes que cheguem no mesmo ouvido e no mesmo instante (na mesma amostra do sinal correspondente à resposta impulsiva) devem ter suas amplitudes somadas.

Note que esta simulação usando o modelo de fontes virtuais deve preencher a resposta impulsiva até o instante $I*T/100$ segundos, o que corresponde à distância de $R=340*I*T/100$ metros. Podemos então calcular de forma aproximada a faixa de valores para os índices i, j e k das salas virtuais que pertencem à esfera de raio R considerando a restrição $i * DX \leq R$, ou equivalentemente $i \leq \frac{R}{DX}$, e a restrição esférica

$$(i * DX)^2 + (j * DY)^2 + (k * DZ)^2 \leq R^2.$$

Poderíamos assim escrever o seguinte laço para percorrer todas as salas virtuais que contribuem para a porção inicial da resposta impulsiva (usando uma sintaxe do tipo C/Java):

```
NI = floor(R/DX);
for (i=-NI; i<=NI; i++) {
  NJ = floor(sqrt(R*R-i*i*DX*DX)/DY);
  for (j=-NJ; j<=NJ; j++) {
    NK = floor(sqrt(R*R-i*i*DX*DX-j*j*DY*DY)/DZ);
    for (k=-NK; k<=NK; k++) {
      /* processa o raio da sala (i,j,k) */
    }
  }
}
```

Naturalmente este laço terá que ser repensado usando os objetos do Pure Data. É importante notar que o uso de [triggers] será crucial para ordenar a sequência de valores (i,j,k) produzidos.

Usando o objeto [print] você deve indicar o início do processamento das salas virtuais, e indicar a cada sala os índices (i,j,k), os tempos e amplitudes (t_e, a_e) e (t_d, a_d) produzidos, e também o percentual aproximado de salas processadas até o momento (considerando que nesta parte serão processadas aproximadamente $\frac{\frac{4}{3}\pi R^3}{DX \cdot DY \cdot DZ}$ salas).

1.2 Porção tardia da resposta impulsiva: preenchimento estocástico

A partir do instante $I*T/100$ segundos, trataremos as reflexões de forma aleatória. Para isso consideraremos janelas de 100 ms (ou 0.1 segundos) e anéis que correspondem às salas virtuais que produziram impulsos chegando dentro dessa janela.

Considere o intervalo de tempo $[\Delta, \Delta + 0.1]$. Pensando no modelo de fontes virtuais, as reflexões que chegariam dentro dessa janela teriam se originado a uma distância no intervalo $[340 * \Delta, 340 * (\Delta + 0.1)]$

metros. Considere o anel esférico de raio interno $340 * \Delta$ e raio externo $340 * (\Delta + 0.1)$; o volume total desse anel é de

$$V = \frac{4}{3}\pi(340 * (\Delta + 0.1))^3 - \frac{4}{3}\pi(340 * \Delta)^3 = \frac{4}{3}\pi * 340^3 * [(\Delta + 0.1)^3 - \Delta^3]$$

o que corresponde a $\frac{V}{DX * DY * DZ}$ vezes o volume de uma sala. Podemos concluir então que o número médio de raios provenientes desse anel e que atingem a janela $[\Delta, \Delta + 0.1]$ é de $N = \frac{V}{DX * DY * DZ}$ raios.

Faremos então, para cada ouvido/canal, um sorteio uniforme de N instantes $t_o, t_1, \dots, t_{N-1} \in [\Delta, \Delta + 0.1]$ que serão preenchidos de forma aditiva (ou seja, raios incidentes na mesma amostra são somados), e considerando que o raio do instante t_l terá uma atenuação individual em função da distância ($\frac{1}{340 * t_l}$). Para o tratamento das reflexões, sortearemos o número médio de paredes atravessadas, independentemente para cada raio, dentre 4 valores possíveis: $\frac{340 * t_l}{DX}$ (com probabilidade 1/7), $\frac{340 * t_l}{DY}$ (com probabilidade 1/7), $\frac{340 * t_l}{DZ}$ (com probabilidade 1/7), que correspondem à incidência paralela a um dos eixos X, Y ou Z, e $\frac{340 * t_l}{\sqrt{3}} (\frac{1}{DX} + \frac{1}{DY} + \frac{1}{DZ})$ (com probabilidade 4/7), que corresponde à incidência pela diagonal de algum dos ortantes. Sendo β_l este número sorteado de reflexões, a amplitude do raio que chega no instante t_l será portanto

$$a = \frac{(1 - \alpha)^{\beta_l}}{340 * t_l}.$$

A cada intervalo de tempo de 100 ms seu patch deverá gerar mensagens informando o valor de Δ , o número de raios gerados, e o percentual de intervalos processados nesta porção tardia da resposta impulsiva. Finalmente, a resposta impulsiva (em dois canais) assim construída deve ser gravada em um arquivo wav estéreo. Você pode usar os objetos `[soundfiler]` ou `[writesf~]` para isso.

2 Auralização

O objetivo da auralização é utilizar as respostas impulsivas geradas no primeiro módulo como filtros FIR na convolução com sinais sonoros anecoicos¹ (por exemplo, sinais de instrumentos digitais puros ou gravações feitas em ambientes sem reverberação). Você pode obter gravações anecoicas para usar em seus testes em vários repositórios online (basta googlear anechoic recordings).

O objeto `[auraliza_sala~]` deve usar um parâmetro de inicialização com o nome do arquivo wav contendo a resposta impulsiva estéreo, e processar um sinal anecoico mono (obtido a partir de um `[inlet~]`) produzindo o resultado da auralização estéreo através de dois `[inlet~]`. Este processamento deve ser implementado diretamente (aqui não é para usar `[FIR~]` nem outros objetos de propósito semelhante).

A convolução do sinal de entrada com o arquivo de áudio interpretado como um filtro FIR será computada de maneira análoga ao objeto `[amplitude~]` implementado no primeiro trabalho maior. Naquele objeto era necessário memorizar um trecho do sinal incluindo um ou mais blocos anteriores, em função do tamanho N da somatória. Aqui o N será dado pela duração do arquivo contendo a resposta impulsiva; sendo $x[n]$ o sinal da entrada e $y[n]$ a resposta impulsiva de tamanho N, a saída do filtro de auralização será

$$z[n] = (x * y)[n] = \sum_{m=0}^{N-1} y[n]x[n - m].$$

Isso será feito utilizando-se o mesmo sinal de entrada $x[n]$ e cada um dos canais $y_e[n]$ e $y_a[n]$ que correspondem às respostas impulsivas simuladas.

Teste seu simulador/auralizador com configurações diversas, por exemplo com salas pequenas (e.g. $3 \times 4 \times 5$), médias (e.g. $12 \times 16 \times 20$) e grandes (e.g. $25 \times 50 \times 200$), com durações de resposta impulsiva estimadas pela expressão do T_{60} de Sabine

$$T_{60} = \frac{0.16 \cdot V}{\sum a_i s_i} = \frac{0.16 \cdot DX \cdot DY \cdot DZ}{2\alpha(DX \cdot DY + DX \cdot DZ + DY \cdot DZ)},$$

com porção inicial da resposta impulsiva pequena (por exemplo 10% do T_{60}) e grande (por exemplo 40% do T_{60}) e com coeficientes de absorção $A = 0.1$ e $A = 0.2$. Para criar simulações binaurais verossímeis, é importante manter uma distância entre os pontos OE e OD compatíveis com as dimensões humanas (aproximadamente 0.21 m).

Bom Trabalho!

¹É possível usar também quaisquer arquivos de áudio neste processo, porém nesse caso haveria um acoplamento entre a reverberação presente no sinal original e aquela introduzida pela sala simulada.