

AVALIAÇÃO DA QUALIDADE DAS SIMULAÇÕES E AURALIZAÇÕES DA PLATAFORMA VIRTUSOUND

D. Alarcão, J. L. Bento Coelho
CAPS, DEEC, Instituto Superior Técnico
Av. Rovisco Pais, 1
P-1049-001 Lisboa, Portugal
diogo.alarcao@ist.utl.pt

Resumo

A plataforma Virtusound foi desenvolvida para permitir a simulação da acústica de salas e a correspondente auralização em tempo real através da utilização da tecnologia binaural. O módulo de cálculo do sistema utiliza um método das fontes imagem acelerado e um método de radiosidade hierárquica dependente do tempo para a computação das respostas ao impulso, binaurais, dos recintos a simular. O módulo de convolução utiliza as respostas preparadas pelo módulo de auralização para efectuar convoluções segmentadas de baixa latência com sinais áudio arbitrários. Testes informais preliminares mostraram a precisão do sistema e que o som tridimensional criado era bastante realístico, tanto do ponto de vista da espacialização sonora tridimensional como do ponto de vista da assinatura reverberante dos recintos. No entanto, entendeu-se ser necessário efectuar uma avaliação mais exaustiva e precisa a fim de se garantir a qualidade geral da plataforma. Este artigo apresenta pois os primeiros resultados dos trabalhos de avaliação das capacidades da plataforma Virtusound em termos dos parâmetros objectivos previstos e em termos das auralizações geradas.

Palavras-chave: auralização, tempo real, acústica virtual, tecnologia binaural.

Abstract

The Virtusound platform allows the real time simulation and auralization of virtual enclosures through binaural technology. The calculation module of the system uses an accelerated mirror image sources method together with a time dependent hierarchical radiosity method to compute the binaural room impulse responses of the simulated enclosures. The segmented convolution module uses the impulse responses prepared by the auralization module for a low latency segmented convolution step with arbitrary audio signals. Informal listening tests showed the system to be overall precise and that the generated 3D sound was conceivable realistic, both from the point of view of sound spatialization and from the point of view of the rooms' reverberant signatures. However, it became clear that a more thorough evaluation of the system was needed in order to guarantee the overall quality of the platform. This paper presents thus a first assessment on the capabilities of the Virtusound system in terms of the predicted objective parameters and in terms of the produced auralizations.

Keywords: auralization, real time, virtual acoustics, binaural technology.

PACS no. 43.55.KA, 43.55.HY

1 Introdução

O presente artigo apresenta a plataforma Virtusound, uma plataforma recentemente desenvolvida e implementada para a simulação e auralização da acústica de salas em tempo real, e também os resultados de uma primeira avaliação objectiva e subjectiva da qualidade das simulações e das auralizações obtidas.

Sistemas para auralização em tempo real foram já desenvolvidos anteriormente por outros grupos de investigação [1, 2], mas, no entanto, a plataforma Virtusound possui características únicas, sendo as mais importantes a capacidade de simular em tempo real as reflexões especulares e difusas de um recinto arbitrário e estar implementada num único computador multi-threaded, não existindo portanto a necessidade de redes de comunicação.

O módulo de cálculo da plataforma utiliza uma combinação de um método de fontes imagem acelerado com um método de radiosidade hierárquica [4] dependente do tempo para a computação das respostas ao impulso, binaurais, dos recintos simulados [3, 5]. O módulo de auralização prepara as respostas binaurais para o módulo de convolução (streaming) segmentada efectuar a filtragem com sinais áudio arbitrários.

O sistema Virtusound foi desenvolvido pelo Grupo de Acústica e Controlo de Ruído do CAPS – Instituto Superior Técnico de Lisboa.

2 Arquitectura do sistema

O diagrama de blocos do sistema Virtusound encontra-se ilustrado na Figura 1.

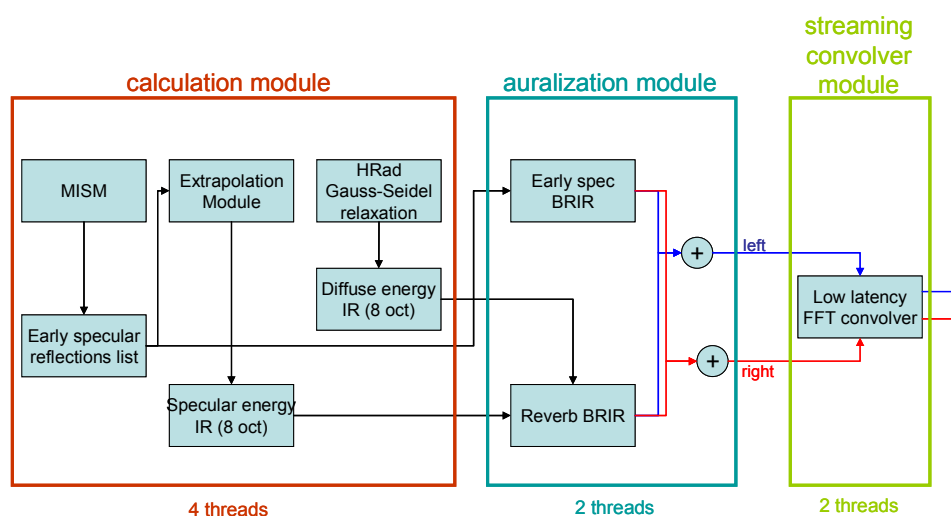


Figura 1 – Arquitectura do sistema Virtusound

Como pode ser constatado, o sistema é composto por 3 módulos principais: cálculo, auralização e convolução. O sistema encontra-se presentemente implementado num computador baseado em CPU Intel I7 Extreme, correndo sobre 8 threads simultâneas.

3 Descrição do sistema

3.1 Módulo de cálculo

O módulo de cálculo combina dois métodos. Um primeiro método, referente ao cálculo das componentes reflectidas de forma especular, e um outro referente às componentes reflectidas de forma difusa pelas paredes do recinto.

A determinação das componentes especulares do som reflectido foi implementada usando-se um método de fontes imagem acelerado. Por outro lado, as componentes difusas resultam do estudo da radiosidade hierárquica dependente do tempo [4] do recinto em questão. Este último permite ao utilizador definir os parâmetros necessários para garantir a máxima precisão e rapidez do mesmo. Ambos os métodos são executados de forma independente através do uso de diferentes threads de computação.

3.1.1 Método das fontes imagem acelerado

Sabendo-se que o número de fontes-imagem aumenta de forma exponencial com a ordem de reflexão requerida:

$$M + M(M - 1) + \dots + M(M - 1)^{K-1} = \frac{M(M - 1)^K - 1}{M - 2} \approx (M - 1)^K \quad (1)$$

onde M representa o número de paredes do espaço em questão, e K a ordem reflexão, o cálculo geométrico das fontes imagem de um determinado espaço, costuma ser limitado a uma ordem de reflexão não superior a 5, tipicamente de 3 a 5.

3.1.1.1 Técnicas de aceleração do método das fontes imagem

A fim de se obter a lista de fontes imagem no menor tempo possível, quatro técnicas de aceleração são utilizadas.

- “*Back-Face Culling*”

“Back-face culling” é de implementação simples e permite reduzir o número de cálculos de possíveis fontes imagem uma vez que limita esse cálculo apenas às paredes cuja orientação é visível da fonte a ser espelhada. Esta decisão é feita através do uso dos vectores normais aos polígonos que compõem cada parede.

- *Combinações impossíveis*

Esta técnica permite ter conhecimento à partida de combinações impossíveis de pares de polígonos (uma parede é constituída por um ou mais polígonos), ou seja, situações geométricas em que o som reflectido por um polígono não pode atingir directamente um outro. Desta forma, é possível limitar as hipóteses de reflexão para cada polígono, ou seja, o número de possíveis novas fontes-imagem e por sua vez eliminar os cálculos de validade e visibilidade das mesmas. A lista completa das combinações impossíveis é armazenada num pré-processamento do sistema, não tendo qualquer custo computacional durante a execução do mesmo.

- “*View-Frustum*”

O uso da técnica denominada por *View-Frustum* permite eliminar de uma forma simples um conjunto de fontes imagem obtidas de forma válida mas que não seriam visíveis de nenhuma posição dentro da sala em questão e, portanto, não participariam na resposta impulsiva da mesma. Na prática, a técnica consiste em, pegando numa determinada fonte imagem construir o seu *View-Frustum* (como demonstra a Figura 2). De seguida, apenas se consideram como paredes válidas para o cálculo de fontes-imagem de ordem superior aquelas que estejam dentro do *View-Frustum* construído. Esta

técnica permite então, aumentar o desempenho e rapidez do método, uma vez que anula automaticamente muitas possíveis fontes imagem.

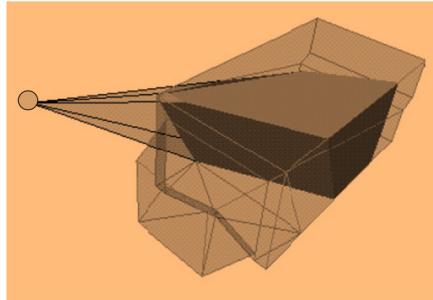


Figura 2 – Exemplo de um View-Frustum para eliminar paredes na construção de novas fontes-imagem de ordem de reflexão superior

- *Aglomeração de Polígonos*

Esta técnica utiliza o facto de que na maioria dos recintos existentes, todos os seus polígonos de entrada (que podem ser em grande número), pertencem a um pequeno conjunto de planos tridimensionais. Portanto, numa fase de pré-processamento, um algoritmo de “*clustering*” constrói um único *polígono-pai* incluindo todos os polígonos do mesmo plano tridimensional – ver Figura 3. Este algoritmo reduz então a complexidade do problema, uma vez que o número de combinações entre polígonos é automaticamente reduzido.

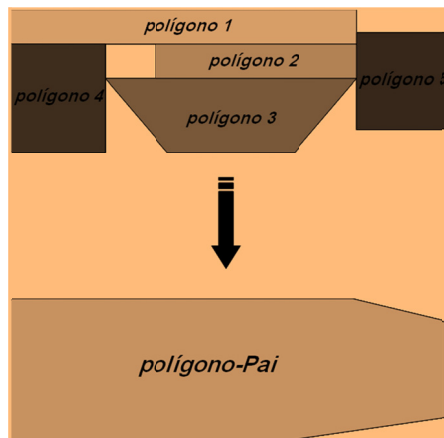


Figura 3 – Exemplo de aglomeração de uma série de polígonos num único polígono-pai

Na prática, este algoritmo baseia-se num estudo simplificado da sala, através do uso dos *polígonos-pai* (convexos) para o cálculo das fontes imagem válidas e visíveis. Desta forma, apenas quando uma fonte imagem é considerada válida e visível em relação a um dado *polígono-pai*, é que se verifica, que polígono de entrada específico a originou, e se conclui acerca das características da reflexão em questão. Esta verificação não tem um grande custo computacional, uma vez que apenas se necessita de determinar a que polígono de entrada corresponde o ponto de intercepção do raio sonoro com o *polígono-pai*. Obviamente, e tendo em conta a complexidade dos polígonos de cada sala, pode dar-se o caso de uma fonte imagem ser válida e visível no estudo simplificado de *polígonos-pai* e não o ser usando todos os polígonos e entrada separadamente. Um exemplo disto é quando o ponto de intercepção coincide com um “buraco” no *polígono-pai* – ver Figura 2. No entanto, estes casos constituem normalmente uma minoria, já que os *polígonos-pai* são construídos como o *convex-hull* de todos os polígonos de entrada pertencentes a um dado plano tridimensional.

A Figura 4 ilustra o módulo de interface gráfica do sistema implementado, nela podemos observar os caminhos sonoros *fonte-receptor* considerados válidos e visíveis para uma ordem de reflexão igual a 3, no Centro de Congressos do Instituto Superior Técnico de Lisboa.

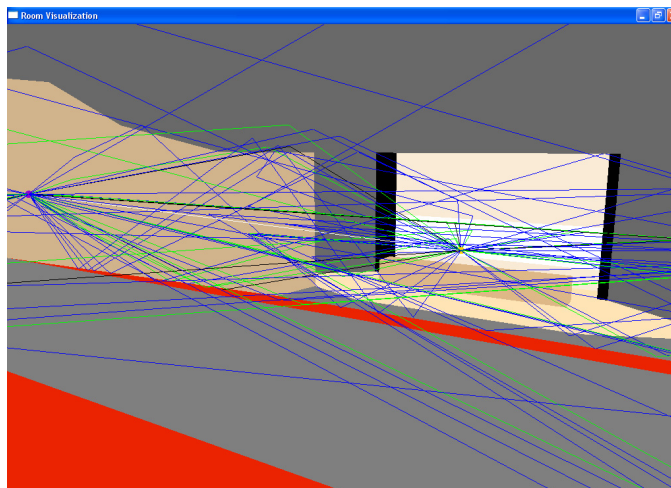


Figura 4 – Interface gráfica mostrando um exemplo dos “caminhos sonoros” fonte-receptor, calculados até à terceira ordem de reflexão

3.1.1.2 Técnicas para ordens de reflexão superiores

Uma vez que em termos computacionais existe a limitação de não se exceder a quinta ordem de reflexão no cálculo geométrico das fontes imagem, sugere-se a utilização dos resultados obtidos, numa extrapolação estatística e determinística para ordens de reflexão superiores. Deste modo, consegue-se construir a resposta impulsiva completa da sala em estudo, com a reverberação devida às componentes reflectidas de forma especular.

O módulo de extrapolação recorre aos seguintes parâmetros, obtidos durante a fase de construção geométrica das fontes-imagem:

- > Número de imagens visíveis, por ordem de reflexão;
- > A distância média do receptor às fontes-imagem, para cada ordem de reflexão;
- > O desvio padrão das distâncias, por ordem de reflexão;
- > O coeficiente de reflexão médio, por ordem de reflexão;
- > O desvio padrão dos coeficientes de reflexão, por ordem de reflexão;
- > O coeficiente de difusividade médio, por ordem de reflexão;
- > O desvio padrão dos coeficientes de difusividade, por ordem de reflexão.

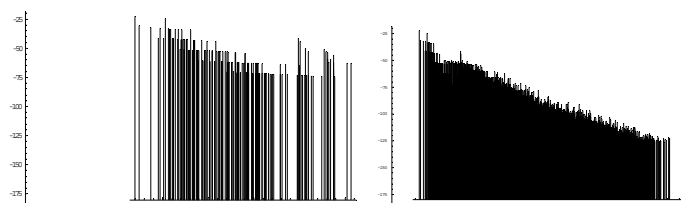


Figura 5 – Ecogramas derivados de reflexões especulares. Esquerda: calculado até à ordem 6. Direita: calculado por extrapolação até à ordem 45

Desta forma a resposta impulsiva especular completa da sala será obtida através da soma das componentes calculadas exactamente como descrito em 3.1 com as componentes calculadas através de extrapolação.

3.1.2 Método da Radiosidade Hierárquica dependente do tempo

O cálculo das componentes reflectidas de forma difusa pelas paredes da sala, numa abordagem de elementos finitos, levanta dois problemas principais:

Problema 1: Entende-se pelo exposto anteriormente, que a situação ideal para o método das fontes-imagem implementado é que os polígonos que constituem a geometria de cada parede da sala sejam o maior possível, enquanto que numa aproximação por elementos finitos é desejável que os polígonos constituam uma rede de vários fragmentos (“*patches*”) mais pequenos.

Problema 2: Se M paredes iniciais são divididas em n fragmentos, então o número de ligações para troca de energia, isto é, o número aproximado de factores de forma, será da ordem $O(n^2)$, e, consequentemente, o esforço de cálculo será elevado.

Uma solução para ambos os problemas é adoptar uma abordagem de multi-resolução através de ligação hierárquica, em que os fragmentos são gerados adaptativamente a partir da geometria das paredes. Nesta abordagem, o número de ligações (factores de forma) é $O(M^2+n)$, permitindo assim uma considerável economia computacional.

O critério de paragem na subdivisão hierárquica é baseado em primeiro lugar na área mínima absoluta dos fragmentos (definida pelo utilizador) e, em segundo lugar, numa condição limiar para os factores de forma entre pares de fragmentos (também definida pelo utilizador).

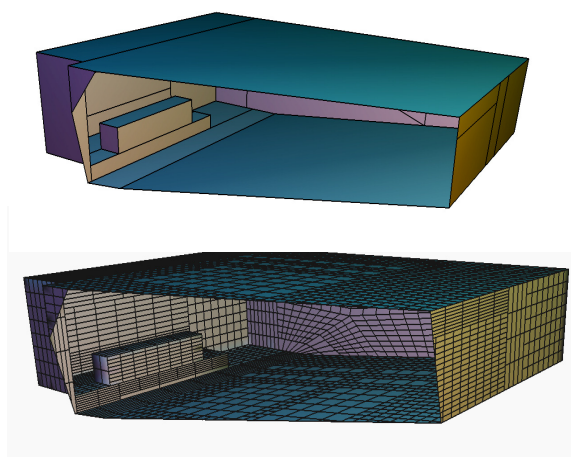


Figura 6 – Subdivisão hierárquica. Em cima: polígonos de entrada. Em baixo: Patches resultantes.

Os factores de forma são calculados usando-se a fórmula analítica do factor de forma de uma área diferencial para um disco paralelo, sendo o factor de visibilidade associado a cada factor de forma (com oclusão) calculado através de um processo de “*ray-casting*”.

A solução para as equações dependentes do tempo é obtida através de relaxação “*Gauss-Seidel*”, usando-se uma operação de “*push-pull*” de energia por cada hierarquia de fragmentos.

De modo a limitar o crescimento exponencial de reflexões difusas com crescentes ordens de reflexão, é usado um algoritmo de aglutinação com uma taxa de amostragem interna.

Em cada iteração, a energia é reunida em pontos de recepção predefinidos, calculada através conhecida lei do co-seno de Lambert.

3.2 Módulo de Auralização

O módulo de auralização sistema Virtusound funciona em threads assíncronas, de forma independente do resto dos outros módulos, necessitando apenas de aceder periodicamente à lista actualizada de

todas as fontes imagem válidas e visíveis e aos ecogramas das componentes difusas da situação acústica em estudo. O módulo de auralização implementa uma abordagem de banda larga em termos da dependência na frequência, e encontra-se implementado de forma a processar o som através da tecnologia binaural.

3.2.1 Auralização das Componentes Especulares

Como requisito para o bom funcionamento deste módulo exige-se que, à lista de fontes imagem previamente calculada, acrescentemos a lista detalhada de todas as paredes que as originaram, bem como os ângulos de incidência dessas mesmas fontes-imagem em relação à posição do receptor. Tendo esta informação disponível, o módulo de auralização constrói os respectivos filtros lineares para caracterizar cada caminho acústico válido entre a fonte real e o receptor.

Elementos tal como a directividade da fonte real, atenuação no ar e absorção das paredes são representados através de filtros lineares invariantes no tempo, numa escala logarítmica de frequência, em bandas de oitava, com 10 valores entre os 31,5 Hz e os 16 kHz. Estes filtros possuem apenas valores de magnitude, através dos quais, e recorrendo-se a uma transformada de Hilbert (FFT), se procede à reconstrução da respectiva fase mínima.

A fim de obter a resposta impulsiva binaural, o resultado dos filtros anteriores é convolucionado no domínio da frequência, após interpolação, com o conjunto de HRIR correspondente ao ângulo de incidência das fontes-imagem no receptor.

O módulo de auralização utiliza o banco de dados HRIR do IRCAM. Esta base de dados tem uma resolução de 15° para o ângulo de azimute, enquanto a resolução vertical é constituída por 10 ângulos de elevação, começando em -45° e terminando em 90°, ou seja, com intervalos de 15°. Como um todo, existem 187 pares de HRIR. Cada conjunto de dados HRIR tem uma taxa de amostragem de 44,1 kHz e 24 bits de quantização, o que consiste em 8192 amostras. Cada conjunto de dados HRIR é pré-processados a fim de se obterem três novos conjuntos de dados: o primeiro com 512 amostras, o segundo com 128 amostras, e o terceiro com 32 amostras. O conjunto de 512 amostras é utilizado para o processamento do som directo, enquanto o conjunto de 128 amostras é utilizado para o processamento das componentes especulares da lista de imagens visíveis. O terceiro conjunto é utilizado para processamento no módulo de extrapolação.

Todas as convoluções são feitas no domínio da frequência dupla usando-se radix-2 FFTs.

O resultado final desta etapa é a resposta impulsiva binaural das componentes especulares do som reflectivo, tendo em conta o par fonte-receptor em estudo.

3.2.2 Auralização das componentes difusas

O resultado do método de radiosidade hierárquica dependente do tempo consiste num ecograma de energia para cada uma das 10 bandas de oitava.

Cada ecograma é considerado como o envelope do quadrado da resposta impulsiva da pressão, para cada banda de frequências. De modo a reconstruir a resposta impulsiva equivalente em termos da pressão sonora, implementou-se um procedimento que usa um ruído branco filtrado.

Detalhadamente, uma janela temporal gaussiana com uma largura de 10 ms é aplicada em conjunto com um filtro de magnitude equivalente, obtido a partir do cálculo dos ecogramas de energia, ao ruído branco. Este procedimento é repetido em intervalos de tempo cada vez maiores até que a janela de tempo fique fora do intervalo de tempo relevante. Para a simulação da resposta binaural, são criadas duas respostas não-correlacionadas que são equalizadas através de uma HRTF média de acordo com uma incidência difusa.

3.3 Módulo de convolução

As respostas impulsivas binaurais para as componentes especulares e difusas são combinadas em duas respostas finais, uma para cada ouvido. Estas respostas são então, convolucionadas com um sinal de áudio com uma frequência de amostragem de 44,1 kHz e 16 bits de quantização, segundo uma abordagem de baixa latência através do método de *overlap-add* tipo Gardner, utilizando-se a biblioteca FMOD [6].

A reprodução do sinal de áudio resultante é feita utilizando-se uns auscultadores de alta-fidelidade com equalização de campo difuso.

4 Avaliação do sistema Virtusound

4.1 Desempenho computacional

A Tabela 1 inclui alguns valores temporais típicos para a duração de diferentes tarefas realizadas pelo sistema Virtusound em “*run-time mode*”.

Tabela 1: Tempos de cálculo típicos relativos a diferentes tarefas do sistema Virtusound

Task	Calc Time [ms]
MISM early images (order 5)	85
Extrapolation module	0,06
HR Gauss Seidel relaxation	500
Diffuse energy IR at receiver	6
Early specular refl BRIR	5,5
Reverb BRIR	11
FFT Convolver (1024 samples)	10

4.2 Avaliação objectiva

Para se obter uma avaliação objectiva das capacidades e da qualidade do sistema Virtusound, foram realizadas medições acústicas em 5 salas existentes no Campus da Alameda do IST, as quais foram depois comparadas com os valores dos mesmos parâmetros obtidos por simulação pelo sistema Virtusound para modelos dessas salas.

O conjunto das salas consideradas englobou desde salas de aula pequenas até ao auditório do centro de congressos do IST (volumes entre 150 m³ e 220 m³). Os parâmetros objectivos da qualidade de acústica de salas considerados foram: T₃₀, EDT, D₅₀, C₈₀ e L_p.



Figura 7 – Fotografias de algumas salas consideradas nas medições acústicas

A Figura 8 ilustra os resultados obtidos para a sala de aula denominada VA2 (primeira sala da Figura 7). Como se pode constatar, os valores obtidos por simulação pelo sistema Virtusound apresentam uma boa concordância com os valores obtidos por medição, excepto nas duas bandas de frequência

mais baixas, onde efeitos de origem modal começam a ter importância, e que não são considerados nos modelos utilizados no módulo de cálculo.

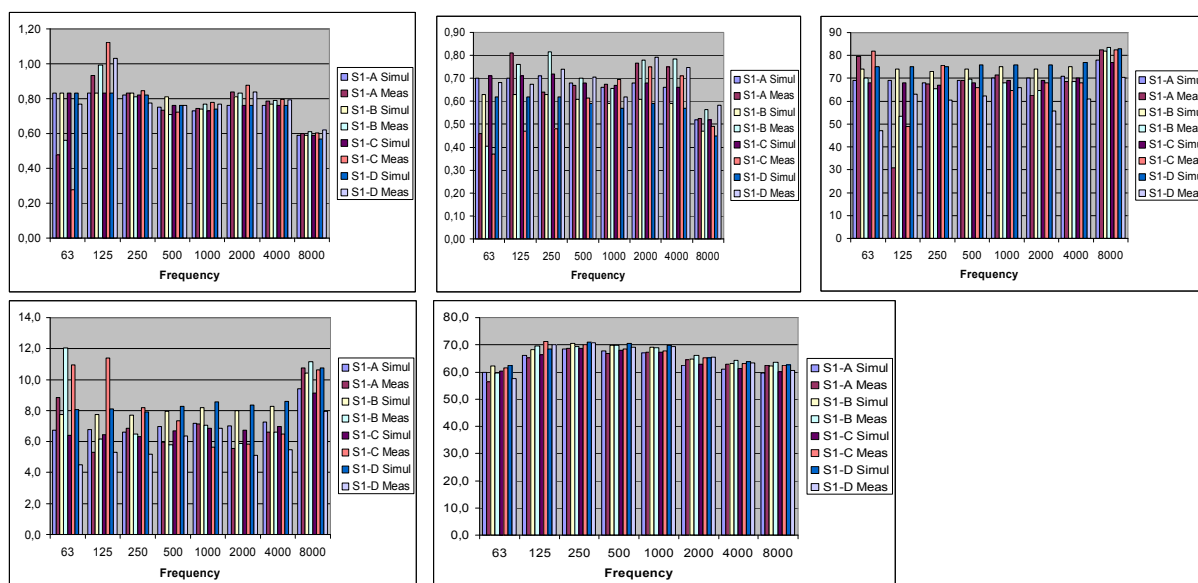


Figura 8 – Comparação de parâmetros objectivos T_{30} , EDT, D_{50} , C_{80} e L_p para a sala VA2

Adicionalmente, as mesmas salas foram também modeladas num conhecido software comercial de modelação da acústica de salas, e novamente procedeu-se à comparação entre os parâmetros calculados por este software e pelo sistema Virtusound. Na Figura 9 mostra-se os resultados desta comparação obtidos para três das salas consideradas.

4.3 Avaliação subjectiva

A fim de se obter uma avaliação subjectiva das capacidades da plataforma Virtusound em termos das auralizações geradas foram efectuados vários testes de audição por um painel de 15 pessoas composto por investigadores, estudantes, músicos e pessoal administrativo. As audições foram efectuadas em frente de um écran de computador com a interface do sistema activa (OpenGL GUI [7]), podendo cada sujeito interagir directamente com o recinto em simulação, através de deslocação e rotação do receptor e da fonte (*“interactive walkthroughs”*) – Figura 10.

Os testes utilizaram placas de som Echo Gina 3G PCI, auscultadores Beyerdynamic DT990 Pro e diferentes excertos de fala e de música (cada um com 5 minutos de duração).

Foram colocadas 3 questões a cada um dos sujeitos:

- a) O som 3D parece natural (sem colorações)?
- b) A direcção do som percebida é consistente com a localização relativa da fonte sonora na interface visual?
- c) A reverberação e o *loudness* são consistentes com o recinto simulado e com a localização da fonte e do receptor?

Os resultados destes testes revelaram que as auralizações geradas eram convincentes e agradáveis de ouvir, e em termos das respostas às 3 questões colocadas, os resultados foram os seguintes:

- i) 15 em 15 julgaram as auralizações como naturais
- ii) 13 em 15 julgaram as direcções do som consistentes com a localização da fonte sonora (*“front-back confusion”*)
- iii) 10 em 15 julgaram a reverberação percebida e o *loudness* como correctos dado o tipo de recinto e a localização relativa entre a fonte e o receptor.

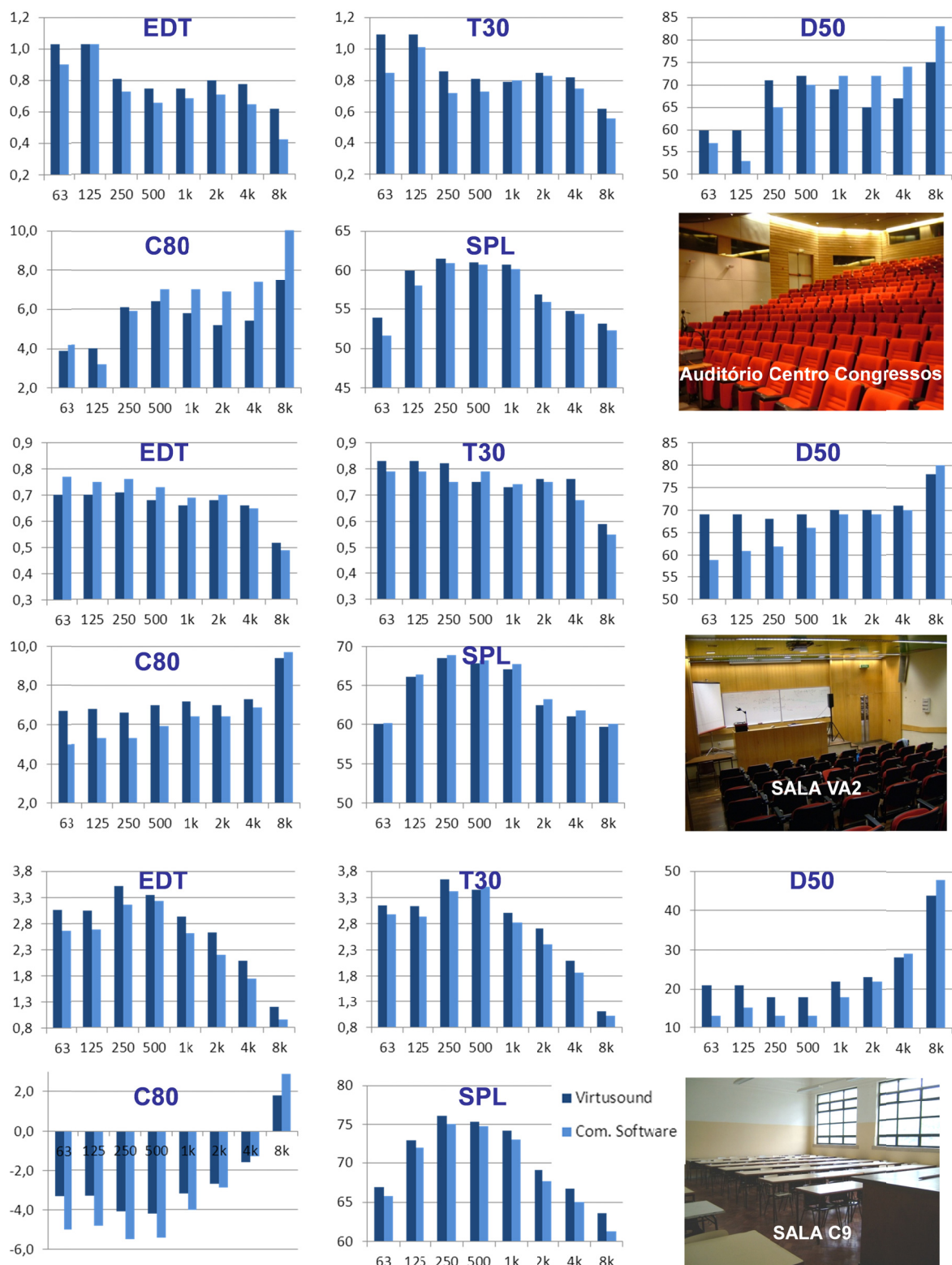


Figura 9 – Comparação de parâmetros objectivos T_{30} , EDT, D_{50} , C_{80} e L_p obtidos pelo sistema Virtusound e por software comercial



Figura 10 – Testes auditivos para avaliação subjectiva da plataforma Virtusound

5 Conclusões

A plataforma Virtusound descrita neste trabalho permite a simulação e auralização em tempo real da acústica de salas, permitindo ainda a interação do utilizador ao nível do movimento do receptor e da fonte sonora.

Testes efectuados para avaliação objectiva e subjectiva do sistema permitem concluir das boas capacidades e da boa qualidade tanto das simulações do campo sonoro dentro de recintos como das auralizações geradas.

Referências

- [1] T. Lentz, D. Schröder, M. Vorländer e I. Assenmacher; Virtual Reality System with Integrated Sound Field Simulation and Reproduction. *EURASIP Journal on Advances in Signal Processing*, Vol. 2007
- [2] M. Noisternig, B. Katz, S. Siltanen e L. Savioja; Framework for Real-Time Auralization in Architectural Acoustics. *Acta Acustica united with Acustica, Technical and Applied Papers*, Vol. 94 (2008), 1000 – 1015
- [3] D. Alarcão e J.L. Bento Coelho; A Combined Method for Room Acoustics integrating Specular and Diffuse Reflections. *In Proceedings of 18th. International Conference on Acoustics*, Kyoto, Japan, April, 2004
- [4] P. Hanrahan, D. Salzman e L. Aupperle; A Rapid Hierarchical Radiosity Algorithm. *Computer Graphics*, 25(4):197-206, 1991
- [5] D. Alarcão, Acoustic Modelling for Virtual Spaces; *PhD Thesis*, Instituto Superior Técnico, TULisbon, 2005
- [6] <http://www.fmod.org>
- [7] <http://www.opengl.org>