

# Análise de Desempenho de Modelos de Regressão Entre Avaliações Subjetivas e Objetivas Aplicados a Sistemas de Baixo Virtual

Danilo O. Carvalho  
Dpto. de Engenharia de Sistemas  
Universidade de Pernambuco  
Recife, Pernambuco  
Email: doc@poli.br

Carmelo J. A. Bastos-Filho  
Dpto. de Engenharia de Sistemas  
Universidade de Pernambuco  
Recife, Pernambuco  
Email: carmelo.filho@upe.br

Sergio C. Oliveira  
Dpto. de Engenharia de Sistemas  
Universidade de Pernambuco  
Recife, Pernambuco  
Email: sergio.campello@upe.br

**Abstract**—Due to their physical constraints, small-size loudspeakers cannot reproduce low frequencies free of unexpected effects such as distortion. In this article, a psycho-acoustic phenomena-based system was developed to process bass sounds in tracks with different features. The resulting signals are assessed by a set of volunteers and an objective algorithm. Based on these data, some regression processes (linear, gaussian, neural network, regression trees for example) are used to provide functions to generate similar volunteer scores. Validation experiments were performed considering the perception of human beings and automatic methods using 5 songs with different characteristics to evaluate the regression process using a speaker with low frequency reproduction restrictions. Their performance numbers are compared and presented to validate the effectiveness of the system, where gaussian regression models lead to better results.

## I. INTRODUÇÃO

No âmbito dos sistemas de reprodução de áudio, existem diversos tipos de equipamentos envolvidos, dentre os quais destacam-se os alto-falantes. Devido às características dimensionais e físicas dos materiais empregados [1], estes componentes são capazes de reproduzir, com clareza e definição aceitáveis, uma faixa de frequências limitada [2] [3], o que justifica a grande variedade de modelos disponíveis no mercado. Em certos casos, onde o espaço físico para instalação destes dispositivos não representa uma restrição, é comum utilizar alto-falantes de diversos tamanhos, a fim de cobrir tanto quanto possível a faixa do espectro audível da perspectiva humana. Porém, em aplicações com dimensões reduzidas, como em aparelhos de TV e telefones móveis, a reprodução de baixas frequências pode ter sua qualidade comprometida [4] [5].

Para superar estas restrições, foram desenvolvidas técnicas baseadas em fenômenos psicoacústicos [6], que podem ser compreendidos como a distorção sensorial da percepção dos sons ou ilusões auditivas. Dentre estas, destaca-se a *missing fundamental* (MF) [7] [8], na qual as harmônicas superiores produzem, no sistema auditivo humano, a mesma percepção de *pitch* de uma determinada frequência fundamental. Esta técnica foi utilizada com sucesso e implementada em sistemas como o MaxxBass [9] e UltraBass [10], permitindo que alto-

falantes de dimensões reduzidas produzam tons percebidos como graves, mesmo sem a sua presença física.

Visto que os efeitos psicoacústicos produzidos pelo sistema de baixo virtual (VBS - *Virtual Bass System*) são obtidos através da adição de harmônicos gerados por funções não lineares, como o retificador de onda completa (FWR - *full-wave rectifier*) e integrador de onda completa (FWI - *full-wave integrator*) [11], recorte suave (SC - *soft clipping*) [12] e a raiz quadrada do arco tangente (ATSR - *arch-tangent square root*) [13], pode-se obter resultados com diversos níveis de distorção, dependendo do ajuste empregado [14]. Para quantificar estas características de forma subjetiva, o método MUSHRA (*MULTI Stimulus Test with Hidden Reference and Anchor*) [15] proposto pela União Internacional de Telecomunicações (ITU - *International Telecommunication Union*), realiza, por meio de voluntários, a comparação entre diferentes processamentos de uma determinada faixa de áudio em relação a uma referência, onde cada um receberá uma nota indicando melhora, piora ou não alteração na qualidade. Considerando o aspecto objetivo, o método Avaliação Perceptual de Qualidade de Áudio (PEAQ - *Perceptual Evaluation of Audio Quality*) [16] realiza uma análise quantitativa de determinadas características do sinal, como distorção, ruído e largura de banda dos sinais de referência e teste. Realizados estes dois processos, torna-se viável, por meio de métodos de regressão, a obtenção de modelos capazes de relacionar os resultados objetivos, baseadas em indicadores quantificáveis das faixas, e subjetivos, cujos valores representam mais fielmente as preferências dos usuários. Além de permitir a comparação dos indicadores de performance de diversos modelos de regressão, são obtidos modelos computacionais aptos a gerar notas equivalentes às dos voluntários, viabilizando sua utilização em aplicações diversas, como sistemas de baixo virtual auto ajustáveis, por exemplo. Este artigo tem como objetivo apresentar uma análise de desempenho de diversas técnicas de regressão aplicadas para avaliações entre as análises subjetivas e objetivas, visando gerar um modelo que possa avaliar de forma automática o procedimento que tipicamente é realizado por seres humanos. Experimentos de

validação foram executados considerando a percepção de seres humanos e métodos automáticos utilizando 5 músicas com características diferentes para avaliar o processo de regressão utilizando um alto-falante com restrições de reprodução de baixas frequências.

O restante do artigo é organizado da seguinte maneira: Seção II introduz o sistema de baixo virtual, utilizado para geração de faixas com diferentes respostas nas baixas frequências, assim como padrões subjetivos e objetivos de avaliação de qualidade de áudio. Na Seção III, descreve-se a implementação, configurações e utilização deste sistema, a escolha das faixas para processamento bem como a realização dos experimentos de avaliação com voluntários e o algoritmo empregado na avaliação objetiva. Também se descrevem os tratamentos dados aos resultados dessas avaliações. Na Seção IV, definem-se os métodos de regressão aplicados aos resultados dos testes e seus resultados. A Seção V apresenta as conclusões do artigo.

## II. SISTEMA DE BAIXO VIRTUAL E AVALIAÇÃO DE QUALIDADE DE ÁUDIO

### A. Sistema de Baixo Virtual

Com o objetivo de obter faixas de áudio nos quais a percepção das baixas frequências se dá através do efeito psicoacústico do fenômeno MF, Aarts [6] propõe o sistema de processamento de sinal observado na Figura 1, no qual se implementa uma sequência de filtros em conjunto com um dispositivo não linear, responsável por gerar o resultado desejado.

Cada elemento desempenha funções específicas [11], descritas a seguir :

- HPF 1: filtro passa alta localizado no caminho primário do processamento, cuja função é retirar do sinal original apenas a porção de frequências que podem ser executadas pelo alto falante sem geração de distorções harmônicas, isolando do áudio apenas a parte que não necessita de processamento pelo dispositivo não linear. A frequência de corte deste filtro deve ser igual a frequência de ressonância do alto falante.
- LPF: este filtro passa baixa, cujo flanco é ajustado para a frequência de corte do alto falante, extrai do sinal original toda a porção de baixas frequências que geram distorções e perda de qualidade, as quais serão posteriormente processadas pelo dispositivo não linear.
- NLD: do inglês *non linear device*, este dispositivo não linear opera sobre o sinal de baixa frequência resultante do LPF, sendo responsável pela produção dos efeitos psicoacústicos do MF através da geração dos componentes harmônicos de frequência acima da faixa de corte do alto falante, preservando o tom original.
- HPF 2: após a operação do NLD, o sinal resultante é composto pelas componentes de baixa frequência do áudio original e suas harmônicas, geradas no processo. A fim de evitar a distorção, este filtro, cuja frequência de corte é igual a frequência de ressonância do alto

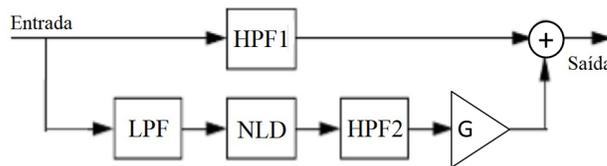


Fig. 1. Fluxograma de processamento de sinal do Sistema de Baixo Virtual.

falante, retira apenas a porção do espectro responsável pelo efeitos psicoacústicos.

- G: tendo em vista que as frequências harmônicas geradas pelo dispositivo não linear possuem amplitudes consideravelmente menores que o sinal original, o volume deste áudio tende a ser significativamente menor do que o gerado pelo HPF1. Para que estes sejam compatíveis, o sinal resultante do HPF2 sofre um ganho, que pode ser parametrizado.

Para se obter os efeitos do baixo virtual, o dispositivo não linear é o componente de maior importância no sistema. São encontradas na literatura diversas formas de implementação, dentre as quais o FWR é uma das mais citadas, calculando o valor absoluto do sinal de entrada [11]. Para uma onda senoidal simples, esta função gera um sinal de saída com o dobro da sua frequência, apresentando apenas seus harmônicos pares. Este processo proporciona uma melhora em relação ao áudio original, porém seu desempenho ainda deixa a desejar se comparada a outras opções, sendo mais indicado para aplicações que disponham de recursos limitados.

Tendo em vista a limitação do FWR, o FWI realiza a integração do módulo do sinal de entrada a cada ciclo, reiniciando o cálculo a cada cruzamento positivo da função no eixo do tempo. Esta operação gera como saída um sinal de frequência fundamental igual a função original, cujo espectro contém tanto os harmônicos ímpares quanto os pares e, por consequência, produz uma impressão mais intensa e nítida dos tons graves [11].

A observação de efeitos de baixo virtual também pode se dar por meio de outras equações não lineares. O SC aparece como uma opção relevante, a qual obteve melhores avaliações do que o FWR e FWI em testes subjetivos[12].

Além do SC, outra função que pode ser utilizada como NLD é o ATSR [13], cujos parâmetros permitem uma grande variedade de ajustes e, conseqüentemente, diferentes respostas em relação a geração de harmônicos, influenciando diretamente na percepção dos efeitos psicoacústicos. As funções correspondentes podem ser visualizadas na Tabela I.

### B. Avaliação de Qualidade de Áudio

Tendo em vista que o sistema de baixo virtual é composto por etapas, onde a principal delas é realizada por um dispositivo não linear que pode ser parametrizado, assim como o ganho após o filtro HPF2, os resultados deste processamento podem ser satisfatórios, mas também imperceptíveis ou mesmo muito distorcidos. Para que a melhor resposta seja alcançada, é necessário realizar avaliações de qualidade sobre os áudios

TABELA I  
FUNÇÕES DOS DISPOSITIVOS NÃO LINEARES

NLD	Função
FWR	$f(x) =  x $
FWI	$f(x) = \int_a^b  x $
SC	$f(x) = \frac{x}{K x +1}$
ATSR	$f(x) = \alpha \arctan(\beta x) + \psi \sqrt{1 - (\zeta x)^2} - \psi$

obtidos a partir de diversas combinações de configurações. Uma das abordagens que pode ser adotada é a realização de avaliações subjetivas, nas quais os próprios ouvintes classificam cada uma das amostras do conjunto.

Com o propósito de estabelecer uma metodologia formal para execução destas avaliações subjetivas de qualidade, a União Internacional de Telecomunicações (ITU - *International Telecommunications Union*) introduz na recomendação ITU-R BS.1534-1 [15], um método para avaliação subjetiva de sistemas de codificação de qualidade intermediária (*Method for subjective assessment of intermediate quality level of coding systems*) baseado em uma ferramenta de teste multi estímulo com âncora e referência oculta (MUSHRA - *MULTI Stimulus test with Hidden Reference and Anchor*), gerando resultados precisos e confiáveis. Esta recomendação descreve um sistema no qual o avaliador deve analisar um conjunto de faixas de áudio pré processadas, quantificando a qualidade de cada uma delas em relação a uma referência sem processamento. Vale ressaltar que neste conjunto está a própria referência (não identificada), o que permite que o aplicador do teste detecte classificações muito discrepantes e realize as compensações necessárias. Além disso, também são definidos aspectos como escopo de aplicação, critérios para seleção dos avaliadores, sinais de teste (quantidade e duração de amostras), tratamento estatístico dos resultados entre outros.

Em contrapartida à abordagem subjetiva, as análises objetivas de qualidade também se estabelecem como alternativas a serem consideradas, de maneira que os resultados independem de variáveis ligadas a características humanas como humor, preferências musicais, capacidade auditiva entre outras, proporcionando pouca ou nenhuma variação nos seus resultados. Neste contexto, a recomendação ITU-R BS.1387-1 [16] propõe um método para medição objetiva de percepção de qualidade, no qual diversas características do sinal são medidas, gerando um conjunto de valores denominado modelo de variáveis de saída (MOV - *Model Output Variables*), como observado na Tabela II.

### III. ELEMENTOS DO SISTEMA DE TESTES E SELEÇÃO DAS FAIXAS MUSICAIS

Considerando que a etapa fundamental deste trabalho se dá em torno da avaliação de qualidade sonora de sistemas de baixo virtual, torna-se necessária a definição dos equipamentos envolvidos na produção destes sons. Para tal, foi utilizado um arranjo composto por um computador, que armazena as faixas musicais e proporciona aos usuários a interface de

TABELA II  
DESCRIÇÃO DOS MOV'S

Referência	Nome	Descrição
MOV 1	$BandwidthRef_B$	Largura de banda do sinal de referência
MOV 2	$BandwidthTest_B$	Largura de banda do sinal de teste
MOV 3	$TotalNMR_B$	Relação ruído máscara
MOV 4	$WinModDiff1_B$	Diferença de modulação em janela
MOV 5	$ADB_B$	Distorção média do bloco
MOV 6	$EHS_B$	Estrutura harmônica do erro
MOV 7	$AvgModDiff1_B$	Diferença média de modulação
MOV 8	$AvgModDiff2_B$	Diferença média de modulação
MOV 9	$RmsNoiseLoud_B$	Distorção de volume
MOV 10	$MFPD_B$	Probabilidade filtrada máxima detecção
MOV 11	$RealDistFrames_B$	Frames relativamente perturbados

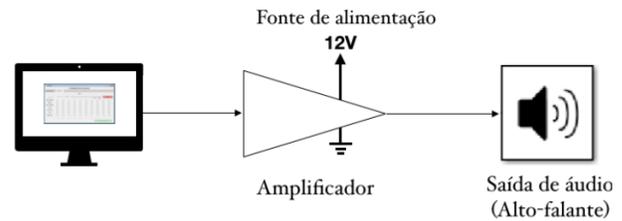


Fig. 2. Diagrama de interligação de equipamentos do sistema de teste.

software para executá-las, seguido pelo amplificador, que eleva a amplitude dos sinais que serão finalmente reproduzidos pelo alto-falante, representados na figura 2. Para a análise, foi empregado um computador com sistema operacional Windows 7 Professional 64 bits, com processador Dual Core 2.3 GHz, memória RAM DDR3 de 4Gb e saída de áudio via plug tipo P2. Para a etapa de amplificação do sinal, foi utilizado um modelo Taramps TL-500 com resposta de frequência (-3dB) de 80Hz a 40KHz, tensão de alimentação de 8 a 16V.

O alto falante utilizado para validação dos experimentos possui frequência de corte por volta de 200Hz, 8 ohms e 1W. Sua curva de resposta foi obtida através de experimento prático [17] e pode ser observada na Figura 3, indicando a relação entre frequência (Hz) e nível de pressão de som (dB).

Com base neste arranjo, foram selecionadas cinco faixas nas quais se observa a presença de componentes harmônicos abaixo do limiar de 200Hz. Considerando o aspecto do gênero, foram priorizadas músicas de rock e pop, uma vez que estes são os estilos onde os sistemas de baixo virtual são mais frequentemente aplicados [11]. Com base nestes critérios, foram selecionadas as músicas e as respectivas bandas listadas na Tabela III.

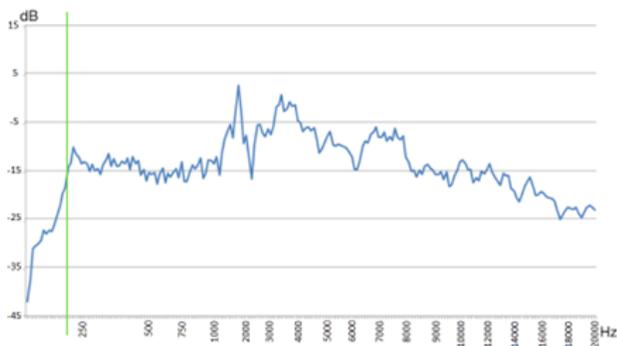


Fig. 3. Curva de resposta do alto falante.

 TABELA III  
FAIXAS UTILIZADAS NOS EXPERIMENTOS

Número	Nome da faixa	Artista
Faixa 1	Bad	Michael Jackson
Faixa 2	Billie Jean	Michael Jackson
Faixa 3	How many more times	Led Zeppelin
Faixa 4	Other side	Red Hot Chilli Peppers
Faixa 5	The chicken	Jaco Pastorius

#### IV. IMPLEMENTAÇÃO DO VBS E EXPERIMENTOS DE AVALIAÇÃO DE QUALIDADE DE ÁUDIO

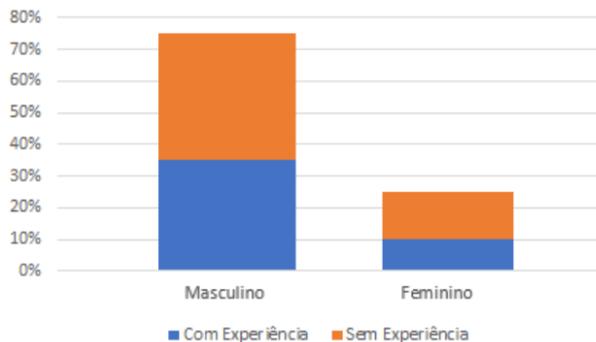
##### A. Experimentos subjetivos

A fim de investigar as preferências de um conjunto de voluntários em relação aos efeitos de melhoria na percepção de baixas frequências promovidas por sistemas de baixo virtual, os experimentos subjetivos de avaliação de qualidade sonora foram divididos em duas etapas, onde a primeira foca na verificação de qual função não-linear gera as melhores avaliações e a segunda na sua melhor configuração. Para tal, estes voluntários avaliaram versões das faixas exibidas na III processadas por diversas configurações de VBS com o auxílio de uma interface de software.

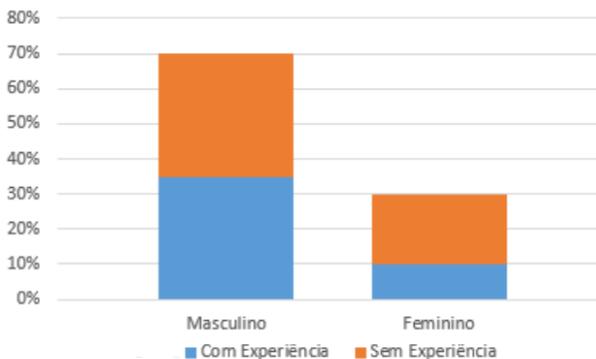
Na Figura 4, pode-se observar a representação gráfica do perfil dos voluntários para a primeira (a) e segunda (b) etapas, para as quais foram selecionados 20 voluntários em cada, com idade entre 18 e 40 anos, mais concentrados na faixa de 19 a 24 anos, de ambos os sexos, com e sem experiência musical.

Cada faixa de áudio selecionada foi processada de acordo com o esquema da Figura 1, onde foram parametrizadas as frequências de corte  $\omega_{c1} = 250Hz$  e  $\omega_{c2} = 300Hz$  para o LPF e  $\omega_{c1} = 250Hz$  e  $\omega_{c2} = 200Hz$  para o HPF, de acordo com a frequência de corte do alto falante observada na Figura 3.

As avaliações foram realizadas através de interface gráfica desenvolvida em Matlab, capaz de reproduzir as versões processadas de cada faixa incluindo a versão original, armazenando as notas dadas por cada avaliador, considerando da faixa de -100, indicando piora, à 100 indicando melhora na qualidade sonora. Para a segunda etapa, foram realizadas melhorias com base em sugestões dos voluntários, restringindo



(a)



(b)

Fig. 4. Quantidade de voluntários por sexo e experiência musical.



Fig. 5. Interface do software de avaliação subjetiva de qualidade.

às possibilidades de avaliação para “muito pior”, “pior”, “pouco pior”, “igual”, “pouco melhor”, “melhor” e “muito melhor”, correspondendo aos valores -3, -2, -1, 0, 1, 2, e 3, respectivamente, como observado na tela da interface gerada apresentada na Figura 5.

1) *Primeira etapa*: Especificados os equipamentos do sistema de reprodução sonora e configurações dos filtros do VBS bem como a seleção dos voluntários, resta a definição das funções não lineares e seus parâmetros. Para tal, foram utilizadas diversas configurações para as funções descritas na Seção II, listadas na Tabela IV.

As notas das avaliações foram armazenados pela interface

TABELA IV  
CONFIGURAÇÕES DOS NLD'S NO PRIMEIRO TESTE SUBJETIVO.

NLD	G	$\alpha$	$\beta$	$\psi$	$\zeta$	K
ATSR 1	1	1,1	0,5	0,5	0,4	-
ATSR 2	1	1,2	0,5	0,9	2,0	-
ATSR 3	3	3,0	1,5	1,5	3,0	-
ATSR 4	0,5	2,5	0,9	0,9	2,5	-
SC 1	0,5	-	-	-	-	1,3
SC 2	0,5	-	-	-	-	4,2
SC 3	0,25	-	-	-	-	2,5
FWR 1	0,5	-	-	-	-	-
FWI 1	0,25	-	-	-	-	-

TABELA VII  
CONFIGURAÇÕES DOS NLD'S NO SEGUNDO TESTE SUBJETIVO.

NLD	G	$\alpha$	$\beta$	$\psi$	$\zeta$
ATSR 1	1,5	1,5	0,45	0,45	1,5
ATSR 2	2,0	1,5	0,45	0,45	1,5
ATSR 3	2,5	1,5	0,45	0,45	1,5
ATSR 4	1,5	2,5	0,9	0,9	2,5
ATSR 5	2,0	2,5	0,9	0,9	2,5
ATSR 6	2,5	2,5	0,9	0,9	2,5
ATSR 7	1,5	3,5	1,8	1,8	3,5
ATSR 8	2,0	3,5	1,8	1,8	3,5
ATSR 9	2,5	3,5	1,8	1,8	3,5

TABELA V  
MÉDIAS DA PRIMEIRA AVALIAÇÃO SUBJETIVA.

NLD	Faixa 1	Faixa 2	Faixa 3	Faixa 4	Faixa 5
ATSR 4	39,30	-6,95	-1,2	25,65	17,20
ATSR 2	37,80	-19,60	9,2	7,50	6,95
SC 4	-14,65	3,65	-0,1	4,80	34,20
ATSR 1	23,15	-4,50	22,7	6,40	-26,95
SC 2	-3,20	-3,10	-14,0	-13,80	16,95
SC 3	-0,25	-14,40	-15,1	-42,75	29,60
FWR	-1,55	-35,00	-7,9	-47,15	13,80
ATSR 3	-60,95	-79,85	-1,5	-77,60	-70,50
FWI	-86,10	-82,55	-93,8	-93,25	-13,95

TABELA VIII  
MÉDIAS DA SEGUNDA FASE DE AVALIAÇÕES SUBJETIVAS.

NLD	Faixa 1	Faixa 2	Faixa 3	Faixa 4	Faixa 5
ORIGINAL	0,10	0,25	0,45	0,80	0,40
ATSR 1	0,30	0,80	0,75	0,20	0,35
ATSR 2	0,55	0,60	0,50	0,65	0,40
ATSR 3	2,50	1,50	0,45	0,45	1,50
ATSR 4	0,55	-0,60	0,80	-0,15	0,65
ATSR 5	0,55	-1,65	-0,15	-0,40	0,80
ATSR 6	-0,30	-2,10	-1,20	-1,30	0,85
ATSR 7	-1,70	-2,65	-1,90	-2,40	-0,40
ATSR 8	-2,45	-2,75	-2,55	-2,85	-1,50
ATSR 9	-2,60	-2,95	-2,70	-2,95	-1,95

TABELA VI  
RESULTADOS DO TESTE DE WILCOXON PARA PRIMEIRA FASE SUBJETIVA.

NLD	Faixa 1	Faixa 2	Faixa 3	Faixa 4	Faixa 5
ATSR 4	0,00175	0,96339	0,95052	0,06991	0,03650
ATSR 2	0,0027	0,99200	0,80656	0,51490	0,1272
SC 2	0,7717	0,94373	0,98255	0,99582	0,04171
SC 3	0,73381	0,98969	0,97711	0,99976	0,00418
SC 4	0,94189	0,8140	0,89104	0,77801	0,00200
ATSR 1	0,13419	0,85236	0,39677	0,78895	0,99730
ATSR 3	0,99992	0,99995	0,95450	0,99996	0,99995
FWR	0,76385	0,99924	0,99370	0,99994	0,0651
FWI	0,99994	0,99995	0,99995	0,99996	0,88753

de software e as médias dos resultados são exibidos na Tabela V, onde os resultados positivos são destacados em verde e os negativos em vermelho. A média geral de cada NLD também é apresentada.

Por fim, o teste de classificação pareada de Wilcoxon, demonstrado pela Tabela VI, exhibe os valores-p calculados para cada uma das amostras de áudio do conjunto, onde o ATSR4 apresenta significância estatística maior que 95% para as faixas 1 e 5, se aproximando também para a faixa 4. Em seguida, o ATSR2, SC2, SC3 e SC4 alcançaram resultados relevantes para apenas uma das faixas. As demais variações não atingiram índices estatisticamente expressivos em nenhum dos casos.

Desta forma, o algoritmo ATSR foi selecionado para a

realização da segunda etapa de avaliações subjetivas, onde foram testadas mais variações de parametrização.

2) *Segunda etapa:* A partir dos resultados obtidos na primeira etapa, uma nova bateria de testes foi realizada apenas com variações do ATSR, listadas na Tabela VII, visto que suas versões obtiveram avaliações superiores em relação aos outros NLDs.

De maneira análoga a primeira etapa, as médias dos resultados da segunda bateria de avaliações subjetivas podem ser verificadas na Tabela VIII.

### B. Experimentos objetivos

Para realização da fase objetiva da avaliação de qualidade de áudio, foram utilizados os algoritmos de cálculo de MOV's desenvolvidos pelo laboratório de telecomunicações e processamento de sinais da universidade McGill [18], implementados em Matlab. Neste caso, foram avaliadas as mesmas faixas utilizadas na segunda avaliação subjetiva, apenas com variações de processamento do ATSR. Com base nos resultados obtidos neste processo, disponíveis na Tabela XI, foram observados os pontos listados a seguir:

- O MOV 10 gerou apenas valores iguais a 1, inviabilizando sua utilização, uma vez que não acrescenta informações relevantes para a fase de regressão.
- As diferentes configurações do ATSR apresentadas na Tabela VII interferem minimamente na largura de banda do sinal de referência e de teste.

- Os resultados de relação ruído-máscara são semelhantes para todas as faixas, com exceção da Faixa 2, que apresenta pouca dispersão.
- Os MOVs 7 e 8, que indicam a diferença média de modulação, possuem características em comum, onde as faixas 1 e 2 apresentam baixa dispersão e resultados próximos a 800 e 200. Para as faixas 3, 4 e 5, os índices são similares para os dois casos.

#### V. REGRESSÃO ENTRE AVALIAÇÕES OBJETIVA E SUBJETIVA

Com o objetivo de obter modelos computacionais capazes de qualificar uma faixa de forma similar a um avaliador humano, os dados obtidos nos procedimentos descritos na Seção IV-B foram submetidos a processos de regressão. Para viabilizar execução desta etapa, foi utilizada a aplicação *regression learner* do *software* MATLAB versão R2017a, onde os valores de entrada para treinamento correspondem aos resultados dos MOV's (com exceção do MOV10) e a saída esperada é indicada pela nota da segunda etapa avaliação subjetiva, ambos normalizados. Neste procedimento, foram utilizados todos os algoritmos disponibilizados por esta plataforma, permitindo uma avaliação comparativa para diferentes técnicas. A Tabela IX apresenta os modelos disponíveis que foram testados nos experimentos realizados. Pode-se observar que existem modelos diversos para cinco diferentes classes, como por exemplo, os modelos lineares, gaussianos, baseados em árvores, baseados em vetor suporte e redes neurais artificiais. Para validação, foi utilizado o método de *cross-validation* com o parâmetro de *k-fold* ajustado para 25 (máximo disponível) em todos os casos.

A partir da análise dos indicadores de desempenho geralmente utilizados para avaliar processos de regressão. RMSE, R-Squared, MSE e MAE disponibilizados pelo *regression learner*, pode-se verificar que os estimadores baseados em modelos gaussianos geraram os melhores resultados, com os quatro melhores desempenhos dentre os 24 modelos disponíveis e testados, exibidos na Tabela X. Como destaque, a Regressão Gaussiana (exponencial) obteve erro médio absoluto de 5,6% e a raiz do erro médio quadrático abaixo de 7%.

Para melhor explicitar visualmente a qualidade do processo de regressão entre as avaliações subjetivas e medidas objetivas, são apresentados os gráfico de dispersão na Figura 6. Pode-se observar a relação entre a resposta real das médias da segunda avaliação subjetiva e a estimativa dada pelo modelo de regressão gaussiana exponencial após seu treinamento e validação cruzada. Neste, pode-se verificar que todos os pontos de distribuem em torno da reta central, indicando seu nível de acerto.

#### VI. CONCLUSÃO

Neste trabalho foi realizado um estudo detalhado para analisar de a relação de avaliações subjetivas e objetivas no problema de reprodução psicoacústica de baixas frequências em alto-falantes que não conseguem reproduzir estas baixas frequências. Para geração de dados, foi desenvolvido um estudo experimental considerando especificamente um alto

TABELA IX  
MODELOS DE ESTIMADORES DE REGRESSÃO DISPONÍVEIS NO *regression learner*.

Modelo	Tipo
Linear	<ul style="list-style-type: none"> <li>• linear</li> <li>• <i>interactions linear</i></li> <li>• <i>robust linear</i></li> <li>• <i>stepwise linear</i></li> </ul>
Gaussianos	<ul style="list-style-type: none"> <li>• <i>squared exponential GPR</i></li> <li>• <i>matern 5/2 GPR</i></li> <li>• <i>exponential GPR</i></li> <li>• <i>rational quadratic GPR</i></li> </ul>
Árvores	<ul style="list-style-type: none"> <li>• <i>fine tree</i></li> <li>• <i>medium tree (minimum leaf size: 12)</i></li> <li>• <i>coarse tree (minimum leaf size: 36)</i></li> <li>• <i>boosted trees (minimum leaf size: 8)</i></li> <li>• <i>bagged trees (minimum leaf size: 8, number of learners: 30)</i></li> </ul>
SVM	<ul style="list-style-type: none"> <li>• <i>linear SVM</i></li> <li>• <i>quadratic SVM</i></li> <li>• <i>cubic SVM</i></li> <li>• <i>fine gaussian SVM</i></li> <li>• <i>medium gaussian SVM</i></li> <li>• <i>coarse gaussian SVM</i></li> </ul>
Redes neurais artificiais	<ul style="list-style-type: none"> <li>• <i>narrow neural network (number of fully connected layers: 1, first layer size: 10)</i></li> <li>• <i>medium neural network (number of fully connected layers: 1, first layer size: 25)</i></li> <li>• <i>wide neural network (number of fully connected layers: 1, first layer size: 100)</i></li> <li>• <i>bilayered neural network (number of fully connected layers: 2, first layer size: 10 Second layer size: 10)</i></li> <li>• <i>trilayered neural network (number of fully connected layers: 3, first layer size: 10, second layer size: 10, third layer size: 10)</i></li> </ul>

TABELA X  
INDICADORES DE PERFORMANCE DOS MODELOS DE REGRESSÃO PARA K-FOLD=25.

Modelo	RMSE	R-Squared	MSE	MAE
Regressão Gaussiana (exponencial)	0,0696	0,96	0,0048442	0,056304
Regressão Gaussiana (matern 5/2)	0,073597	0,95	0,0054166	0,058333
Regressão Gaussiana (Quadratica racional)	0,075615	0,95	0,0057177	0,059711
Regressão Gaussiana (Quadratica exponencial)	0,078808	0,95	0,0062107	0,063936

falante que apresenta resposta não adequada em frequências abaixo de 200Hz, um sistema para produção de efeitos psi-

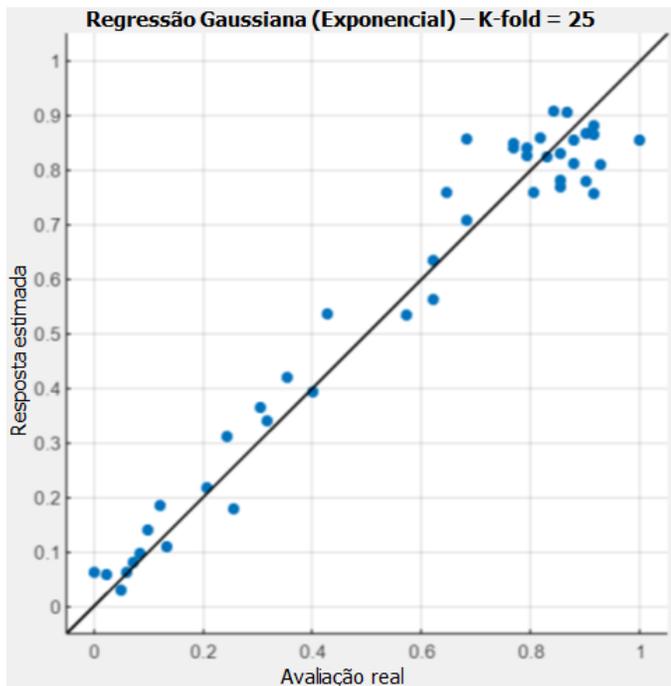


Fig. 6. Gráfico de dispersão do modelo de regressão gaussiana para k-fold=25.

coacústicos aplicado a esta faixa do espectro, o qual se baseia em um conjunto de filtros e um dispositivo não linear, sendo este último seminal para o bom funcionamento do sistema. Utilizando quatro funções com diferentes configurações para geração de harmônicos no dispositivo não linear, foi obtido um conjunto de faixas com respostas mais perceptíveis para frequências graves.

Posteriormente, as faixas geradas a partir destes processamentos foram avaliadas de forma subjetiva, por meio de voluntários, e de forma objetiva, através de algoritmos. Após estas análises, o dispositivo ATSR apresentou os melhores resultados, sendo utilizado em uma segunda bateria de avaliações.

Após a obtenção dos dados relativos à segunda etapa, foram realizados processos de regressão com diversos modelos e funções *kernel*, a fim de buscar a melhor forma de relacionar resultados objetivos e subjetivos e, por conseguinte, obter um algoritmo capaz de gerar classificações compatíveis às dos voluntários de forma automática. Este processo demonstrou que os estimadores gaussianos, especialmente o modelo baseado em função exponencial conseguem obter os melhores resultados. Nos experimentos realizados, o melhor modelo de baseado em regressão gaussiana conseguiu obter valores de RMSE inferior a 7% para avaliações em um cenários em cinco músicas foram consideradas.

Para trabalhos futuros, pretende-se ampliar a quantidade de músicas e gêneros musicais testados para avaliar a consistência dos resultados. Também está no escopo de investigações futuras a aplicação deste modelo de regressão que emula a avaliação automática para o desenvolvimento de novas funções para geração de harmônicos no dispositivo não linear.

## AGRADECIMENTOS

Os autores agradecem o apoio financeiro da FITec e CNPq, e o apoio institucional da Universidade de Pernambuco e CAPES.

## REFERENCES

- [1] H. F. Olson, *Acoustical Engineering*. Van Nostrand, Prince-ton, New Jersey, 1957.
- [2] W. S. Gan, S. M. Kuo, e C. W. Toh, "Virtual bass for home entertainment, multimedia PC, game station and portable audio systems" *IEEE Trans. Consumer Electron*, v. 47, pp. 787–796, Novembro 2001.
- [3] S. Zhang, L. Xie, Z. H. Fu e Y. Yuan, *A Hybrid Virtual Bass System with Improved Phase Vocoder and High Efficiency*. 9th International Symposium on Chinese Spoken Language Processing, Singapura, 2014.
- [4] S. Cecchi, E. Moretti e F. Piazza, *A New Approach to Bass Enhancement Based on Prony's Method*. 15th Intl. Conf. on Digital Signal Processing, Estados Unidos, 2007.
- [5] J. Lee, E. Song, Y. Park e D. Youn, "Effective Bass Enhancement Using Second-order Adaptive Notch Filter" *IEEE Transactions on Consumer Electronics*, v. 54, Maio 2008.
- [6] R. Aarts, *The Application of Illusions and Psychoacoustics to Small Loudspeaker Configurations*. AES 22ND UK CONFERENCE, Reino Unido, 2007.
- [7] H. Mu e W. S. Gan, *A Virtual Bass System with Improved Overflow Control*. International Conference on Acoustics, Speech, and Signal Processing, Austrália 2015.
- [8] N. Ono, K. Miyamoto, J. Le Roux, H. Kameoka e S. Sagayama, *Separation of a monaural audio signal into harmonic/percussive components by complementary diffusion on spectrogram*. Proc. EUSIPCO, Suíça 2008.
- [9] D. Ben-Tzur e M. Colloms, *The effect of MaxxBass psycho-acoustic bass enhancement on loudspeaker design*. 106th Audio Eng. Soc. Convention, 1999.
- [10] G. F. M. D. Poortere, C. M. Polisset e R. M. Aarts, *Ultra bass*. U.S. patent, 613 433 017, Outubro 2000.
- [11] R. Aarts, E. Iarsen e D. Schoben, *Improving Perceived Bass and Reconstruction of High Frequencies for Band Limited Signals*. MPCA, Bélgica 2002.
- [12] B. T. Gerstel, *Tunable Virtual Bass Enhancement*. Maio 2009.
- [13] N. Oo, W. S. Gan e W. T. Lim, *Generalized Harmonic Analysis of Arctangent Square Root (ATSR) Non-linear Device for Virtual Bass System*. ICASSP, Estados Unidos 2010.
- [14] H. Mu, W.S. Gan e E. L. Tan, "An Objective Analysis Method for Perceptual Quality of a Virtual Bass System" *IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING*, v. 23, pp. 840–850, Maio 2015.
- [15] ITU-R BS.1534-1, *Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA)*. International Telecommunications Union, Suíça 2001.
- [16] ITU-R BS.1387-1, *Method for objective measurements of perceived audio quality*. International Telecommunications Union, Suíça 2001.
- [17] J. P. G. Oliveira e H. Lima, *Análise de dados experimentais da caracterização dos alto-falantes*. FITEC, 2016. Disponível em <http://www.fittec.org.br/>. Acesso em: 03/07/2020.
- [18] P. Kabal, *An examination and interpretation of ITU-R BS.1387: perceptual evaluation of audio quality*. TSP Lab Technical Report, Department of Electrical and Computer Engineering, Dezembro 2003.

TABELA XI  
RESULTADOS DOS TESTES OBJETIVOS.

Faixa	NLD	MOV1	MOV2	MOV3	MOV4	MOV5	MOV6	MOV7	MOV8	MOV9	MOV10	MOV11
Faixa 1	ATSR 1	419,3	417,9	10,5	911,3	285,1	1,5	732,3	168,5	11,4	1	1
Faixa 1	ATSR 2	419,2	417,9	11	911,8	285,7	146,8	731,5	169,2	11,6	1	1
Faixa 1	ATSR 3	419,2	417,8	11,6	913	286,5	146,8	731,4	170	12,1	1	1
Faixa 1	ATSR 4	419,3	417,9	148,8	924	291,5	146,2	733,4	173,9	143,9	1	1
Faixa 1	ATSR 5	419,1	417,8	167,9	933,1	294,5	1,5	735,3	176,2	147,3	1	1
Faixa 1	ATSR 6	418,7	417,6	184,3	947,7	297,2	1,7	745,3	183,9	14,8	1	1
Faixa 1	ATSR 7	416,2	415,7	215,7	99	305,4	293	763,1	197,1	142,8	1	1
Faixa 1	ATSR 8	415,1	414,8	226,1	999,1	311,6	370,1	756,8	199,1	136,5	1	1
Faixa 1	ATSR 9	414,3	414	232,5	990,6	316,6	435,6	749,2	194	130,1	1	1
Faixa 2	ATSR 1	423,1	419,9	29,2	857,5	307	697,2	824,8	144,4	190,7	1	1
Faixa 2	ATSR 2	423,1	419,9	292,3	853,4	307,4	732,8	823,3	143,9	193,4	1	1
Faixa 2	ATSR 3	422,8	419,7	292,4	851,7	308,1	731	823,6	144	195,1	1	1
Faixa 2	ATSR 4	421,2	419,8	292,7	891,5	313,5	2,5	837,4	147,2	179,2	1	1
Faixa 2	ATSR 5	420,1	419,4	293,2	928,8	317,8	369,3	844,7	148,9	173,4	1	1
Faixa 2	ATSR 6	419	418,6	293,9	949,3	3,2	459,3	847	148,1	17,3	1	1
Faixa 2	ATSR 7	416,8	416,5	297	995	331,6	704,5	86,3	148,5	193,2	1	1
Faixa 2	ATSR 8	416	415,6	300,3	98,7	336,8	823,6	848,6	145,8	213,5	1	1
Faixa 2	ATSR 9	415	414,5	304,3	984,4	340,8	905,8	838,5	145,3	228,4	1	1
Faixa 3	ATSR 1	701,6	701,6	-118	402,8	167	12,2	397	597,7	387,5	1	764,327
Faixa 3	ATSR 2	701,5	701,5	-909,3	432,8	187,1	12,8	431,8	607	501,6	1	926,316
Faixa 3	ATSR 3	701,7	701,7	-680,5	446,8	202,7	128,4	450,7	588	607,4	1	969,591
Faixa 3	ATSR 4	701,4	701,4	770,9	567,6	243,6	147,1	570,5	614,5	158,9	1	997,661
Faixa 3	ATSR 5	699,7	699,7	3,8	8,8	258,6	215,9	799,7	979,6	249,2	1	997,661
Faixa 3	ATSR 6	696,7	696,7	643,2	140	274,7	486	12,9	182,1	322,8	1	997,661
Faixa 3	ATSR 7	687,2	687	131,8	288,9	309,8	303,9	282,2	457,9	5,5	1	99,883
Faixa 3	ATSR 8	682,2	681,8	164,7	337	321,6	463,7	335,8	566,7	663,6	1	99,883
Faixa 3	ATSR 9	678,5	677,2	18,9	368,7	329	585,7	368,4	631,8	716,7	1	1
Faixa 4	ATSR 1	748,6	748,4	-122,7	364,6	159,9	494	346,1	663,6	28,2	1	760,768
Faixa 4	ATSR 2	748,1	748,1	-956,3	389,4	180,4	451,3	3,8	705,2	412,9	1	864,959
Faixa 4	ATSR 3	748	748	-717,5	409,2	196,6	397,8	400,6	743,9	56,1	1	968,568
Faixa 4	ATSR 4	747,8	747,8	489,3	575,7	239,2	28,4	560,5	10,8	148,3	1	994,761
Faixa 4	ATSR 5	741,4	741,4	357,5	12,5	255,6	145,3	11	268,3	195,4	1	995,343
Faixa 4	ATSR 6	727,3	727,3	623,2	253,3	2,7	408,1	215,6	630,9	222,6	1	997,672
Faixa 4	ATSR 7	665,6	665,2	165	549,1	314,8	981,6	42,8	159	394,1	1	998,836
Faixa 4	ATSR 8	624,2	623,6	212	51,1	328,6	113	428,8	162,4	451,7	1	1
Faixa 4	ATSR 9	600,8	599,8	240,5	489,9	336,6	119,1	443,5	172,1	479	1	1
Faixa 5	ATSR 1	762,2	762,2	-109,4	327,5	153,6	314,9	322,5	308,7	501	1	749,129
Faixa 5	ATSR 2	762,2	762,1	-814,7	357,8	172,4	322,8	3,6	343,6	785,3	1	872,242
Faixa 5	ATSR 3	762,2	762,1	-587,7	377,7	186,1	268,6	3,8	367,7	109,4	1	933,798
Faixa 5	ATSR 4	761,9	761,9	127,1	453,1	221,6	317,6	453,4	4,7	251,6	1	99,187
Faixa 5	ATSR 5	761,9	761,9	415,3	496,8	233,7	385	497,1	528,3	323,6	1	996,516
Faixa 5	ATSR 6	762,1	762,1	633,8	540,4	242,5	439,4	539	594,9	375,3	1	997,677
Faixa 5	ATSR 7	761,7	761,7	11,2	840,3	261,8	811,7	756,4	961,2	470,8	1	1
Faixa 5	ATSR 8	760,7	760,7	138,7	134,8	272,5	150,3	10,4	151,4	504,4	1	1
Faixa 5	ATSR 9	758,6	758,5	158,5	190,8	281,3	275,2	142,1	22,7	512,4	1	1