Auralização em Tempo Real num Ambiente de Realidade virtual

Ricardo Casaleiro
DETI/IEETA

Paulo Dias
DETI / IEETA

Guilherme Campos
DETI / IEETA

Beatriz Sousa Santos
DETI / IEETA

Universidade de Aveiro
Campus Universitário de Santiago, 3810-193 Aveiro
{ a26197, agrcampos, paulo.dias, bss }@ua.pt

Sumário

A realidade virtual (VR) é uma área de crescente importância, com uma gama de aplicações também ela crescente. Nos últimos anos têm surgido várias empresas que disponibilizam hardware e software para ambientes cada vez mais complexos, imersivos e com dispositivos de interacção cada vez mais avançados. Não obstante o seu grande potencial, este tipo de ambientes ainda tem muitos desafios a ultrapassar ao nível da interacção e do realismo.

No desenvolvimento desses ambientes é dado particular ênfase à vertente visual, mas uma imersão e interacção só poderá ser completa se incluir também os outros sentidos. Em particular a audição permite aumentar de uma forma significativa tanto a sensação de imersão como a capacidade de localização num ambiente virtual. O trabalho apresentado neste artigo visa o desenvolvimento de algoritmos de auralização realistas e eficientes para permitir estimular em tempo real os utilizadores enquanto os mesmos se encontram imersos num ambiente de Realidade Virtual. São apresentados os principais algoritmos utilizados para calcular o som directo e a reverberação tardia assim como alguns testes perceptuais preliminares para tentar validar as opções tomadas e eventualmente corrigir algumas delas.

Palavras-chave

Auralização, Head Related transfer Functions, Modelos 3D, Realidade Virtual, Reverberação.

1. INTRODUÇÃO

A recriação artificial de um ambiente sonoro é chamada *auralização*. Consiste em reproduzir (normalmente através de auscultadores) os estímulos sonoros que atingiriam os tímpanos do ouvinte se este se encontrasse na sala que se pretende simular, para uma dada posição da(s) fonte(s) e do próprio ouvinte.

Assumindo que existe um caminho directo entre a fonte e o ouvinte, o ouvinte irá ouvir primeiro o ***som directo***, seguido das suas ***primeiras reflexões***. Após algumas dezenas de milissegundos, o número de reflexões torna-se bastante elevado e cria o que é denominado por ***reverberação tardia***. Esta é independente da posição sonora e fornece informação espacial sobre o ambiente onde o ouvinte se encontra permitindo assim ao utilizador uma melhor localização dos sons virtuais [1]. Na Figura 1 é possível ter uma noção temporal mais clara sobre as distintas fases da percepção do som.

Para permitir uma auralização realista é necessário ter em conta todos estes aspectos. Neste artigo são apresentados os métodos e algoritmos utilizados para modelar todas as fases do som [1]: a utilização de *Head Related Transfer Functions* (HRTF) para a auralização em espaço aberto, assim como para simular as primeiras reflexões, e alguns dos algoritmos mais comuns para calcular a reverberação tardia. No final do artigo são apresentados alguns testes perceptuais preliminares assim como conclusões e trabalho futuro.

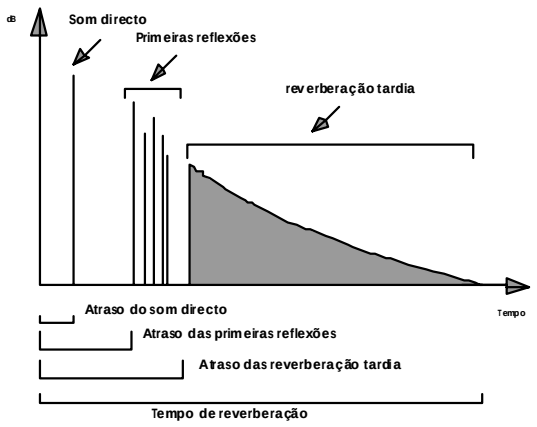


Figura 1: Fases distintas da percepção do som [6]

2. AURALIZAÇÃO EM ESPAÇO ABERTO

2.1 Algoritmo

O comportamento acústico de uma sala é descrito pela sua *resposta impulsional* (*RIR-Room Impulse Response*) – dita *binaural* no caso de se considerar um par de sinais (um para cada ouvido). A auralização baseia-se na convolução da *RIR* binaural com gravações anecóicas do som emitido pela fonte evitando assim a influência de uma sala específica numa dada gravação.

É possível gravar a função de transferência relacionada com a cabeça (HRTF) tendo em conta uma posição específica no espaço. No nosso software, utilizamos um conjunto de pares de HRTF bastante completas recolhidas através de um '*KEMAR dummy head microphone*' pelo *Massachusetts Institute of Technology*. Estes dados estão disponíveis para pesquisa através da página do MIT [7]. Este conjunto de amostras foram apenas recolhidas com o ouvido esquerdo do boneco KEMAR, dado que as HRTF para ângulos simétricos em relação aos hemisférios esquerdo e direito são semelhantes.

As HRTF são usadas para simular a posição específica de uma fonte no espaço. Isso é conseguido através da convolução do som da fonte sonora e as HRTF da posição específica a simular. O som de entrada é duplicado, gerando dois sinais, um para cada ouvido, aos quais são aplicado o par de HRTF pretendido. Por fim, o resultado é reproduzido ao utilizador através de auscultadores.

Quando o utilizador se movimenta, alterando a informação de um dos ângulos de rotação descritos no ponto anterior, é necessário actualizar a informação que o mesmo tem sobre o mundo sonoro, ou seja, cada fonte virtual vai alterar a sua posição em relação ao utilizador. Para conseguir esse efeito é apenas necessário calcular os ângulos de azimute e elevação tendo em conta a nova posição do utilizador e usar a HRTF correspondente. No entanto, se este processo for feito em tempo real e enquanto a fonte virtual estiver activa, vão ser gerados artefactos (esse fenómeno é normalmente designado por *clicking*) em cada transição de filtros HRTF (ver Figura 2).

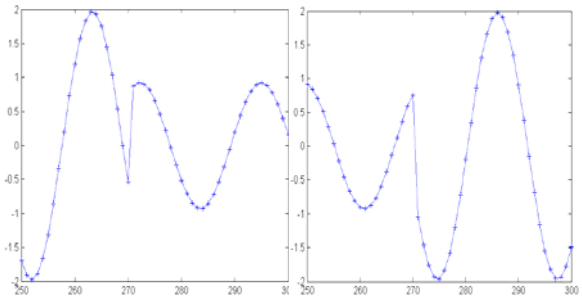


Figura 2: *Clicking* - quebra na amplitude do som no canal direito e esquerdo quando ocorre uma transição entre HRTF

2.2 Eliminação do *clicking*

A solução para eliminar estes artefactos passou por duas fases. Primeiro, sempre que ocorrer um movimento do utilizador, é calculada a convolução da nova HRTF com o último segmento de som usado com a HRTF antiga. De seguida, são substituídas as amostras que sobraram da última convolução calculada com a HRTF antiga pelas amostras calculadas com a nova HRTF (ver Figura 3). Com esta modificação, verifica-se que o ruído deixa de existir no início de cada transição. No entanto, agora é criado uma quebra de amplitude no resultado de saída quando existe a transição entre as HRTF (ver Figura 4a). A segunda fase da solução consistiu na interpolação entre a transição das HRTF para resolver a quebra de amplitude. Efectua-se uma interpolação linear de forma a obter uma transição suave entre os *buffers* com HRTF diferentes efectuando assim um *fade in* e um *fade out* entre os diferentes *buffers* em N amostras.

O processo completo de cálculo do som directo é apresentado na Figura 3, que mostra a forma como o sinal de saída é gerado para os dois primeiros segmentos. O sinal final obtido é apresentado na Figura 4b.

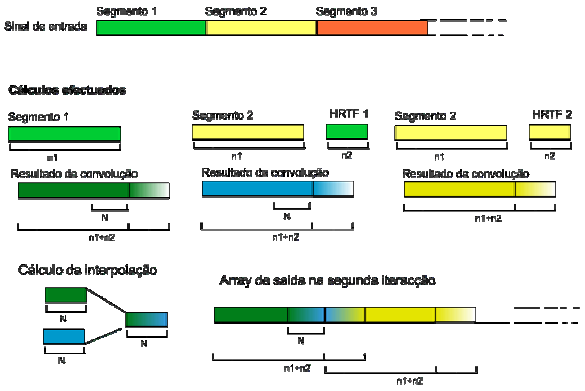


Figura 3 : *Buffer* de saída com a interpolação para remover o *clicking*

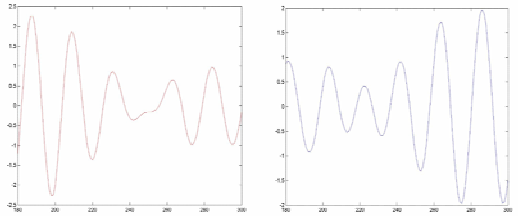


Figura 4: Resultado final antes e depois da interpolação

3. REVERBERAÇÃO INICIAL: 1as REFLEXÕES

As primeiras reflexões provêm são posições diferentes do som directo (devido a reflexões do som nas paredes da sala por exemplo) e apresentam tempos de atraso ligeiramente maiores. Existem diferentes métodos de simulação dessas reflexões. Essa simulação pode ser feita

dessas reflexões. Essa simulação pode ser feita através de métodos geométricos, em que as reflexões são calculadas tendo em conta o modelo da sala, ou através de filtros que tentam recriar a realidade (filtros *tapper delay*)[8], não sendo tão precisos como os métodos geométricos. Dado que o objectivo deste trabalho consiste em simular com o máximo de realismo a acústica de uma sala optou-se por usar um método de cálculo baseado no modelo geométrico, tendo-se desenvolvido uma aplicação que permite associar as características acústicas aos vários materiais constituintes do modelo [9]. É então calculada a posição das primeiras reflexões usando o método das imagens virtuais [10] (Figura 5). O algoritmo pode ser aplicado de forma iterativa para calcular reflexões de ordem superior, mas dado o aumento exponencial do número de reflexões esse cálculo torna-se rapidamente incomportável (Figura 6).

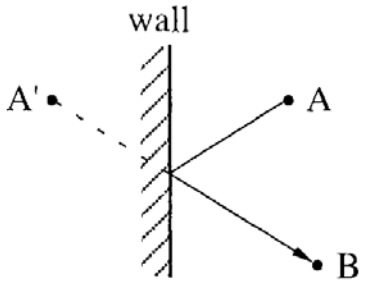


Figura 5: Método das imagens virtuais

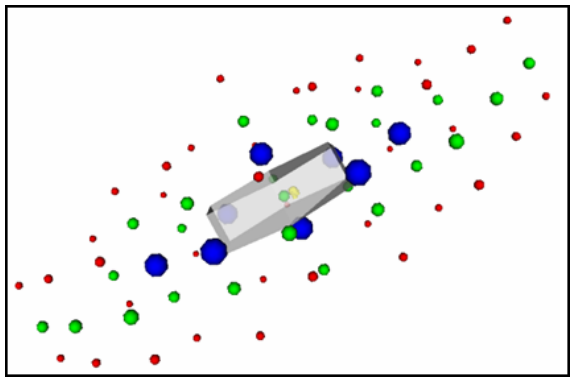


Figura 6: Exemplo de reflexões até a 3ª ordem calculadas com o método das imagens virtuais

Aplicando este método é possível considerar as reflexões como fontes distintas. A única diferença reside no facto dos sinais de saída dessas fontes serem atenuados tendo em conta o coeficiente de absorção dos materiais onde o som é reflectido. Como cada material tem coeficientes de absorção distintos para cada banda de frequências, é necessário filtrar o sinal de entrada em bandas específicas, atenuando cada uma separadamente. Dessa forma cria-se o som que irá ser reproduzido em cada reflexão que foi obtida. A Figura 7 apresenta o algoritmo utilizado.

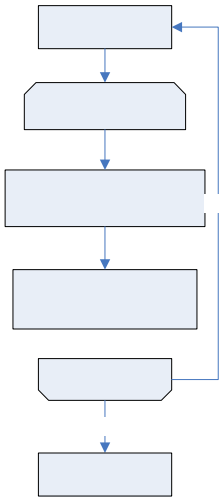


Figura 7: Fluxograma do algoritmo para atenuação do som por bandas

4. REVERBERAÇÂO TARDIA

A maior parte da vida do ser humano é vivida em ambientes reverberantes. Independentemente de o ouvinte estar a assistir a um concerto num auditório ou ao ar livre, o som percepcionado é sempre acompanhado por reflexões do som vindas de diversos pontos. Se essas reflexões ocorrem muito cedo (se o ouvinte se encontrar numa sala pequena, por exemplo) o resultado destas não são percebidas como sons vindos de posições distintas. Em vez disso, as reflexões modificam a percepção do som, alterando o seu timbre, a sua amplitude e as suas características espaciais. As reflexões tardias, muito comuns em ambientes reverberantes, tais como salas de espectáculo e catedrais, produzem um som bastante distinto do original.

A presença de reverberação é um efeito bastante comum na música que sem reverberação torna-se muito seca e perde um pouco da sua vivacidade. Por outro lado, muita reverberação ou o tipo errado de reverberação pode fazer com que uma boa performance musical seja pouco inteligível, perdendo assim o seu encanto.

A reverberação difusa ou tardia é um tópico estudado há vários anos existindo por isso vários artigos e unidades de reverberação diferentes. Os primeiros passos foram dados por *Schroeder* em 1962 com um artigo no qual são descritos os parâmetros necessários para criar reverberação incolor, ou seja que tenha uma resposta em frequência uniforme para todas as frequências[11]. Este artigo descreve algumas das propriedades mais importantes das salas de espectáculo de grande dimensão, tais como a resposta em frequência e a forma como um som impulsional é percepcionado. Encontra-se ainda descrito o filtro *all-pass*, os motivos que levaram à sua implementação e uma unidade de reverberação baseada em filtros *all-pass*.

Segundo *Schroeder*, para se conseguir recriar a reverberação difusa é necessário simular um mínimo de 1000 reflexões. Tentar simular essas reflexões com o método

das imagens virtuais é impossível, devido ao peso computacional envolvido. É então necessário criar algoritmos com um menor custo computacional.

As condições necessárias a atingir em reverberação artificial são [11]:

1- A resposta em frequência tem que ser uniforme para todas as bandas de frequência.
2- A reverberação tem que cobrir a gama de frequência existentes no som a ser reverberado.
3- A quantidade de ecos após curto intervalo de tempo desde o início da reprodução do som tem que ser suficientemente grande para que os sons não sejam perceptíveis individualmente pelo ouvido.
4- Não devem existir 'ecos' periódicos
5- A resposta amplitude – frequência não pode conter nenhuma periodicidade aparente.

Moorer, em 1987, resume e explica as primeiras unidades de reverberação descobertas de *Schroeder* tais como o filtro *comb* e *all-pass*, e sugere combinações criadas com esses filtros para melhorar a resposta audível, tais como os filtros *comb* e *all-pass* oscilatórios, discutindo as vantagens e desvantagens do seu uso. A optimização que *Moorer* propõe a uma das combinações de filtros de *Schroeder* ficou conhecida como reverberador de *Moorer*. Os parâmetros necessários para a sua implementação, tais como o atraso de cada filtro usado na unidade de reverberação e o uso de filtros passa baixo para simular a atenuação do ar, são discutidos no mesmo artigo. Outros dos pontos que foi inicialmente descrito neste artigo e que ainda hoje continua em estudo é a possibilidade da criação de resposta impulsionais sintéticas de salas recorrendo para isso a ruído branco (tendo em conta que a resposta impulsional de uma sala se assemelha a ruído branco com um decaimento exponencial) [13].

Em [14], *Angelo Farina* descreve como gerar reverberação por convolução de respostas impulsionais previamente gravadas com sons anecóicos, apresentando resultados sobre o peso computacional que esta operação exige para diferentes sistemas.

No livro *Applications of Digital Signal Processing to Audio and acoustics*[12] existe um capitulo que descreve com bastante detalhe a evolução dos algoritmos de reverberação existentes desde os filtros *comb* passando pelas *Feedback delay Networks* (FDN) e *Digital waveguide Mesh* (DWM), até aos algoritmos com variações temporais, explicando detalhadamente os passos necessários para a sua implementação.

Mais recente, o livro *Digital Audio Effects*[15] fornece também alguns algoritmos de reverberação dando maior ênfase à FDN apresentando mais detalhes técnicos sobre a sua implementação.

Um dos principais objectivos para o uso das FDN consiste no aumento da densidade de ecos simulados. Em 1997, *Davide Rochesso* descreveu uma matriz de realimentação para FDN baseada na sequência de *Galois* e ordenada numa matriz circular de forma a maximizar a densidade de ecos. Devido à particularidade dessa sequência e da matriz circular é possível optimizar o peso computacional das FDN [16].

Em 2000, *Luke Dahl* e *Jean Marc Jot* descrevem um filtro baseado numa nova implementação do filtro *all-pass*. Esta implementação tem como objectivo aumentar a precisão do tempo de decaimento e o nível de energia dos filtros *all-pass*. Descrevem ainda uma unidade de reverberação baseado num *Unitary* FDN (UFDN) que usa esses filtros[17].

No âmbito deste trabalho todas as unidades de reverberação aqui descritas foram implementadas cronologicamente, desde as mais básicas (filtros *Comb*) até às unidades *Multirate*. O objectivo deste estudo era determinar um algoritmo que simule a reverberação tardia real. Algumas das unidades de reverberação foram implementadas apenas para ter uma noção do resultado produzido, conhecendo-se à partida os defeitos das mesmas. No entanto, optou-se por esta estratégia para assim se ter uma relação qualidade/custo de vários algoritmos e verificar qual o mais indicado para cada ambiente virtual criado.

5. TESTES PERCEPTUAIS

Foram efectuados testes perceptuais preliminares para verificar a capacidade do utilizador determinar a posição espacial das fontes sonoras usando o software criado. Os testes efectuados tiverem como principal objectivo verificar a capacidade de localização do utilizar no plano vertical e horizontal, nas seguintes condições:

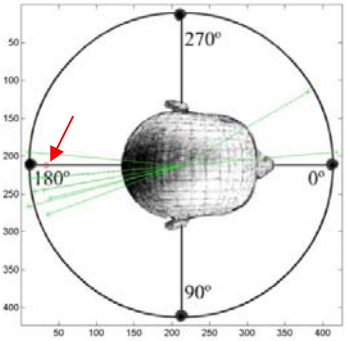
- Apenas som directo
    - Utilizador estático
    - Utilizador dinâmico (pode mudar a orientação da cabeça sendo o som actualizado de acordo)
- Com primeiras reflexões
    - Utilizador estático
    - Utilizador dinâmico

A experiência foi feita com um grupo relativamente reduzido de 11 pessoas. No entanto planeia-se alargar a experiência a mais utilizadores assim como implementar testes adicionais para validar também os vários algoritmos de reverberação implementados. Na Figura 8 encontra-se uma captura do ambiente usado com as primeiras reflexões activas e o utilizador estático (movimentos da cabeça não provocavam alterações no som). As primeiras reflexões foram calculadas tendo em conta uma sala hexagonal (Figura 6). Os objectos coloridos que se encontram na figura são usados para permitir ao utilizador indicar qual a posição que considera correcta para a fonte sonora. Nos testes em que o utilizador estava estático perante o mundo não foi dada qualquer informação visual ao utilizador.

Figura 8: Teste perceptual efectuado no plano horizontal (primeiras reflexões activas)

Dado o número reduzido de utilizadores que colaboraram nestes testes preliminares, os resultados obtidos não podem ser muito conclusivos. No entanto, a com base numa análise global é de realçar que quando é dada a possibilidade de interacção com o ambiente (utilizado dinâmico), a sua capacidade de localização melhora de uma forma significativa quando analisado o plano horizontal. No plano vertical, os resultados são inconclusivos, sendo difícil em qualquer das situações ao utilizador determinar a origem da fonte sonora.

Na Figura 9 são apresentados exemplos dos resultados obtidos durante os testes perceptuais para o som directo. O ponto vermelho em cada imagem indica a posição correcta das fontes e os raios verdes os dados recolhidos durante a mesma. Verifica-se que, nos testes do utilizador dinâmico, existe um erro bastante menor nos dados obtidos. No entanto, para analisar os dados do utilizador estático é necessário ter em conta o problema da "confusão frente-trás", ou seja, o utilizador só tem capacidade de determinar se o som está à sua frente ou atrás de si se puder interagir com o mundo, devido a simetria das HRTF utilizadas.



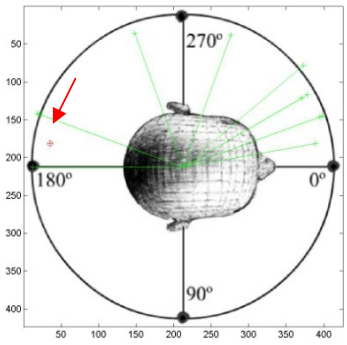


Figura 9 : Resultado dos testes para um utilizador estático (em cima) e com possibilidade de movimento da cabeça (em baixo)

5.1 Conclusões

Neste artigo são apresentados os vários algoritmos desenvolvidos para permitir uma auralização em tempo real numa ambiente de realidade virtual imersivo.

Foi implementado um algoritmo para determinar o som directo e as primeiras reflexões e para atenuar cada reflexão tendo em conta os coeficientes de absorção do material reflector. Este algoritmo tem um peso computacional elevado. No entanto, dado que os cálculos são efectuados *off-line* antes do início da interacção, não este problema não é muito significativo para a posterior utilização em tempo real. Tem ainda a vantagem de ser possível guardar os resultados obtidos, caso os mesmos tenham que ser usados em diferentes ocasiões.

Foram ainda implementados vários algoritmos para simular a reverberação tardia. Através da criação da resposta impulsional sintética é possível obter um som reverberante bastante natural, sendo um dos algoritmos a ter em consideração, caso o resultado não necessite de ser em tempo real. Os algoritmos para criar reverberação em tempo real que convém realçar são as FDN e o algoritmo multirate. Através da FDN é possível simular um tempo de reverberação médio bastante preciso em que o resultado é bastante natural com a vantagem de ter um peso computacional relativamente pequeno. Contudo, falta ainda testar e validar os algoritmos mais promissores para permitir uma reverberação configurável por banda e integrar este algoritmo com os algoritmos de som directo e reflexões de forma a ter um sistema de auralização completo.

Os resultados de testes perceptuais preliminares permitiram evidenciar a importância de poder interagir com o ambiente sonoro permitindo uma melhor localização do som. Estão planeados, para o futuro, testes adicionais para avaliar e validar os vários algoritmos de reflexão implementados.

6. REFERÊNCIAS

[1] Durand Begault, “Perceptual Effects of Synthetic Reverberation on Three-Dimensional Audio Systems,” J. Audio Eng. Soc., vol. 40, no. 11, pp. 895-904, 1992.

[2] Durand Begault (1994) *‘3-D Sound for virtual reality and multimedia’*. Academic Press

[3] Frederic A. White, *‘Our acoustic Environment’* John Wiley & Sons

[4] Gary S. Kendall (1995) *‘A 3-D Sound Primer: Directional Hearing and Stereo Reproduction’*, Computer Music Journal, 19(4), 23-46.

[5] William G. Gardner (1999), *’3D Audio and Acoustic Environment Modeling’*, Wave Arts, Inc.

[6] F. Alton Everest (2001), ‘*The master handbook of acoustics*’,McGraw-Hill

[7] WWW ‘*HRTF Measurements of a KEMAR Dummy-Head Microphone* ‘. URL (Junho 2008): http://sound.media.mit.edu/KEMAR.html

[8] Young-cheol Park, Taik-sung Chio, Jae-woong Jung Dae-hee Youn, Siwook Nam, and Jungmin Song, Low Complexity 3D Audio Algorithms for Handheld Devices, AES, international conference Seoul, Korea, 2006

[9] Dias P., Campos G., Santos V., Casaleiro R., Seco R., Sousa Santos B. "3D Reconstruction and Auralization of the “Painted Dolmen” of Antelas". In Proceedings of the Electronic Imaging 2008 conference, SPIE Vol. 6805, 6805OY, Three-Dimensional Image Capture and Applications 2008, San Jose, California, USA. 28-29 Janeiro 2008.

[10] J. B. Allen and D. A. Berkl_ey, "Image Method for Efficiently Simulating Small-Room Acoustics," J. Acoust. Soc. Am., vol. 65, pp. 943-950 (1979)

[11] *M.R.Schroeder, Colorless Artificial Reverberation, J. Audio Eng. Soc.*, vol. 9, pp. 192-197; July, 1961

[12] W.G. Gardner Chapter 3. Reverberation Algorithms, in Kahrs, M. and Brandenburg, K. Editors. Applications of Digital Signal Processing to Audio and Acoustics. Kluwer Academic Publishers. 1998

[13] J. A. Moorer, "About This Reverberation Business," Foundation of Computer Music, The MIT press, pp.605-639, 1987.

[14] Farina, A.- “*Convolution of anechoic music with binaural impulse response*” - Proc. of PARMA-CM Users Meeting, Parma, 1-2 luglio 1993

[15] Udo Zölzer, *DAFX: Digital Audio Effects*, John Wiley & Sons, Inc., 2002

[16] Davide Rochesso, *Maximmaly diffusive, Yet Efficient Feedback Delay Networks for artificial reverberation*, IEEE signal processing letters, vol. 4, nº 9, September 1997

[17] Luke Dahl, Jean Marc Jot, *A reverberator based on absorbent all-pass filters*, COST G-6 Conference on Digital Audio Effects (DAFX-00), Verona, Italy, December 7-9, 2000