

Técnica rápida para geração de aurilizações utilizando redes neurais artificiais

Taminato, F.O.¹, Tenenbaum, R.A.² e Melo V.S.G.³

¹Laboratório de Instrumentação em Dinâmica, Acústica e Vibrações – LIDAV, Programa de Pós-Graduação em Modelagem Computacional, Universidade do Estado do Rio de Janeiro, Nova Friburgo, RJ, filipeot@gmail.com

²Programa de Pós-Graduação em Engenharia Civil, Universidade Federal de Santa Maria, Santa Maria, RS, ratenenbaum@gmail.com

³Engenharia Acústica, Programa de Pós-Graduação em Engenharia Civil, Universidade Federal de Santa Maria, Santa Maria, RS, viviane.melo@eac.ufsm.br

Resumo

Um dos objetivos do desenvolvimento e do aperfeiçoamento de técnicas numéricas em sistemas de geração de realidade virtual acústica e produção de aurilizações fidedignas, consiste na redução do custo computacional e, simultaneamente, garantir a qualidade sonora da simulação. Neste artigo, é apresentada uma nova técnica para modelar as funções de transferência associadas à cabeça humana, necessárias ao cômputo das respostas impulsivas biauriculares. São utilizadas redes neurais artificiais do tipo funções de base radial. Um conjunto dessas redes é treinado e testado de modo a cobrir todo o espaço auditivo no entorno da cabeça. Cada rede neural para uma dada direção tem como entrada o espectro da frente de onda sonora que atinge o receptor e, como saída, a resposta impulsiva associada à cabeça humana filtrada, para a direção correspondente. Desse modo, trabalha-se diretamente no domínio do tempo, contornando-se a necessidade de convoluções com uma redução do custo computacional em cerca de 90%. A técnica proposta é comparada com o método das convoluções, tanto no domínio do tempo como no domínio da frequência. Os resultados simulados demonstram a eficiência da técnica proposta, com valores de correlação bem próximos de um. Para validar o resultado, testes preliminares utilizando índices de articulação para comparar a percepção da fala em uma sala real e na mesma sala simulada computacionalmente utilizando a modelagem das respostas impulsivas filtradas aqui descrita foram conduzidos, com resultados plenamente satisfatórios.

Palavras-chave: realidade virtual acústica, aurilização em salas, geração de BRIRs, redes neurais artificiais, simulação de campo acústico em salas, índice de articulação.

PACS: 43.55.-n, 43.55.Br, 43.55.Ka.

Fast technique for auralization generation using artificial neural networks

Abstract

The main goal in development of numerical techniques in acoustic virtual reality systems and production of reliable auralizations is to reduce the computational cost and, at the same time, to guarantee the sound quality. In this paper, a new technique for modeling head-related transfer functions are presented. Artificial neural networks of the radial basis functions type are used. A set of these networks is trained and tested to cover the entire auditory space around the head. Each neural network for a given direction has as input the spectrum of the sound ray that reaches the receiver and, as output, the filtered head-related impulse response, for the corresponding direction working directly in the time domain and circumventing the need for convolutions with a computational cost reduction of 90%. The proposed technique is compared with the convolution method in both the time and frequency domain. The results show the efficiency of the proposed technique, with correlation values very close to one. To validate the result, preliminary tests using articulation indices to compare speech intelligibility in an actual and virtual room were conducted, with fully satisfactory results.

Keywords: acoustic virtual reality, room auralization, BRIRs generation, artificial neural networks, room acoustic simulation, articulation index.

1. INTRODUÇÃO

A técnica conhecida como Realidade Virtual Acústica (RVA) tem como objetivo, recriar ambientes sonoros a partir de dados medidos, sintetizados ou simulados [1]. Seu leque de aplicações é bem vasto, varrendo desde o controle de ruído em plantas industriais, a avaliação subjetiva de isolamento acústico entre recintos, o auxílio ao projeto acústico de salas para múltiplos usos, a melhoria da inteligibilidade da fala em salas de aula, dentre muitos outros. O resultado mais proeminente em RVA consiste na aurilização de salas que nada mais é do que a geração de sinais sonoros a serem escutados, geralmente por meio de fones de ouvido, que permitem criar a sensação auditiva de imersão no ambiente que foi simulado [2]. Isso é muito útil, por exemplo, para projetar determinado ambiente, de forma a que se possa prever o comportamento acústico da sala, antes de sua construção. Na correção acústica de ambientes existentes, a aurilização é uma ferramenta de grande aplicabilidade para verificar se as soluções acústicas adotadas resultarão no campo sonoro desejado.

Para obter uma aurilização a partir de dados simulados é necessário, naturalmente, dispor de um simulador de acústica de salas que forneça de forma adequada e confiável, para um dado par fonte-receptor, as respostas impulsivas biauriculares da sala (BRIRs – da sigla em inglês, *Binaural Room Impulse Responses*), núcleo do procedimento de aurilização [3]. Para tal fim, o simulador necessita modelar as fontes, os receptores, as superfícies de contorno da sala, com suas características acústicas (coeficientes de absorção e de espalhamento) e os diversos fenômenos intervenientes na propagação de ondas acústicas em recintos [4,5]. Cumpre notar que o receptor a ser modelado não constitui um microfone onidirecional, mas sim a cabeça de uma pessoa. A modelagem desse receptor é feita por meio de funções direcionais conhecidas como respostas impulsivas associadas à cabeça humana (HRIRs – da sigla

em inglês, *Head Related Impulse Responses*) ou das funções de transferência associadas à cabeça (HRTFs – da sigla em inglês, *Head Related Transfer Functions*), estas últimas dadas no domínio da frequência. São essas funções que conferem as características de filtragem em frequência resultante da inserção da cabeça humana no campo sonoro, as principais responsáveis pela sensação de espacialidade da audição humana [3], sendo fortemente dependentes da direção de chegada das ondas sonoras. Cumpre notar, ainda, que essas funções diferem um pouco entre pessoas, pois cada indivíduo possui dimensões antropométricas levemente diferenciadas, particularmente no que toca ao formato da cabeça, da parte superior do torso e do pavilhão auricular.

As BRIRs são, então, obtidas a partir das informações das frentes de onda – que, doravante, denominaremos como *raios acústicos* ou, simplesmente, *raios* –, obtidas a partir da simulação acústica da sala, e das HRIRs. Por fim, na última etapa de um sistema de aurilização, é feita a convolução de um sinal monofônico e anecoico com as BRIRs. A saída obtida é, então, reproduzida, utilizando-se fones de ouvido de boa qualidade, preferencialmente equalizados.

Diante do exposto, dois elementos essenciais para gerar aurilização são a simulação acústica da sala e a modelagem das HRTFs. Para a simulação acústica da sala são utilizados os métodos baseados na teoria ondulatória, na acústica geométrica ou na acústica estatística. A teoria ondulatória permite o estudo dos modos de baixa frequência da sala e tem sido negligenciada pelos desenvolvedores de programas de simulação em salas [6]. A principal razão é que somente geometrias retangulares possuem solução analítica fechada e todas as demais demandam a utilização de métodos numéricos, tais como o método dos elementos finitos ou o método dos elementos de contorno, com um alto custo computacional. Já os métodos da acústica geométrica são os mais utilizados em

simulação de salas [7]. Os métodos de traçado de raios [8–10], traçado de cones ou pirâmides [11] e o método das imagens ou fontes virtuais [12] têm sido os mais utilizados nessa linha, com bons resultados. Contudo, em se tratando de aurilização, as reflexões difusas desempenham um papel fundamental no preenchimento da cauda reverberante [13] e necessitam ser cuidadosamente modeladas. Um método que tem se mostrado acurado nessa linha é o método dos caminhos aleatórios, também conhecido como da radiosidade [14,15], a despeito de seu custo computacional ser mais elevado do que os métodos da acústica geométrica.

Quanto às HRIRs/HRTFs há diversas maneiras de caracterizá-las. A primeira delas é, naturalmente, através de medições. Desde a década de 1990 vêm sendo publicados bancos de dados de HRIRs com diferentes refinamentos. O banco de dados do manequim KEMAR [16], por exemplo, apresenta dados de medição para 720 direções, para cada ouvido. O mais recente e completo banco de dados publicado apresenta medições para 64.442 direções para cada ouvido do manequim FABIAN [17]. Isso corresponde a uma HRIR de grau em grau, tanto no azimute quanto na elevação, resultando em uma acurácia superior ao limiar do observável, ou (JND – da sigla em inglês, *Just Noticeable Difference*) [18].

Há, também, diversas técnicas de modelagem das HRIRs/HRTFs [19–22]. A questão básica na modelagem dessas funções fundamenta-se no binômio eficácia versus fidedignidade. Em outros termos, trata-se de produzir modelos que permitam uma maior eficiência computacional na geração das BRIRs, mantendo, sobretudo, a fidedignidade das aurilizações.

Cumprir notar que o cálculo das BRIRs envolve não só a modelagem das HRIRs (ou HRTFs), mas, também, em uma etapa subsequente, o cômputo das HRIRs filtradas. Isso corresponde ao produto de convolução entre o raio que chega ao receptor e a HRIR da

direção correspondente. O produto de convolução torna o custo computacional dessa etapa de geração de aurilizações extremamente custoso, devido à grande quantidade de raios sonoros que chegam no receptor e a consequente necessidade de realização de numerosas convoluções.

Diante do exposto, este artigo tem como objetivo principal apresentar uma modelagem que permita recuperar as HRIRs filtradas utilizando redes neurais artificiais (RNAs). O uso das RNAs traz vantagem em relação ao método tradicional de geração das BRIRs que utiliza convoluções, por apresentar um menor custo computacional. Para tal, foi desenvolvido um conjunto de redes neurais artificiais do tipo funções de base radial (RBF – da sigla em inglês, *Radial Basis Functions*) [23,24] especializadas na geração de HRIRs filtradas. Cada rede está associada a uma direção de chegada e possui como entrada os componentes do espectro do raio acústico, em bandas de oitava, e como saída a HRIR filtrada, tal como descrito na Seção 4. O núcleo do procedimento de aurilização é o simulador de acústica de salas. Este deve ser alimentado com as informações sobre a geometria da sala que, por sua vez, deve conter os dados dos coeficientes de absorção e espalhamento das superfícies de acabamento, informações sobre as fontes sonoras, incluindo seus dados de direcionalidade, e informações sobre os receptores, que devem incluir o banco de dados das HRIRs/HRTFs da cabeça receptora. A saída do simulador – no que toca à geração de aurilizações – são as BRIRs, para cada par fonte-receptor previamente selecionado. Finalmente, escolhendo-se um dado sinal anecoico, sua convolução com as BRIRs fornecerá o sinal sonoro aurilizado desejado, também para cada par fonte-receptor. A Figura 1 ilustra o esquema geral para a geração de uma aurilização.

Na Seção 2, discute-se o método tradicional de geração das BRIRs; na Seção 3 é feita uma brevíssima apresentação do simulador de acústica de salas utilizado; a Seção 4 descreve sucintamente o modelo das redes neurais

artificiais; a Seção 5 apresenta o modelo para geração das BRIRs aqui proposto. Na Seção 6 são apresentados os principais resultados, comparando-se as HRIRs/HRTFs filtradas calculadas pelo método da convolução e pelo método das redes neurais artificiais, bem como ensaios preliminares de validação via testes de articulação. As principais conclusões do trabalho são discutidas na Seção 7.

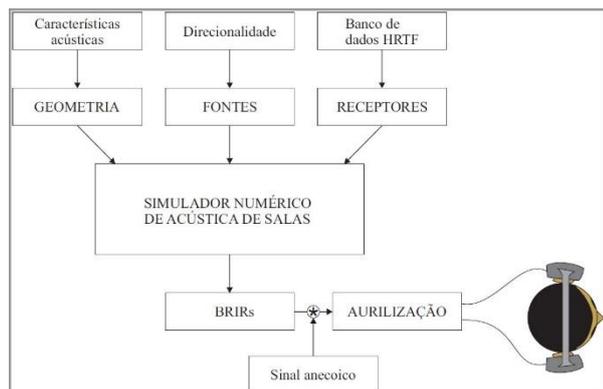


Figura 1: Esquema geral para gerar um som aurilizado.

2. SOBRE O MÉTODO TRADICIONAL DE GERAÇÃO DAS BRIRs

Para o procedimento de geração das respostas impulsivas biauriculares, $b_E(t)$ e $b_D(t)$, onde E e D referem-se aos ouvidos esquerdo e direito, respectivamente, e t é o tempo, são necessárias as seguintes informações: a) para cada raio acústico que chega ao receptor, seu espectro de energia (E_n) por banda de oitava (63 Hz – 16 kHz), o tempo de chegada (τ_n) e os ângulos de chegada (φ_n – azimute e θ_n – elevação), sendo o índice n um contador dos raios; b) as HRIRs, dos dois ouvidos, $h_E(\theta_n, \varphi_n, t)$ e $h_D(\theta_n, \varphi_n, t)$ da cabeça que está sendo simulada.

A Figura 2 ilustra o procedimento para a geração das BRIRs, segundo o método das convoluções.

Observe-se que o simulador de salas gera um conjunto de raios acústicos que atingem o receptor com três informações: o tempo de chegada; o espectro de energia; e a direção de chegada, para cada raio. Esta última informação seleciona a HRTF do banco de dados da direção mais próxima. Em seguida, é

efetuado o produto complexo entre o espectro em bandas de oitava do raio e a HRTF selecionada (dois canais, um para cada ouvido). Para efetuar esse produto, na técnica clássica, cada uma das nove bandas de oitava em que o espectro do raio é dividido é multiplicada pela média espectral, na banda correspondente, da HRTF em questão.

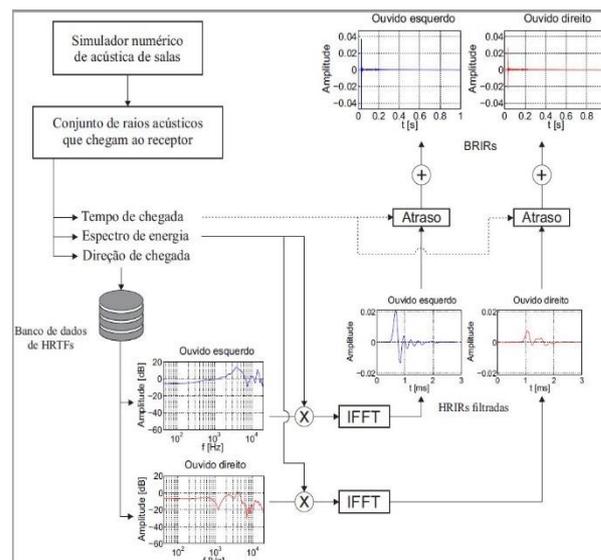


Figura 2: Panorama geral do cômputo de uma BRIR utilizando o método das convoluções.

Na sequência, aplica-se a transformada inversa rápida de Fourier (IFFT), para se obter as HRIRs filtradas para os ouvidos esquerdo e direito. Estas, por sua vez, são atrasadas segundo o tempo de chegada de cada raio e, finalmente, adicionadas para a obtenção da BRIR. O produto complexo acima mencionado, seguido da aplicação da IFFT nos dois canais, equivale à convolução do raio acústico com a HRIR da direção correspondente e, por essa razão, esse procedimento é chamado de *método da convolução*.

3. O CÓDIGO RAIOS 7

O código computacional para simulação de acústica de salas RAIOS (*Room Acoustics Integrated and Optimized Software*) vem sendo desenvolvido desde o final da década de 1990 pelos autores, estando atualmente em sua Versão 7. O código opera um algoritmo híbrido que utiliza o método de traçado de

raios (*ray-tracing*) para o cálculo das reflexões especulares, incluindo a absorção nas superfícies e a atenuação do ar, com o método das matrizes de transição – também conhecido como método da radiosidade – para o cálculo das reflexões difusas [5].

A Figura 3 ilustra o método das matrizes de transição, que não é usual em programas de simulação em salas.

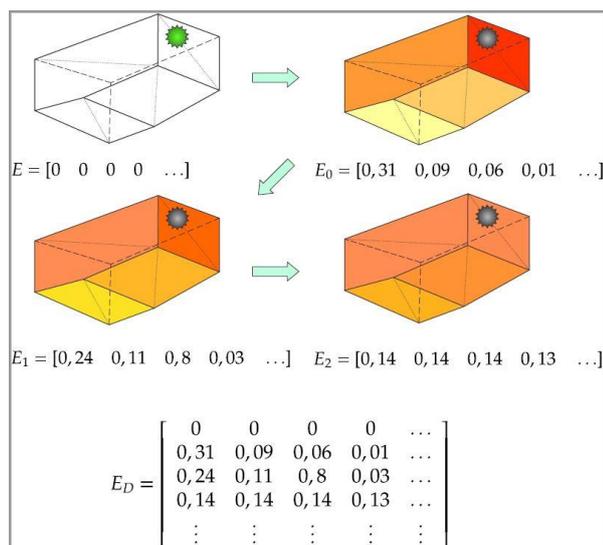


Figura 3: Esquema ilustrativo do método de cálculo das reflexões difusas por meio de matrizes de transição.

No instante inicial, a energia sonora está toda contida na fonte; passado um intervalo de tempo, essa energia é distribuída entre as diversas superfícies segundo o ângulo sólido com que cada uma delas é vista pela fonte (a parede mais próxima à fonte fica mais energizada, como mostra a Figura 3). Na sequência, todas as superfícies emitem para todas as demais, sempre segundo os ângulos sólidos relativos, de modo que, aos poucos, a energia sonora vai se distribuindo mais uniformemente pela sala. O procedimento é descrito matematicamente e numericamente por meio de matrizes – cujas colunas representam as diversas superfícies em que o contorno da sala é subdividido e cujas linhas representam o tempo discretizado – que vão se atualizando ao longo do processo de difusão. O resultado final é uma cauda reverberante mais densa, ou melhor preenchida, e mais próxima daquela que é observada experimentalmente.

O programa RAIOS participou da comparação internacional de programas de simulação de acústica de salas, *Round Robin 3* (RR3) [6,25], tendo obtido excelentes resultados na média dos diversos parâmetros de qualidade acústica simulados, confrontados com os mesmos parâmetros medidos, como ilustra a Figura 4. Contudo, naquela ocasião (2002), não se avaliou a capacitação dos programas participantes para gerar aurilizações.

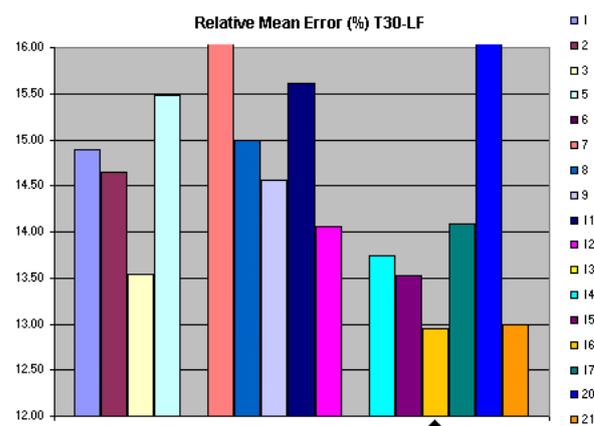


Figura 4: Média dos erros médios relativos para diversos parâmetros de qualidade acústica, com respeito aos medidos, entre os participantes do RR3. O RAIOS (em sua Versão 3) é o software de número 16, indicado com a seta.

4. DESCRIÇÃO DO MODELO DAS RNAs

Redes neurais artificiais (RNAs) constituem a base daquilo que hoje é conhecido como inteligência artificial ou inteligência computacional [26]. RNAs são muito utilizadas em engenharia para resolver problemas de natureza não linear, inclusive na área de acústica [27]. Além disso, RNAs são também muito úteis para o reconhecimento de padrões. E é justamente essa característica que pode ser largamente explorada para modelar as respostas impulsivas associadas à cabeça humana.

O tema das RNAs é por demais vasto para ser analisado aqui. Para maiores informações remete-se a [28] e [29]. Todavia, algumas informações básicas são discutidas a seguir. Primeiramente, para uma RNA aprender um determinado padrão de funções, é necessário o seu *treinamento* ou *procedimento de*

aprendizagem, que pode ser ou não supervisionado. A Figura 5 ilustra o esquema de aprendizagem supervisionado que foi utilizado neste trabalho. A rede é alimentada com um padrão de entrada e a saída corrente é comparada com a saída desejada. A partir do erro encontrado, os pesos da rede são ajustados segundo um dado algoritmo de minimização, até que um determinado critério de parada seja alcançado.

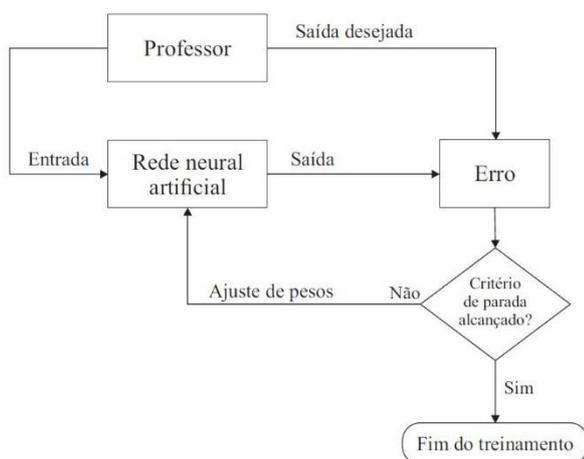


Figura 5: Esquema de treinamento supervisionado para uma rede neural artificial.

Uma vez treinada, a rede necessita ser testada, ou seja, é preciso verificar se ela aprendeu o padrão adequadamente. Redes superdimensionadas podem gerar os erros de *overfitting*, enquanto redes subdimensionadas podem causar *underfitting*. No primeiro caso, a rede memoriza os padrões de tal forma que perde a capacidade de generalização; no segundo caso, a rede não consegue aprender os padrões de forma satisfatória. Os padrões utilizados no teste das redes necessariamente deverão ser diferenciados daqueles utilizados no treinamento. O banco de dados de HRIRs utilizado para o treinamento das RNAs foi o de [17] e um conjunto de 64.442 redes neurais foram treinadas e testadas, uma para cada direção, com discriminação de um grau.

Dentre os diversos tipos de redes disponíveis para as mais variadas aplicações, adotamos, por sua maior facilidade de treinamento e, principalmente, por seu baixo custo computacional na fase de execução, as redes

conhecidas como funções de base radial (RBF – da sigla em inglês, *Radial Basis Functions*). As redes do tipo RBF possuem apenas uma camada intermediária. Constituem redes não retroalimentadas, ou progressivas, também chamadas de redes não recorrentes.

A Figura 6 ilustra a arquitetura da rede RBF utilizada neste trabalho. Observe-se que há apenas uma camada intermediária, onde cada neurônio φ_j , $j = 1, 2, \dots, k$ recebe informação dos nós de entrada e as processa para os neurônios de saída.

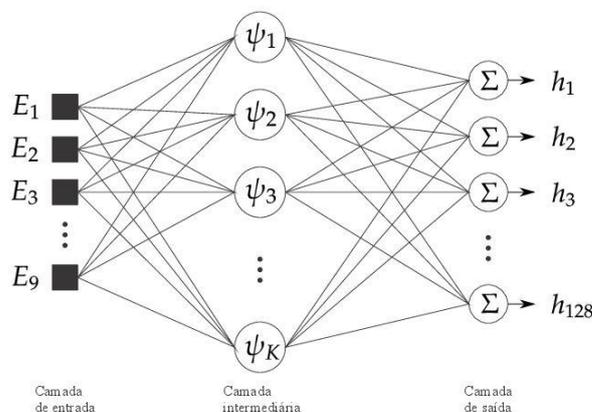


Figura 6: Arquitetura da rede neural do tipo RBF utilizada neste trabalho.

O número de nós na entrada da rede adotada é nove, correspondendo à energia do raio que chega ao receptor filtrada em nove bandas de oitava (63 Hz a 16 kHz). A camada de saída possui 128 neurônios, que correspondem à amplitude da HRIR filtrada discretizada em 128 amostras por canal. A camada intermediária possui k neurônios, onde k é um número a ser investigado. Os melhores resultados em termos de custo-benefício foram obtidos para $k = 5$.

5. SOBRE O MÉTODO DE GERAÇÃO DAS BRIRs UTILIZANDO RNAs

Apresenta-se nesta seção um novo método de geração das BRIRs utilizando-se RNAs do tipo RBF, com a arquitetura apresentada na Figura 6. A ideia central desse método é modelar as HRIRs, diretamente no domínio do tempo, contornando a necessidade de

convoluções e, conseqüentemente, minimizando o custo computacional. A Figura 7 ilustra o procedimento geral de geração das BRIRs pelo método das RNAs, aqui proposto.

A parte inicial é idêntica, ou seja, a simulação acústica da sala fornece o conjunto de raios que atingem o receptor com as informações de tempo, espectro e direção. No lugar do banco de dados das HRTFs, utiliza-se o banco de dados de RNAs, descrito na Seção 4. Com a informação da direção é selecionada a rede RBF previamente treinada para aquela direção.

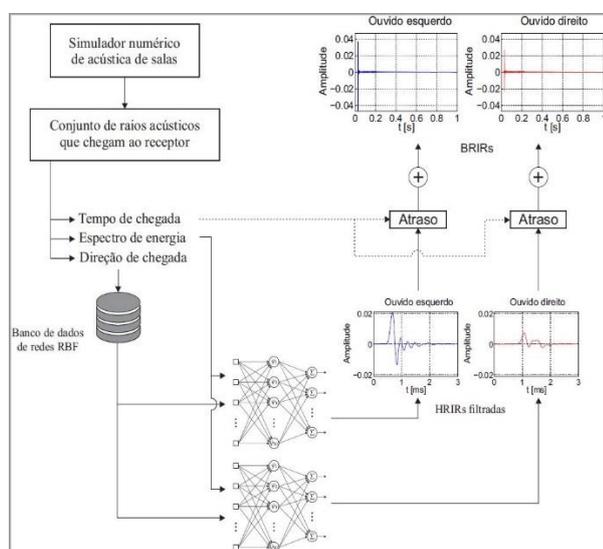


Figura 7: Panorama geral do cômputo de uma BRIR utilizando o método das redes neurais artificiais.

A informação do espectro do raio, então, passa a ser o dado de entrada (nove bandas) da rede, cuja saída será a HRIR filtrada (dois canais), com 128 amostras em cada canal. O procedimento a seguir, como pode ser observado na Figura 7, é o mesmo, ou seja, as HRIRs filtradas são atrasadas segundo o tempo de chegada do raio ao receptor e sucessivamente adicionadas para compor a resposta impulsiva biauricular.

Por trabalhar diretamente no domínio do tempo e evitar o procedimento de convolução, o ganho computacional é de aproximadamente 90%, ou seja, o cálculo da BRIR é efetuado em cerca de 1/10 do tempo daquele despendido pelo método das convoluções [30].

É importante observar que, na simulação computacional de acústica de salas, há dois gargalos fundamentais, ou seja, há dois blocos de cálculo que são computacionalmente mais custosos. O primeiro, diz respeito à simulação da sala para gerar o conjunto de raios acústicos que chegam a cada receptor. Para se ter um resultado mais preciso, principalmente se a sala for grande e de geometria complexa, pode ser necessário, no cômputo das reflexões especulares via traçado de raios, um número substancial de raios emitido pela fonte, podendo ser este da ordem de $10^5 \sim 10^6$. Além disso, o cálculo das reflexões difusas, envolvendo o fenômeno do espalhamento – que no código computacional RAIOS 7 é efetivado pelas matrizes de transição – também tem um custo computacional alto, principalmente quando geometrias complexas exigem um grande número de superfícies planas para sua correta modelagem. No algoritmo híbrido do código RAIOS 7, as reflexões difusas também são convertidas em raios sonoros e, portanto, compõem o conjunto do segundo bloco das Figuras 2 e 7. Nesse primeiro gargalo, portanto, não há participação das redes neurais artificiais.

É no segundo gargalo que as redes promovem uma economia computacional substancial: no cálculo das BRIRs. A etapa da geração das respostas impulsivas biauriculares para cada par fonte-receptor, principalmente para diversos pares, costuma ter um custo computacional equivalente ao da simulação da sala. Nessa etapa, portanto, o modelo das RNAs mostra-se extremamente conveniente.

Se, digamos, para uma sala de geometria medianamente complexa, p.e., o Cenário 10 utilizado no *Round-Robin 4* [31], o cálculo da simulação da sala, gerando todos os raios, especulares e difusos, que atingem cada receptor e o cálculo das respostas impulsivas monoauriculares para, digamos, dez pares fonte-receptor pode levar a um tempo de cálculo de cerca de três horas em um *laptop*. O cômputo das respostas impulsivas biauriculares para os dez receptores – pelo método clássico das convoluções – pode

chegar a um tempo de processamento equivalente. Contudo, como será visto nas próximas seções, utilizando-se o método das RNAs, esse tempo de cálculo cai para menos de vinte minutos.

6. RESULTADOS

Nesta seção são apresentados resultados comparativos entre as HRIRs filtradas (e suas correspondentes no domínio da frequência) calculadas pelo método tradicional das convoluções (MC), tal como ilustrado na Figura 2 e pelo método das redes neurais artificiais (MRN), tal como ilustrado na Figura 7.

São também apresentados resultados comparativos entre índices de articulação reais e virtuais, estes gerados a partir de simulações de uma sala com o código computacional RAIOS 7, utilizando o método das RNAs, para gerar as BRIRs.

Como já ficou demonstrado em trabalhos anteriores [32,33], índices de articulação constituem uma métrica confiável para avaliar aurilizações. Nesse caso, a ideia é comparar índices de articulação medidos em uma sala real com os índices obtidos a partir de aurilizações – avaliadas com fones de ouvido – da mesma sala simulada pelo código computacional RAIOS 7 com a modelagem das HRIRs filtradas com redes neurais artificiais.

6.1. HRIRs e HRTFs filtradas

Como para cada uma das 64.442 direções há uma RNA do tipo RBF dedicada, seria inviável apresentar os resultados comparativos para todas as direções. Foram escolhidas para serem mostradas aqui apenas quatro das direções, representativas do entorno da cabeça humana, cada uma delas concernente a um dos quadrantes em que se pode particionar a esfera. Dentro de cada quadrante a escolha da direção foi feita de forma aleatória. Cada direção é definida por um ângulo de azimute, φ , e de elevação, θ , no padrão usualmente utilizado

para as coordenadas esféricas no entorno da cabeça [1].

Embora a comparação entre as HRIRs filtradas para as direções selecionadas já daria uma boa ideia de como os dois métodos de cálculo são equivalentes em termos de acurácia, apresentam-se também as HRTFs filtradas (complexas) correspondentes, tanto em magnitude como em fase.

De modo a com mais precisão avaliar a diferença entre as curvas geradas pelos dois métodos, uma vez que a mera inspeção visual não evidencia as diferenças, adota-se o coeficiente de correlação cruzada normalizado (NCCC – da sigla em inglês, *Normalized Cross Correlation Coefficient*) [34,35].

O NCCC indica a similaridade entre duas determinadas funções, sendo definido por

$$r = \frac{\sum_{j=1}^N u_j(p)y_j(p)}{\sqrt{\left\{ \sum_{j=1}^N [u_j(p)]^2 \right\} \left\{ \sum_{j=1}^N [y_j(p)]^2 \right\}}}, \quad (1)$$

em que u_j e y_j são as duas funções cuja similaridade está sendo comparada e N é o número de pontos de discretização das funções. No nosso caso, $N = 128$. Quanto mais similares forem as duas funções, sendo comparadas, mais próximo de um será o coeficiente r .

As Figuras 8 a 11 apresentam os resultados comparativos entre as HRIRs filtradas (domínio do tempo) e as HRTFs filtradas correspondentes, em magnitude e fase (domínio da frequência) para quatro direções escolhidas, calculadas tanto pelo método clássico da convolução (MC) como pelo método aqui proposto das redes neurais artificiais (MRN). Todavia, essas comparações foram efetuadas para todas as 64.442 direções, varrendo o amplo intervalo ($0 < \varphi < 360$), ($-90 < \theta < 90$), e obteve-se valores de coeficientes de correlação cruzada

normalizado variando entre 0,9985 e 1,000 [36].

O que se evidencia como caráter geral, observando a Figura 8, é que as três curvas praticamente superpõem-se duas a duas, ou seja, não há diferença observável, tanto na função temporal como nas frequenciais, entre os resultados obtidos pelos dois métodos. O valor calculado do NCCC, $r \approx 1$, confirma esse fato.

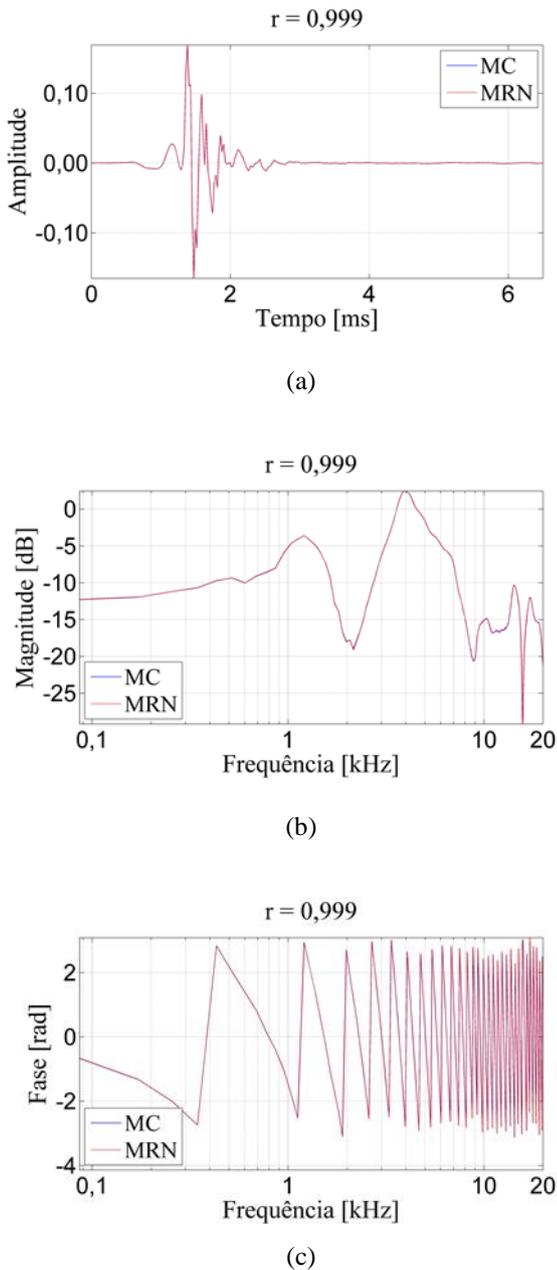


Figura 8: Comparação entre a HRIR/HRTF filtrada, calculada pelos dois métodos (MC e MRN), para a direção $\varphi = 7^\circ$ e $\theta = 23^\circ$. (a) Amplitude; (b) magnitude; e (c) fase.

Como se pode observar das Figuras 8 a 11, tanto no domínio do tempo quanto no da frequência (amplitude e fase) a similaridade entre as curvas é muito alta, com o coeficiente de correlação cruzada normalizado indicando o valor $r = 0,999$ para todos os gráficos.

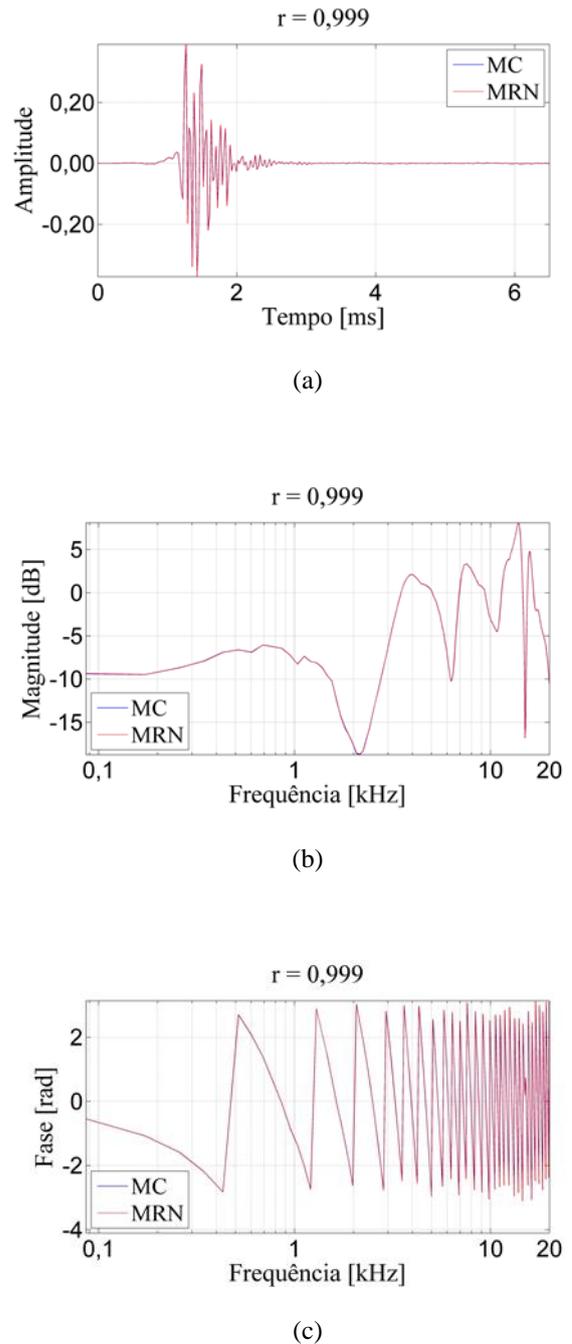
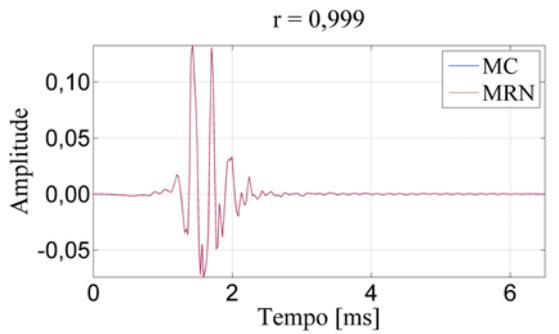
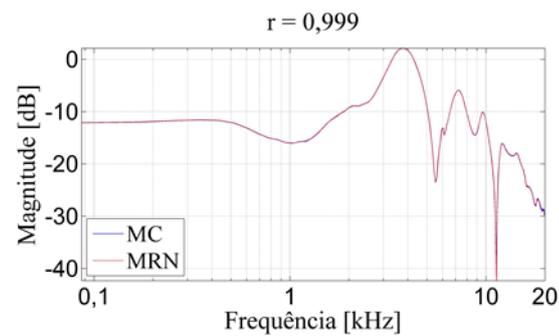


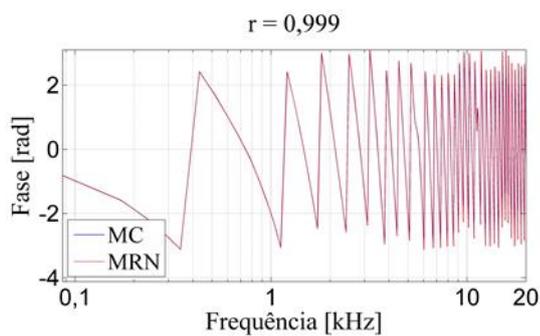
Figura 9: Comparação entre a HRIR/HRTF filtrada, calculada pelos dois métodos (MC e MRN), para a direção $\varphi = 39^\circ$ e $\theta = -21^\circ$. (a) Amplitude; (b) magnitude; e (c) fase.



(a)

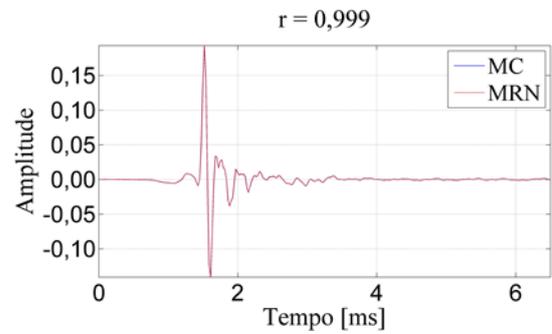


(b)

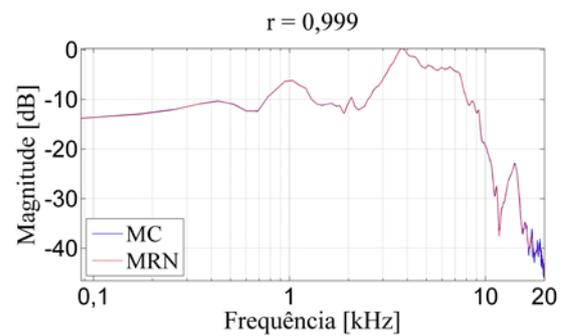


(c)

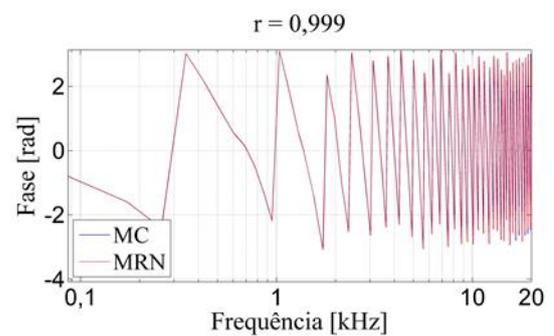
Figura 10: Comparação entre a HRIR/HRTF filtrada, calculada pelos dois métodos (MC e MRN), para a direção $\varphi = 342^\circ$ e $\theta = -65^\circ$. (a) Amplitude; (b) magnitude; e (c) fase.



(a)



(b)



(c)

Figura 11: Comparação entre a HRIR/HRTF filtrada, calculada pelos dois métodos (MC e MRN), para a direção $\varphi = 321^\circ$ e $\theta = 58^\circ$. (a) Amplitude; (b) magnitude; e (c) fase.

6.2. Testes de articulação

Aurilizações, por se tratarem de sinais sonoros a serem ouvidos por humanos, possuem um caráter eminentemente subjetivo e multi-dimensional e, conseqüentemente, oferecem uma grande dificuldade em serem avaliadas por meio de métricas objetivas. Diversas tentativas têm sido feitas no sentido de definir métricas para aurilizações, sendo a mais geral e completa delas o *Spatial Audio Quality Inventory* (SAQI), que reúne um conjunto de 42 termos, em língua inglesa, para avaliação subjetiva da qualidade espacial de um sinal de áudio [36]. Contudo, somente termos bastante vagos como ‘maior’ ou ‘menor’, ‘mais intenso’ ou ‘menos intenso’, são utilizados nesse gênero de avaliação.

Alguns autores têm ido em outra direção, buscando atributos mais objetivos, com vistas a uma métrica que possa fornecer um *valor*. Esse é o caso do *índice de articulação* [32,37]. Índices de articulação, uma métrica utilizada para avaliação de inteligibilidade da fala em salas, embora constituam um pequeno subconjunto da multidimensionalidade do som espacial, permitem uma valoração da aurilização quando confrontado com os mesmos índices avaliados na sala real.

A metodologia utilizada neste trabalho para o levantamento dos índices de articulação de uma sala, real e virtual, é similar à apresentada por [37]. Em essência, em uma sala de aula de médias dimensões (~180 m²), 36 voluntários, estudantes de graduação de ambos os sexos, são posicionados, em quatro grupos de nove, no entorno de quatro posições tomadas, como referência, como indica a Figura 12. Essas quatro posições são as utilizadas posteriormente como as coordenadas dos receptores na simulação da sala.

Listas de cem monossílabos, gravados em câmara anecoica, são emitidos por um orador artificial, ver Figura 13, que reproduz proximamente a direcionalidade de um ser humano. Os ditados são anotados pelos voluntários e posteriormente corrigidos para

identificar o índice de acerto de cada um. A média dos acertos para cada grupo de nove pessoas constitui, então, o índice de articulação real (IAR) daquele par fonte-receptor na sala.

A mesma sala é simulada no código computacional RAIOS 7 – utilizando a modelagem das HRIRs filtradas por meio de RNAs –, para cada um dos quatro pares fonte-receptor e são computadas as BRIRs para essas posições.

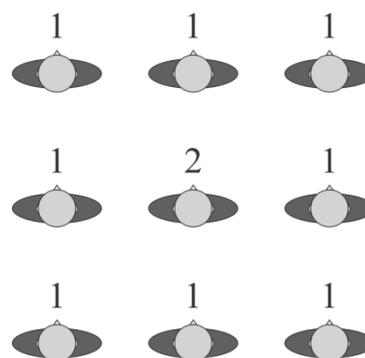


Figura 12: Disposição de cada grupo de nove voluntários com um na posição central (2), posição utilizada na simulação do receptor, e os demais (1) em torno.



Figura 13: Orador artificial utilizado nos testes de articulação reais.

Utiliza-se a mesma lista de palavras do teste de articulação real, porém com a ordem embaralhada. Esse sinal anecoico, então, é convoluído com cada uma das quatro BRIRs, gerando os quatro sinais de aurilização que serão objeto dos ensaios virtuais.

A seguir, esses sinais são escutados por meio de fones de ouvido equalizados por nove voluntários para cada par fonte-receptor e os monossílabos anotados. Finalmente, os ditados

virtuais são corrigidos, atribuindo-se um índice de articulação virtual (IAV) para cada uma das quatro posições fonte-receptor.

A Tabela 1 apresenta a comparação entre os índices de articulação reais e virtuais, para os quatro pares de posição fonte-receptor.

Tabela 1: Comparação entre os índices de articulação reais e virtuais para quatro pares fonte-receptor na sala. Todos os valores estão em percentual (%).

	PAR 1	PAR 2	PAR 3	PAR 4
Medido (IAR)	92	89	88	90
Simulado (IAV)	95	93	90	93
Diferença (R-V)	-3	-4	-2	-3

Como se pode depreender da Tabela 1, os IARs mostraram-se, para todos os quatro pares fonte-receptor, menores do que os IAVs, embora a média da diferença seja de somente 3%. A melhor ideia que se tem por ora sobre essa tendência sistemática é que a reverberação de baixa frequência da sala não é integralmente reproduzida na simulação, que não contempla acústica modal. Contudo, este é um aspecto ainda sob investigação.

7. CONCLUSÕES

Foi apresentada uma nova técnica para modelar as funções de transferência associadas à cabeça humana, que constitui o cerne da geração de respostas impulsivas biauriculares para a produção de aurilizações.

O modelo é baseado em redes neurais artificiais do tipo funções de base radial e foi implementado no código computacional RAIOS 7, resultando em uma economia de tempo computacional de cerca de 90% no cômputo das BRIRs, ou seja, a aurilização pode ser obtida em cerca de 1/10 do tempo.

Como em qualquer procedimento de simulação de acústica de salas há dois blocos de maior custo computacional, a saber: o cálculo dos raios acústicos que chegam a cada receptor; e

o cômputo das respostas impulsivas biauriculares da sala, ao menos um desses blocos beneficiou-se da modelagem via redes neurais, com significativa economia de tempo computacional.

Os resultados comparativos entre as HRIRs/HRTFs filtradas, obtidas pelo método clássico das convoluções e pelo novo modelo das redes neurais artificiais mostraram que as funções são, para todos os efeitos práticos, indistinguíveis, apresentando coeficiente de correlação cruzada normalizada praticamente igual à unidade, tanto no domínio do tempo como no da frequência.

Contudo, como se tratam de aurilizações, o aspecto subjetivo precisa ser considerado e uma métrica envolvendo a intervenção humana se faz necessária. Para tal, foram realizados, em uma única sala, testes de articulação reais e virtuais para quatro posições de recepção (e uma única posição da fonte sonora). Os resultados comparativos entre os índices de articulação reais – medidos diretamente na sala – e os índices de articulação virtuais – obtidos a partir da simulação no código computacional RAIOS 7 – apresentaram diferenças menores que 5%, com média de 3%.

Os testes de articulação deverão ser utilizados como métrica de validação de aurilizações, em diversas salas, com distintos volumes e geometria, de modo a se poder verificar se, efetivamente, as aurilizações geradas pelo código computacional RAIOS 7 mantêm a fidedignidade em situações as mais variadas.

AGRADECIMENTOS

Os autores agradecem aos estudantes de graduação em engenharia que se voluntariaram a participar dos ensaios reais e virtuais.

REFERÊNCIAS

1. VORLÄNDER, M. *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Berlin: Springer-Verlag, 2008.

2. KLEINER, M.; DALENBÄCK, B.I.; SVESSON, P. Auralization - an overview. *J. audio Eng. Soc.*, 41, p. 861, 1993.
3. BLAUERT, J. *Spatial Hearing*. Cambridge: The MIT Press, 1997.
4. RINDEL, J. The use computer modeling in room acoustics. *Journal of Vibroengineering*, 4(3):41–72, 2000.
5. TENENBAUM, R.A.; CAMILO, T.S.; TORRES, J.C.B. and GERGES, S.Y. Hybrid method for numerical simulation of room acoustic: Part 1 – theoretical and numerical aspects. *J. Braz. Soc. Mech. Sci. Engin.*, 29(2):211–221, 2007a.
6. BORK, I. Report on the 3rd Round Robin on room acoustical computer simulation - Part II: Calculations. *Acta Acustica united with Acustica*, 91(4):753–763, 2005.
7. SAVIOJA, L.; SVENSSON, U.P. Overview of geometrical room acoustic modeling techniques. *J. Acoust. Soc. Am.*, 138(2):708–730, 2015
8. EMBRECHTS, J. Randomly traced sound ray techniques. *Acustica*, 51:285–295, 1982.
9. KULOWSKI, A. Algorithmic representation of the ray tracing technique. *Applied Acoustics*, 18:449–469, 1984.
10. ONDET, M; BARBRY, J.L. Modeling of sound propagation in fitted workshops using ray tracing. *J. Acoust. Soc. Am.*, 85(2):787–796, 1989.
11. FARINA, A. RAMSETE – a new pyramid tracer for medium and large-scale acoustic problems. *Proceedings of the Euronoise*, Lyon, 1995
12. ALLEN, J.B.; BERKLEY, D.A. Image method for efficiently simulating small-room acoustics. *J. Acoust. Soc. Am.*, 65, p. 943, 1979.
13. DALENBÄCK, B; KLEINER, M.; SVENSON, P. A macroscopic view of diffuse reflection. *J. audio Eng. Soc.* 42:793–807, 1994.
14. KURZINS, E.; FRICKE, F. The prediction of sound fields in non-diffuse spaces by random walk approach. *J. Sound and Vib.*, 81(4):549–564, 1982.
15. ALARCÃO, D.; BENTO COELHO, J.L.; TENENBAUM, R.A. On modeling of room acoustics by a sound energy transition approach. *Proceedings of EEA Symposium on Architectural Acoustics*, 2000.
16. GARDNER, B.; MARTIN, K. HRTF Measurements of a KEMAR Dummy-Head Microphone. *J. Acoust. Soc. Am.*, vol. 97, n. 6, pp. 3907– 3908, 1995.
17. BRINKMANN, F.; LINDAU, A.; WEINZIERL, S.; VAN DER PAR, S.; OPDAM, R.; VORLÄNDER, M. The FABIAN head-related transfer function database. DOI: 10.14279/depositonce-5718.2, 2017.
18. MILLS, A. On the minimum audible angle. *J. Acoust. Soc. Am.*, 30:237–246, 1958.
19. KISTLER, D. J.; WIGHTMAN, F.L. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *J. Acoust. Soc. Am.* 91(3), 1637–1647, 1992.
20. TORRES, J.C.B; PETRAGLIA, M.R.; TENENBAUM, R.A. An efficient wavelet based HRTF model for auralization. *Acta Acustica united with Acustica*, 90(1):108–120, 2004.
21. HU, H.; ZHOU, L.; MA, H.; WU, Z. HRTF personalization based on artificial neural network in individual virtual auditory space. *Applied Acoustics*, 69(2):163–172, 2008.
22. TENENBAUM, R.A.; TAMINATO, F.O.; MELO, V.S.G.; TORRES, J.C.B. Auralization generated by modeling HRIRs with artificial neural networks and its validation using articulation tests. *Applied Acoustics*, 130, pp. 260–269, 2018.
23. BROOMHEAD, D.; LOWE, D. Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2:321–355, 1988.
24. LI, L.; HUANG, Q. HRTF personalization modeling based on RBF neural network. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, p. 3707–3710, Vancouver, Canada, 2013.
25. TENENBAUM, R.A.; CAMILO, T.S.; TORRES, J.C.B. and STUTZ, L.T. Hybrid method for numerical simulation of room acoustic: Part 2 – validation of the computational code RAIOS 3. *J. Braz. Soc. Mech. Sci. Engin.*, 29(2):223–231, 2007b.
26. FERNANDES, K.M.; TENENBAUM, R.A.; STUTZ, L.T. Problema inverso em propagação de ondas em sólidos com aplicação de método híbrido de otimização. Em: *Técnicas de Inteligência Computacional com Aplicações em Problemas Inversos de Engenharia*. Editora Omnipax, Curitiba, pp. 51–66, 2014.
27. RAHIM, M.G., GOODYEAR, C.C. and KLEIJN, W.B. On the use of neural networks in articulatory speech synthesis. *J. Acoust. Soc. Am.*, v. 93(2):1109–1121, 1992.
28. HAYKIN, S. *Neural Networks and Learning Machines*. 3rd Edition, Prentice Hall, New Jersey, 2009.
29. TAMINATO, F.O. Redes neurais artificiais aplicadas à modelagem de respostas impulsivas associadas à cabeça humana para gerar aurilização. Tese de doutorado, Programa de Pós-graduação em Modelagem Computacional, Universidade do Estado do Rio de Janeiro, 2018.

30. TENENBAUM, R.A.; TAMINATO, F.O.; MELO V.S.G. Fast auralization using radial basis functions type of artificial neural network techniques. *Applied Acoustics*, 157, 106993, 2020.
31. A Ground Truth for Room Acoustic Simulation. <http://dx.doi.org/10.14279/depositonce-6726.2>, outubro 2019.
32. HODGSON, M.; YORK, N.; YANG, W.; BLISS, M. Comparison of predicted, measured, and auralized sound fields with respect to speech intelligibility in classrooms using CATT-acoustic and ODEON. *Acta Acustica united with Acustica*, 94(6):883–890, 2008.
33. MELO, V.S.G.; TENENBAUM, R.A.; NARANJO, J.F.L. A new approach to validate computer modeling auralizations by using articulation indexes. *Acústica & Vibrações*, Vol. 46, pp. 1–8, 2014.
34. POLLOW, M.; BEHLER, G. Variable directivity for platonic sound sources based on spherical harmonics optimization. *Acta Acustica united with Acustica*, 95(6):1082–1092, 2009.
35. KLEIN, J.; POLLOW, M.; VORLÄNDER, M. Optimized spherical sound source for auralization with arbitrary source directivity. *Proceedings of the EAA Joint Symposium on Auralization and Ambisonics*, p. 56–61, Berlin, 2014.
36. LINDAU, A.; ERBES, V.; LEPA, S.; MAEMPEL, H.J.; BRINKMANN, F.; WEINZIERL, S. A spatial audio quality inventory for virtual acoustic environments (SAQI). *Acta Acustica united with Acustica*, 100(5):984–994, 2014.
37. MELO, V.S.G.; LIMA, P.G.; SANTOS, T.C.; TENENBAUM, R.A. Validação de realidade virtual acústica via testes de articulação em salas ruidosas e reverberantes. *Acústica & Vibrações*, 49, pp. 51–57, 2017.