



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA

RICARDO BRUM

**TÉCNICAS DE AQUISIÇÃO E REPRODUÇÃO
BIAURICULAR PARA AVALIAÇÃO DE QUALIDADE
SONORA DE SISTEMAS DE REFRIGERAÇÃO
DOMÉSTICOS**

FLORIANÓPOLIS

2025

Ricardo Brum

**TÉCNICAS DE AQUISIÇÃO E REPRODUÇÃO BIAURICULAR PARA
AVALIAÇÃO DE QUALIDADE SONORA DE SISTEMAS DE
REFRIGERAÇÃO DOMÉSTICOS**

Tese submetida ao Programa de Pós-Graduação
em Engenharia Mecânica da Universidade Fe-
deral de Santa Catarina como requisito parcial
para a obtenção do Grau de Doutor em Enge-
nharia Mecânica.

Orientador: Prof. Arcanjo Lenzi, Ph.D.

Florianópolis

2025

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.
Dados inseridos pelo próprio autor.

Brum, Ricardo

Técnicas de aquisição e reprodução biauricular para
avaliação de qualidade sonora de sistemas de refrigeração
domésticos / Ricardo Brum ; orientador, Arcanjo Lenzi,
2025.

159 p.

Tese (doutorado) - Universidade Federal de Santa
Catarina, Centro Tecnológico, Programa de Pós-Graduação em
Engenharia Mecânica, Florianópolis, 2025.

Inclui referências.

1. Engenharia Mecânica. 2. Qualidade sonora. 3. Gravação
biauricular. 4. Psicoacústica. I. Lenzi, Arcanjo. II.
Universidade Federal de Santa Catarina. Programa de Pós
Graduação em Engenharia Mecânica. III. Título.

Ricardo Brum

Técnicas de aquisição e reprodução binauricular para avaliação de qualidade sonora de sistemas de refrigeração domésticos

O presente trabalho em nível de doutorado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Roberto Aizik Tenenbaum, Dr. Eng.
Universidade Federal de Santa Maria

Prof. Stephan Paul, Dr. Eng.
Universidade Federal de Santa Catarina

Prof. Erasmo Felipe Vergara, Dr. Eng.
Universidade Federal de Santa Catarina

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de Doutor em Engenharia Mecânica.



Documento assinado digitalmente
Henrique Simas
Data: 13/11/2025 18:21:23-0300
CPF: ***.194.639-**
Verifique as assinaturas em <https://v.ufsc.br>

Prof. Henrique Simas, Dr. Eng.
Coordenador do Programa



Documento assinado digitalmente
Arcanjo Lenzi
Data: 13/11/2025 14:43:01-0300
CPF: ***.997.669-**
Verifique as assinaturas em <https://v.ufsc.br>

Prof. Arcanjo Lenzi, Ph.D.
Orientador

Florianópolis, 01 de setembro de 2025.

A todos os cientistas que dedicaram o tempo de sua vida para incrementar o conhecimento humano. A todos os professores que não medem esforços para perpetuar o conhecimento.

AGRADECIMENTOS

AOS PROFESSORES...

Primeiramente, ao meu orientador, Professor Arcanjo. Agradeço imensamente pela orientação precisa, pelo acolhimento e pela confiança depositados em mim durante toda essa caminhada.

Sua imensa qualidade humana, sua preocupação genuína com o bem-estar de cada aluno e sua capacidade de ser um motivador incansável foram fundamentais. Obrigado por valorizar não apenas o trabalho, mas a pessoa por trás dele.

Chefe, muito obrigado por tudo!

Ao professor Roberto pelas contribuições como avaliador e relator do trabalho e aos professores Stephan e Vergara pelas contribuições e ensinamentos durante minha caminhada.

Estendo meus agradecimentos aos demais professores que, de alguma forma, marcaram minha trajetória.

Meu reconhecimento mais profundo, contudo, é reservado a três pilares em minha vida, minhas professoras e apoiadoras incondicionais: minha mãe Simone, minha tia Ana e minha vó Jalva. Os valores que vocês me transmitiram são o alicerce de grande parte do meu sucesso.

AOS COLEGAS...

A todos os colegas do LVA, em especial ao Zinho e ao Jonas por toda a parceria e suporte durante tempos críticos. Todos os conselhos e apoio de vocês proporcionam uma amizade infindável.

Ao Henrique e ao Gabriel pelas ideias trocadas durante o desenvolvimento deste trabalho, me sinto privilegiado em poder contar com a competência e disponibilidade de vocês.

Meu agradecimento especial ao Lucian pelo suporte, e por todo seu empenho e competência no projeto da Petrobras durante a finalização da tese..

Também ao Léo e Robson por toda a ajuda com os experimentos e por me acolher no “senado” durante momentos críticos. Precisamos de muitas paletas de ovelha para comemorar!

Aos bolsistas do laboratório, em especial ao Bruno, Murilo e Vitor pela parceria, empenho e vontade de aprender. Também ao Lucas por ajudar na condução dos testes subjetivos e principalmente à Taiana por todo o empenho no desenvolvimento das interfaces gráficas.

Ao Rafa, o Coiote, pelo suporte técnico e ensinamentos proporcionados durante minha caminhada no LVA.

Ao Olavo por sua competência, humildade e disponibilidade para ajudar. Vou sentir saudades!

Ao Gil pela confiança, parceria, suporte e por todos os fins de semana no laboratório. Viva a balbúrdia!

Também não posso deixar de citar o Thiago pelas ideias trocadas, ao Cauã pelo seu empenho no projeto Embraco e todos os outros colegas do LVA.

AO PATROCINADOR

Agradeço a toda a equipe da célula de acústica da Embraco, em especial ao Mikio que foi um dos grandes idealizadores e motivadores deste trabalho. Serei eternamente grato por todo o aprendizado e crescimento que a empresa me proporcionou.

À FAMÍLIA

Agradeço à minha sogra Guacira e ao meu cunhado Luís por estarem sempre com a porta aberta e disponíveis para ajudar. Ao meu pai, um mestre artesão, por me ensinar que a busca pela perfeição transforma qualquer ofício em arte, não importando a área de atuação. Aos meus tios Vinícius e Tuny pelo apoio e ensinamentos proporcionados desde a infância.

À Lúcia, minha companheira de vida e de jornada. Agradeço por todo o amor, e em especial por sua imensa compreensão, paciência e suporte incondicional durante os desafios desta reta final do doutorado.

Sua presença me enche de motivação, entusiasmo e de uma felicidade que meras palavras jamais conseguiriam descrever. Muito obrigado por estar sempre ao meu lado e por ser parte essencial da minha vida!

IN MEMORIAM

Aos que já partiram para a eternidade. Aos meus avós, Vô Alemão e Vó Marlene e Vô Luís por todos os ensinamentos e valores transmitidos.

Ao meu sogro Carlos, que nos deixou durante a pandemia. Recordo com imenso carinho da parceria que tivemos e serei sempre grato por seu altruísmo.

Embora não estejam mais fisicamente conosco, sempre acreditarei que suas atitudes de caráter, bondade e seus legados viverão em cada um de nós que tivemos o privilégio de conviver com vocês.

Por fim, agradeço a todos os demais que de alguma forma contribuíram para o desenvolvimento deste trabalho.

“Construímos muros demais e pontes de menos.”

(Frase atribuída a Isaac Newton)

RESUMO

Este trabalho apresenta uma investigação sobre técnicas de aquisição e reprodução biauricular aplicadas à avaliação de qualidade sonora em sistemas de refrigeração domésticos. O estudo teve como principal objetivo desenvolver e validar um método capaz de quantificar diferenças perceptíveis entre sinais gravados por microfones convencionais e por simuladores de cabeça e torso (HATS). Inicialmente, foram conduzidos experimentos em câmara semi-anecoica com o objetivo de caracterizar a variabilidade dos sinais acústicos provenientes de fontes mecânicas e eletroacústicas sob diferentes condições operacionais e de reposicionamento dos receptores. As métricas psicoacústicas de *loudness*, *sharpness* e *roughness* foram analisadas de forma a quantificar a estabilidade espectral e a sensibilidade dessas grandezas às variações de montagem e funcionamento dos equipamentos. Os resultados mostraram que, dentro da faixa de operação típica, as diferenças observadas para essas métricas permanecem abaixo dos limiares diferenciais reportados na literatura, indicando que as variações introduzidas por reposicionamento ou regime de operação não são significativas e que os sinais mantêm consistência espectral suficiente para uso em análises comparativas posteriores. Na etapa seguinte, foi implementado um método de correção espectral ponderada, no qual o fator de ajuste α foi controlado por meio do método PEST (Estimativa de parâmetros por teste sequencial) para determinar o fator crítico (α_{crit}), associado ao limiar de discriminação de cada sujeito. A partir desses valores críticos, calcularam-se as diferenças entre o sinal corrigido e o sinal de referência para as métricas de *loudness* (ΔN), *sharpness* (ΔS) e *roughness* (ΔR), bem como para a MCD (*Mel Cepstral Distance*). Esta última, por se tratar de uma métrica de comparação espectral, permitiu estimar uma Diferença de Limiar Observável (DLO) e definir uma zona de confusão associada ao limite de discriminação perceptível entre os sinais. A seguir, foi conduzido um teste MUSHRA (Múltiplos Estímulos com Referência Oculta e Âncora) a fim de validar experimentalmente a correspondência entre as métricas objetivas e as respostas perceptivas do júri. Os resultados indicaram forte correlação entre os escores subjetivos e os valores de MCD, com coeficiente mínimo de determinação $R^2 = 0,96$, confirmando a robustez da métrica como estimador perceptivo de diferença espectral. Por fim, as medições em ambiente típico permitiram avaliar o comportamento dos diferentes HATS e dos critérios de equalização dos fones de ouvido em condições de escuta mais realistas. Verificou-se que a simetria da fonte sonora e o tipo de acoplamento dos fones exercem influência direta sobre a reprodutibilidade das medições e sobre a estabilidade espectral obtida. Conclui-se que a metodologia desenvolvida constitui uma ferramenta eficaz para quantificação e correção de diferenças perceptivas entre sinais sonoros, podendo contribuir significativamente para o avanço das padronizações de medição e reprodução biauricular aplicadas à avaliação de qualidade sonora.

Palavras chave: qualidade sonora; psicoacústica; gravação biauricular.

ABSTRACT

This work presents a investigation on binaural acquisition and reproduction techniques applied to the evaluation of sound quality in household refrigeration systems. The main objective of the study was to develop and validate a method capable of quantifying perceptible differences between signals recorded using conventional microphones and Head and Torso Simulators (HATS). Initially, experiments were carried out in a semi-anechoic chamber in order to characterize the variability of acoustic signals emitted by mechanical and electroacoustic sources under different operating conditions and receiver repositioning configurations. The psychoacoustic metrics of *loudness*, *sharpness*, and *roughness* were analyzed to quantify spectral stability and the sensitivity of these quantities to variations in setup and equipment operation. The results showed that, within the typical operating range, the observed differences for these metrics remained below the differential thresholds reported in the literature, indicating that the variations introduced by repositioning or operating regime were not perceptually significant and that the signals preserved sufficient spectral consistency for use in subsequent comparative analyses. In the following stage, a weighted spectral correction method was implemented, in which the adjustment factor α was controlled through the PEST method (*Parameter Estimation by Sequential Testing*) to determine the critical factor α_{crit} associated with the discrimination threshold of each subject. From these critical values, the differences between the corrected and reference signals were calculated for the *loudness* (ΔN), *sharpness* (ΔS), and *roughness* (ΔR) metrics, as well as for the Mel Cepstral Distance (MCD). The latter, being a spectral comparison metric, allowed for the estimation of a Detectable Limit of Observation (DLO) and the definition of a confusion zone associated with the perceptual discrimination limit between the signals. Subsequently, a MUSHRA test (*Multiple Stimuli with Hidden Reference and Anchor*) was conducted to experimentally validate the correspondence between objective metrics and the perceptual responses of the jury. The results indicated a strong correlation between subjective scores and MCD values, with a minimum determination coefficient of $R^2 = 0.96$, confirming the robustness of this metric as a perceptual estimator of spectral difference. Finally, measurements in a typical acoustic environment made it possible to assess the behavior of different HATS and headphone equalization criteria under more realistic listening conditions. It was found that the symmetry of the sound source and the type of headphone coupling exert a direct influence on the reproducibility of the measurements and on the spectral stability obtained. It is concluded that the methodology developed constitutes an effective tool for the quantification and correction of perceptual differences between sound signals and may contribute significantly to the advancement of measurement and binaural reproduction standardization applied to sound quality evaluation.

Keywords: sound quality; psychoacoustics; binaural recording.

LISTA DE ILUSTRAÇÕES

Figura 1 – Ilustração de um SLIT com uma entrada e uma saída.	8
Figura 2 – Ilustração de um sistema em cascata com seu inverso.	16
Figura 3 – Exemplo de filtro de correção de fones de ouvido	19
Figura 4 – Ilustração genérica de um sistema de gravação biauricular e sua calibração.	25
Figura 5 – Reprodução biauricular com filtros de cancelamento de <i>crosstalk</i>	28
Figura 6 – Exemplo de interface gráfica sugerido pela norma.	42
Figura 7 – Diagrama esquemático da medição da HRTF.	45
Figura 8 – Sinal de excitação utilizado nas medições.	46
Figura 9 – Bancada de medição da HRTF	47
Figura 10 – RI e FRF entre fonte-receptor com janelamento	48
Figura 11 – Comparação das HRTFs utilizando as duas fontes sonoras.	48
Figura 12 – Fontes sonoras posicionadas na câmara semi-anecoica	49
Figura 13 – Comparação entre as medições com HATS e microfone.	51
Figura 14 – Espectros SPL HATS e microfone para sinal de geladeira	53
Figura 15 – Espectros SPL HATS e microfone para sinal de compressor	54
Figura 16 – Análise das medições do monitor de áudio	55
Figura 17 – Espectrograma do sinal de refrigerador	57
Figura 18 – Espectrograma do sinal de compressor - velocidade única	57
Figura 19 – Espectrograma do sinal de compressor - 2400 rpm	58
Figura 20 – Espectrograma do sinal de compressor - 3000 rpm	58
Figura 21 – Espectrograma do sinal de compressor - 4500 rpm	59
Figura 22 – Distribuição dos parâmetros: monitor de áudio compressor 2400 rpm	62
Figura 23 – Distribuição dos parâmetros: monitor de áudio compressor 3000 rpm	62
Figura 24 – Distribuição dos parâmetros: monitor de áudio compressor 4500 rpm	63
Figura 25 – Distribuição dos parâmetros - refrigerador	64
Figura 26 – Distribuição dos parâmetros - compressor	65
Figura 27 – Processo de correção do sinal de teste em função do sinal de referência	67
Figura 28 – Análise de convergência para o fator de correção	69
Figura 29 – Fluxograma das etapas do método PEST para estimativa de limiar.	71
Figura 30 – Interface gráfica utilizada para o teste de juri	72
Figura 31 – Espectros da FRF do sistema de reprodução	73

Figura 32 – Teste de Dunn para α_{crit}	75
Figura 33 – Limiares de discriminação e respectivos parâmetros objetivos	76
Figura 34 – Teste de Dunn para MCD	78
Figura 35 – Interface implementada para o teste MUSHRA.	81
Figura 36 – Resultados MUSHRA - refrigerador	83
Figura 37 – Resultados MUSHRA - compressor	84
Figura 38 – Resultados MUSHRA - compressor 2400 rpm	84
Figura 39 – Resultados MUSHRA - compressor 3000 rpm	85
Figura 40 – Resultados MUSHRA - compressor 4500 rpm	85
Figura 41 – Distribuição dos escores MUSHRA e correlação com os valores de MCD para as fontes mecânicas.	86
Figura 42 – Distribuição dos escores MUSHRA e correlação com os valores de MCD para os sinais de fonte eletroacústica.	87
Figura 43 – Orelha prototipada com o microfone MEMS na entrada do canal auditivo.	91
Figura 44 – Critérios de abertura dos fones de ouvido	92
Figura 45 – Espectros de magnitude médios da FRF dos fones de ouvido abertos, considerando 10 reposicionamentos.	93
Figura 46 – Espectros de magnitude médios da FRF dos fones de ouvido fechados, considerando 10 reposicionamentos.	94
Figura 47 – Atraso de Grupo médio para um fone aberto (HD 58 X) e um fechado (HD 206).	94
Figura 48 – Desvio padrão dos espectros de magnitude da FRF	95
Figura 49 – FRF e desvio-padrão do fone HD 58 X com os 4 HATS	97
Figura 50 – HATS Kemar posicionado no local de medição.	98
Figura 51 – Espectro de magnitude das FRFs entre fonte e receptor utilizando os 3 HATS e as duas fontes sonoras.	99
Figura 52 – Espectro de magnitude das FRFs entre fonte e receptor utilizando os 3 HATS e as duas fontes sonoras com a correção do fone de ouvido.	100
Figura 53 – Variabilidade de reposicionamento do MCD para os 3 HATS e 2 fontes utilizando o sinal do compressor a 4500 rpm.	101
Figura 54 – Comparação da MCD entre HATS	102
Figura 55 – HATS posicionados nos pontos de medição.	104
Figura 56 – Tempo de reverberação da sala em bandas de terço de oitava.	105

Figura 57 – Mapa de medição na sala de reuniões.	106
Figura 58 – Sistema de áudio instalado na sala.	107
Figura 59 – Resposta em frequência da sala utilizando os 3 HATS	108
Figura 60 – Resposta em frequência da sala utilizando os 3 HATS	109
Figura 61 – FRFs biauriculares (Posição 2) com o sistema de áudio da sala	110
Figura 62 – Análise de convergência das métricas psicoacústicas em função do tamanho do bloco para o sinal de compressor de velocidade única.	130
Figura 63 – FRFs biauriculares para a fonte Genelec	131
Figura 64 – FRFs biauriculares para a fonte Hinor	132
Figura 65 – FRFs biauriculares para a fonte Genelec	132

LISTA DE TABELAS

Tabela 1 – Shapiro-Wilk e Mann-Whitney para compressor 2400 rpm	60
Tabela 2 – Shapiro-Wilk e Mann-Whitney para compressor 3000 rpm	60
Tabela 3 – Shapiro-Wilk e Mann-Whitney para compressor 4500 rpm	61
Tabela 4 – Shapiro-Wilk e Mann-Whitney para refrigerador	63
Tabela 5 – Shapiro-Wilk e Mann-Whitney para compressor	64
Tabela 6 – Pares de sinal HATS-microfones utilizados no teste de juri.	66
Tabela 7 – Shapiro-Wilk para α_{crit} de todos os sujeitos	75
Tabela 8 – Medianas dos parâmetros mostrados na Figura 33.	77
Tabela 9 – Médias interauriculares da MCD entre HATS	111
Tabela 10 – Diferenças interauriculares da MCD entre os HATS	113
Tabela 11 – MCD para os sinais de compressor convoluídos com as BRIRs	131

LISTA DE SÍMBOLOS

Símbolos latinos

α	Fator de correção espectral
α_{crit}	Fator de correção espectral crítico - limiar de discriminação
ΔN	Diferença de <i>Loudness</i> entre sinais de referência e teste em sone
ΔR	Diferença de <i>Roughness</i> entre sinais de referência e teste em asper
ΔS	Diferença de <i>Sharpness</i> entre sinais de referência e teste em acum
$\delta(t)$	Distribuição Delta de Dirac
$D_{\text{IS}}(P(\omega), \hat{P}(\omega))$	Distância Itakura-Saito (IS)
D_{LS}	Distância Log-Espectral (DLE)
D_{EUC}	Distância Espectral Euclidiana
$H(f)$	Função resposta em frequência de um sistema
$h(t)$	Função resposta impulsiva de um sistema
N	<i>Loudness</i> em sone
$P_1(f)$	Espectro de pressão sonora no centro da cabeça com essa ausente
$P_2(f)$	Espectro de pressão sonora na entrada do canal auditivo bloqueado sem fone
$P_3(f)$	Espectro de pressão sonora na entrada do canal auditivo aberto sem fone
$P_4(f)$	Espectro de pressão sonora na membrana timpânica sem fone
$P_5(f)$	Espectro de pressão sonora na entrada do canal auditivo bloqueado com fone
$P_6(f)$	Espectro de pressão sonora na entrada do canal auditivo aberto com fone
$P_7(f)$	Espectro de pressão sonora na membrana timpânica com fone
R	<i>Roughness</i> em asper
R^2	Coefficiente de Correlação de Pearson

S *Sharpness* em acum

$x(t)$ Resposta do sistema para $x(t)$ (saída)

$x(t)$ Sinal de excitação de um sistema (entrada)

$y(t)$ Resposta do sistema para $x(t)$ (saída)

MCD *Mel Cepstral distance*

SIGLAS

BEM Método de Elementos de Contorno. [23](#)

BPO Bandas por oitava. [69](#), [70](#)

BRIR Resposta Impulsiva Biauricular da Sala. [2](#), [3](#), [104](#), [107](#), [110](#)

DLO Diferença de Limiar Observável. [3](#), [4](#), [59](#), [61](#), [64](#), [65](#), [70](#), [77](#), [78](#), [81](#), [88](#), [114](#), [115](#), [117](#)

DUT Dispositivo sob Teste. [40](#)

FEC Acoplamento Equivalente de Campo Livre. [26](#), [100](#)

FRF Função Resposta em Frequência. [7](#), [9](#), [10](#), [55](#), [73](#), [99](#), [100](#), [108](#), [109](#)

GDL Atraso de Grupo. [7](#), [93](#)

HATS Simulador de Torso e Cabeça. [3–5](#), [23](#), [27](#), [28](#), [30](#), [36](#), [45](#), [49](#), [56](#), [60](#), [66](#), [73](#), [77](#), [88](#),
[90](#), [98–104](#), [107](#), [109–112](#), [114–116](#)

HRIR Resposta Impulsiva Relacionada à Cabeça. [3](#), [22](#), [45](#), [88](#), [90](#)

HRTF Resposta em Frequência Relacionada à Cabeça. [6](#), [21](#), [36](#), [45](#), [88](#), [110](#), [114](#)

HSC Critério de Seleção de Fones de Ouvido. [4](#), [26](#), [91](#)

IDFT Transformada Inversa Discreta de Fourier. [22](#)

LVA Laboratório de Vibrações e Acústica. [72](#), [90](#), [104](#)

MCD *Mel Cepstral Distance*. [4](#), [22](#), [32](#), [33](#), [69](#), [70](#), [75](#), [78](#), [79](#), [82](#), [86](#), [89](#), [90](#), [101](#), [102](#),
[110–112](#), [114–117](#)

MUSHRA Múltiplos Estímulos com Referência Oculta e Âncora. [4](#), [40](#), [66](#), [80–82](#), [86](#),
[89](#), [115](#)

NPS Nível de Pressão Sonora. [1](#)

NWS Nível de Potência Sonora. [1](#), [2](#)

OI Índice de Oclusão. [26](#)

PDR Razão de divisão de pressões. [4](#), [25](#), [26](#), [91](#)

PEST Estimativa de Parâmetros por Teste Sequencial. [4](#), [43](#), [44](#), [66](#), [70](#), [80](#), [114](#)

RI Resposta em Impulsiva. [7](#), [8](#), [11](#)

RIR Resposta Impulsiva da Sala. [2](#), [8](#)

SLIT Sistema Linear e Invariante no Tempo. [7](#), [9](#), [11](#), [13](#), [117](#)

SNR Relação Sinal Ruído. [8](#), [88](#)

UFSC Universidade Federal de Santa Catarina. [72](#)

SUMÁRIO

1	INTRODUÇÃO	1
1.1	OBJETIVOS	3
1.1.1	Objetivo Geral	3
1.1.2	Objetivos Específicos	3
1.2	CONTRIBUIÇÕES	4
1.3	ESTRUTURA DO TRABALHO	5
2	REFERENCIAL TEÓRICO	6
2.1	TEORIA GERAL DOS SISTEMAS LINEARES E INVARIANTES NO TEMPO	7
2.1.1	Filtros lineares: definições e topologias convencionais	11
2.1.2	Filtros com espectro de magnitude arbitrária	13
2.1.2.1	Método de Janelamento para Projeto de Filtros FIR com Espectro de Magni- tude Arbitrária	14
2.1.2.2	Obtenção da Representação de Fase Mínima via Transformada de Hilbert do Espectro de Magnitude	16
2.1.2.3	Método de Yule-Walker para Projeto de Filtros IIR com Espectro de Magnitude Arbitrária	18
2.2	TECNOLOGIA BIAURICULAR	20
2.2.1	Resposta em frequência relacionada à cabeça	21
2.2.2	Equalização de Sistemas de Reprodução Biauricular	23
2.2.3	Apresentação de sinais utilizando alto-falantes	27
2.2.4	Normas e Padrões para Simuladores de Cabeça e Torso	28
2.2.4.1	Padrões para Audiologia e Eletroacústica (Série IEC 60318)	29
2.2.4.2	Especificações para Medições Acústicas In-Situ (ANSI S3.36)	29
2.2.4.3	Outros referenciais normativos	29
2.2.4.4	Lacunas para Qualidade Sonora	30
2.3	FUNDAMENTOS PSICOACÚSTICOS DAS SENSações SONORAS E PARÂMETROS DE AVALIAÇÃO	30
2.4	MÉTRICAS DE COMPARAÇÃO DE SINAIS	32

2.4.1	Distância espectral euclidiana	32
2.4.2	Distância Itakura–Saito	32
2.4.3	Distância Log-Espectral	33
2.4.4	<i>Mel-Cepstrum Distance</i>	33
2.5	QUALIDADE SONORA DE SISTEMAS DE REFRIGERAÇÃO	35
2.6	MÉTODOS SUBJETIVOS PARA AVALIAÇÃO DE DIFERENÇA EN- TRE SINAIS	37
2.6.1	Teste AB	37
2.6.2	Teste ABX	37
2.6.3	Escolha Forçada de 3 Alternativas	39
2.6.4	Norma ITU-R BS.1116-3 (2015)	40
2.6.5	Teste MUSHRA	40
2.6.6	Método de Ajuste	41
2.7	ESTIMATIVA DE PARÂMETROS POR TESTE SEQUENCIAL . . .	43
3	COMPARAÇÃO DE SINAIS DE SISTEMAS DE REFRIGE- RAÇÃO OBTIDOS COM HATS E MICROFONE EM CÂ- MARA SEMI-ANECOICA	45
3.1	MEDIÇÃO DA HRTF	45
3.2	FONTES SONORAS	49
3.2.1	Compressor	49
3.2.2	Refrigerador	50
3.2.3	Monitor de áudio	50
3.3	RESULTADOS OBTIDOS	51
3.3.1	Comparação espectral entre as gravações obtidas com HATS e microfone.	52
3.4	ANÁLISE DE VARIABILIDADE DAS MÉTRICAS PSICOACÚSTICAS	56
3.4.1	Resultados para monitor de áudio (fonte eletroacústica)	59
3.4.2	Resultados para fontes mecânicas	63
3.5	AVALIAÇÃO SUBJETIVA	66
3.5.1	Amostra sonora	66
3.5.2	Correção espectral ponderada	66

3.5.3	Implementação da PEST para controlar o fator de correção espectral	70
3.5.4	Implementação da plataforma digital para o teste e participantes	71
3.5.5	Ambiente de teste	72
3.5.6	Calibração do sistema de reprodução dos sinais	72
3.5.7	Resultados do teste de juri	74
3.6	TESTE MUSHRA PARA AVALIAÇÃO DA DIFERENÇA PERCEP- TIVA ENTRE OS SINAIS	80
3.6.1	Implementação da interface de teste	80
3.6.2	Amostra sonora para o teste piloto	81
3.6.3	Resultados do teste piloto	82
3.6.4	Resultados do teste final	86
3.7	REVISÃO E SÍNTESE DO CAPÍTULO	88
4	COMPARAÇÃO DE SINAIS OBTIDOS POR DIFERENTES HATS E CARACTERIZAÇÃO DE FONES DE OUVIDO . . .	90
4.1	MEDIÇÃO DOS PARÂMETROS DE FONES DE OUVIDO	90
4.1.1	Critérios de Abertura (PDR e HSC)	91
4.1.1.1	FRF dos fones de ouvido	92
4.1.2	Espectros de atraso de grupo	93
4.1.3	Variabilidade da Magnitude da FRF	95
4.1.4	Síntese e Justificativa da Escolha do Fone	96
4.1.5	Comparativo da FRF do fone utilizando os diferentes HATSs .	96
4.2	MEDIÇÕES EM CÂMARA SEMI-ANECOICA COM OS 3 HATS . . .	98
4.2.1	FRFs entre fontes sonoras e HATS obtidas no experimento . .	99
4.2.2	Comparação de Sinais Convoluídos com as Respostas Impulsivas	101
4.3	GRAVAÇÕES NA SALA DE REUNIÕES	104
4.3.1	Características do ambiente	104
4.3.2	FRFs entre fontes sonoras e HATS obtidas no experimento . .	107
4.3.3	Análise da MCD entre áudios convoluídos	110
4.3.3.1	Análise da Média Interauricular na Comparação entre HATS através da <i>Mel</i> <i>Cepstral Distance</i> (MCD)	111

4.3.3.2	Análise da Assimetria Interauricular na Comparação entre HATS através da MCD	112
5	CONSIDERAÇÕES FINAIS	114
5.1	SUGESTÕES PARA TRABALHOS FUTUROS	116
	Bibliografia	118

APÊNDICES	129
------------------	------------

APÊNDICE A – ANÁLISE DE CONVERGÊNCIA DAS MÉ- TRICAS EM FUNÇÃO DO TAMANHO DO BLOCO	130
--	------------

APÊNDICE B – DADOS OBTIDOS NA SALA DE REUNIÕES	131
---	------------

ANEXOS	133
---------------	------------

ANEXO A – FICHA TÉCNICA DO MICROFONE UTILI- ZADO NA RESTAURAÇÃO DO HATS CORTEX	134
---	------------

1 INTRODUÇÃO

A relevância dos estudos de qualidade sonora tem se expandido significativamente em diversas áreas, impulsionada pela crescente demanda por experiências auditivas aprimoradas e pela necessidade de compreender como os indivíduos percebem o som. As ferramentas de avaliação subjetiva de sinais sonoros são essenciais no desenvolvimento e otimização de sistemas de reprodução sonora e algoritmos de compressão de dados. Como exemplo, pode-se citar a bem-sucedida tecnologia MP3, que demonstrou a possibilidade de reduzir drasticamente o volume de dados. Essa tecnologia proporciona uma percepção sonora muitas vezes considerada indistinguível ao comparar áudios compactados com gravações de referência não comprimidas.

Além disso, as ferramentas de psicoacústica desempenham um papel muito importante em áreas como audiologia e acústica musical, a partir dos estudos que buscam compreender a complexidade da audição humana e os mecanismos envolvidos na apreciação sonora. Outro campo em expansão é a investigação da distinção perceptual entre ambientes sonoros reais e virtuais, especialmente em contextos de simulação acústica imersiva. Estudos como o de Brinkmann et al. (2019) exemplificam esse cenário ao empregar protocolos comparativos baseados em *round robin* para avaliar a percepção auditiva em diferentes configurações de sala simulada, destacando a importância das avaliações subjetivas em ambientes controlados.

No contexto de produtos de consumo como os sistemas de refrigeração domésticos, sabe-se que parâmetros convencionais como **Nível de Potência Sonora (NWS)** e **Nível de Pressão Sonora (NPS)** não são suficientes para caracterizar todos os aspectos sonoros de uma fonte sonora. Dessa forma, a simples redução da potência sonora radiada não significa necessariamente uma melhora da qualidade sonora percebida pelos consumidores.

Nesse sentido, as técnicas utilizadas para avaliação de qualidade sonora também são muito importantes na caracterização dos aspectos sonoros nesse tipo de dispositivo. Portanto, os fundamentos da psicoacústica são essenciais na caracterização sonora de produtos, visando otimizar características sonoras para melhorar o conforto dos usuários.

Apesar das limitações do **NWS** em representar adequadamente todos os aspectos perceptivos de sinais sonoros, a normatização para sua medição é consolidada e amplamente adotada na indústria. As normas ISO 3745 (2012) e ISO 3741 (2010) especificam, respectivamente, procedimentos em câmaras anecoicas ou semi-anecoicas e em salas reverberantes.

Essas normas definem requisitos de ambiente, posicionamento de microfones e da fonte, condições operacionais, correções de ruído residual e caracterização de incertezas. Esses documentos garantem rastreabilidade e comparabilidade entre laboratórios, viabilizando ensaios regulatórios e *benchmarking* (avaliação comparativa) de produtos com base em grandezas físicas bem definidas, como espectro de [NWS](#) e seu respectivo global (L_w).

No entanto, o panorama é distinto para os estudos de qualidade sonora. Nesse caso, existem normas que especificam os modelos para o cálculo de parâmetros psicoacústicos como o *loudness* ([International Organization for Standardization, 2017](#)), *sharpness* ([Deutsches Institut für Normung, 2009](#)) e *roughness* que pode ser obtido a partir de modelos como o de Zwicker ([2013](#)) ou Daniel-Weber ([1997](#)).

Porém, ainda falta padronização quanto à forma de registrar os sinais para avaliações de qualidade sonora e também em relação à apresentação desses sinais em testes de júri. Dessa forma, surgem questionamentos acerca de qual ambiente deve ser utilizado para realizar as medições, como comparar produtos caracterizados em diferentes condições de medição e qual compromisso adotar entre representatividade e rastreabilidade.

Um ambiente típico aproxima as condições reais de uso e preserva interações fonte-sala relevantes à percepção, porém é difícil de controlar e de reproduzir entre laboratórios. A câmara reverberante atende à medição de potência sonora, mas não é adequada para todos os atributos de qualidade sonora, pois tende a reduzir profundidade de modulação, além de não contemplar características de direcionalidade da fonte sonora sob estudo.

Já as câmaras anecoicas ou semi-anecoicas preservam os aspectos de direcionalidade e a modulação do sinal radiado pela fonte. Também possibilitam simular diferentes contextos através da convolução com uma [Resposta Impulsiva da Sala \(RIR\)](#)¹ ou [Resposta Impulsiva Biauricular da Sala \(BRIR\)](#)² associadas a posições fonte-receptor do ambiente de interesse. Assim, quando o objetivo é comparar soluções e isolar o efeito da técnica de aquisição e reprodução, esses ambientes proporcionam melhor rastreabilidade, enquanto a aurilização posterior pode ser empregada para simular as condições do uso representativas em ambientes típicos de utilização.

Nessa direção, estudos de Dal Moro ([2019](#)) e Brum ([2020](#)) mostram que a forma de gravação impacta diretamente na avaliação de qualidade sonora em sistemas de refrigeração: variações de montagem da fonte sonora em câmara semi-anecoica alteram a radiação da

¹ Acrônimo do inglês para *Room Impulse Response*.

² Acrônimo do inglês para *Binaural Room Impulse Response*.

fonte sonora e os respectivos parâmetros psicoacústicos, e há diferenças entre [BRIR](#) medidas e simuladas com efeito na aurilização. Brum (2020) também indica que escolhas da fonte sonora na medição da [BRIR](#) afetam o conteúdo espectral, principalmente devido aos aspectos de direcionalidade. Já as respostas impulsivas obtidas a partir de medições com microfone e convoluídas com [Resposta Impulsiva Relacionada à Cabeça \(HRIR\)](#), embora comprometam a espacialidade do ambiente, preservam características espectrais relevantes e mantêm boa concordância de parâmetros psicoacústicos quando comparados ao [Simulador de Torso e Cabeça \(HATS\)](#).³

Anteriormente, You e Jin (2008) determinaram a [Diferença de Limiar Observável \(DLO\)](#)⁴ para *loudness*, *sharpness*, *roughness* e *fluctuation strength* em sinais de refrigeradores domésticos. Esses valores funcionam como “régua” perceptual para julgar se variações nesses parâmetros, inclusive as decorrentes de diferentes configurações de gravação, aurilização e reprodução discutidas anteriormente, sendo de fato relevantes ao ouvinte.⁵

1.1 OBJETIVOS

Apesar desses avanços, vale ressaltar a área de qualidade sonora para sistemas de refrigeração ainda carece de uma padronização robusta que permita caracterizar produtos no contexto da área. Baseado nisso, os objetivos deste trabalho são os que se seguem.

1.1.1 Objetivo Geral

Quantificar como a técnica biauricular de aquisição e reprodução, incluindo a escolha entre [HATS](#) e microfone com convolução [HRIR](#), o ambiente de medição e a calibração e acoplamento de fones alteram o sinal e o impacto sobre os parâmetros objetivos e julgamentos subjetivos na avaliação de qualidade sonora de sistemas de refrigeração, estabelecendo referências operacionais para comparação entre condições e equipamentos.

1.1.2 Objetivos Específicos

- Comparar gravações obtidas em câmara semi-anecoica por [HATS](#) e por microfone (sinal convoluído com [HRIR](#)), analisando diferenças espectrais e dos parâmetros psico-

³ Acrônimo do inglês para *Head and Torso Simulator*.

⁴ Do inglês: *Just Noticeable Difference*.

⁵ Mais detalhes serão apresentados no referencial teórico desta tese.

acústicos calculados, incluindo estudo de reposicionamento para fonte eletroacústica e variabilidade operacional para fontes mecânicas (compressor e refrigerador).

- Quantificar o **DLO** para *loudness*, *sharpness*, *roughness* e *Mel-Cepstrum Distance* **MCD** por meio da aplicação de teste subjetivo adaptativo baseado no método de **Estimativa de Parâmetros por Teste Sequencial (PEST)**⁶ com correção espectral ponderada.
- Comparar as **DLOs** obtidas com os valores de referência da literatura para validar o método desenvolvido;
- Avaliar diferenças perceptivas em múltiplos níveis de correção espectral por meio do teste **Múltiplos Estímulos com Referência Oculta e Âncora (MUSHRA)**⁷ e relacioná-las com os parâmetros objetivos, discutindo coerência e sensibilidade.
- Caracterizar o acoplamento e a calibração de fones usando **Razão de divisão de pressões (PDR)** e **Critério de Seleção de Fones de Ouvido (HSC)**, e investigar o efeito desses fatores nos sinais biauriculares e na comparabilidade entre sinais gravados com diferentes **HATSS**, utilizando parâmetros objetivos.
- Analisar sinais obtidos com três **HATSS** em câmara semi-anecoica e em sala de reuniões, incluindo posições simétricas e assimétricas, destacando impactos na qualidade sonora.

1.2 CONTRIBUIÇÕES

Este trabalho apresenta as seguintes contribuições:

- Proposição e implementação de um método adaptativo de correção espectral ponderada integrado ao **PEST** para estimativa do limiar de discriminação entre versões de um mesmo sinal.
- Desenvolvimento e aplicação de interfaces experimentais baseadas nos métodos **PEST** e **MUSHRA**, permitindo a condução de testes subjetivos para estimar diferenças perceptíveis em *loudness*, *sharpness*, *roughness* e **MCD** em sinais de sistemas de refrigeração.

⁶ Acrônimo do inglês para *Parameter Estimation by Sequential Test*.

⁷ Acrônimo do inglês para *Multi Stimulus test with Hidden Reference and Anchor*.

- Avaliação integrada do efeito da técnica de aquisição e reprodução: comparação entre gravações por HATS e por microfone com convolução HRIR, entre diferentes HATS e entre ambientes (semi-anecoico e sala), discutindo impactos espectrais/temporais e nos parâmetros psicoacústicos.
- Caracterização do acoplamento de fones por meio de PDR e HSC e aplicação desses critérios à calibração, demonstrando seu efeito na diferença entre sinais obtidos com diferentes HATS.
- Validação prática da MCD como métrica objetiva de dissimilaridade para o domínio estudado e estabelecimento de uma referência operacional (faixa típica no limiar de discriminação) útil para comparações entre condições e equipamentos.

1.3 ESTRUTURA DO TRABALHO

Para facilitar a compreensão do desenvolvimento deste trabalho, sua estrutura está organizada da seguinte forma:

- Capítulo 2: apresenta as principais bases teóricas necessárias para a realização deste trabalho;
- Capítulo 3: apresenta um estudo comparativo utilizando gravação com microfone e HATS em câmara semi anecoica;
- Capítulo 4: apresenta um estudo comparativo entre diferentes HATSs em câmara semi anecoica e ambiente típico bem como os efeitos da equalização dos fones de ouvido;
- Capítulo 5: apresenta as considerações finais, ressaltando as contribuições e sugestões para trabalhos futuros.

2 REFERENCIAL TEÓRICO

Este capítulo estabelece a base teórica essencial para o desenvolvimento do presente trabalho. Serão explorados os conceitos fundamentais que sustentam os métodos propostos e a análise dos resultados. A estrutura deste referencial teórico é organizada nos seguintes tópicos:

- **Teoria geral dos sistemas lineares e invariantes no tempo:** Serão revisados os princípios dos SLITs, com ênfase nas técnicas de projeto de filtros digitais capazes de moldar o espectro de magnitude de forma arbitrária, ferramentas cruciais para a manipulação e equalização de sinais.
- **Fundamentos de Tecnologia Biauricular e Acústica Virtual:** Esta seção abordará os princípios da tecnologia biauricular, explorando como a [Resposta em Frequência Relacionada à Cabeça \(HRTF\)](#) codifica a informação espacial e as técnicas para a síntese de ambientes acústicos virtuais.
- **Fundamentos de Psicoacústica e Parâmetros:** Serão apresentados conceitos da psicoacústica relevantes e os modelos utilizados para quantificar sensações auditivas (*loudness*, *sharpness*, *roughness*), os quais servem de base para a formulação de modelos de avaliação da qualidade sonora.
- **Qualidade sonora de sistemas de refrigeração:** Esta seção revisa os principais estudos desenvolvidos para avaliação de qualidade sonora de sistemas de refrigeração com ênfase em refrigeradores e compressores herméticos utilizados em aplicações domésticas.
- **Métodos em acústica subjetiva com ênfase em técnicas para avaliação de diferenças subjetivas entre sinais:** Por fim, serão discutidos os métodos empregados em testes de avaliação subjetiva em acústica, com foco em técnicas específicas para quantificar as diferenças subjetivas entre estímulos sonoros.

A exploração detalhada destes tópicos fornecerá as bases teóricas necessárias para compreender os métodos utilizados, a análise dos resultados obtidos e a discussão das suas implicações no contexto da acústica biauricular e da avaliação perceptiva.

2.1 TEORIA GERAL DOS SISTEMAS LINEARES E INVARIANTES NO TEMPO

Sistemas Lineares e Invariantes no Tempo (SLITs) constituem um arcabouço teórico fundamental para a análise e o processamento de sinais em diversas disciplinas da engenharia, incluindo a acústica. Um sistema é considerado linear se obedecer ao princípio da superposição, em que a resposta a uma combinação linear de entradas é igual à combinação linear das respostas individuais.

A invariância no tempo, por sua vez, é uma característica de sistemas em que um deslocamento temporal na entrada resulta em um idêntico deslocamento temporal na saída, sem alterar a forma da resposta. Essas propriedades simplificam significativamente a análise e a modelagem de sistemas, permitindo prever seu comportamento sob diferentes excitações. **SLITs** encontram aplicações em modelagem acústica de ambientes ([Beranek; Mellow, 2012](#); [Vorländer, 2020](#); [Cox; D’Antonio, 2020](#)), projeto de filtros para análise e processamento de áudio ([Oppenheim, 2010](#); [Oppenheim; Schafer, 2009](#); [Shin; Hammond, 2008](#); [Zölzer, 2011](#)) e em sistemas de comunicação ([Proakis; Salehi, 2007](#); [Lathi; Ding, 2018](#)).

Sabe-se que um **SLIT** é completamente caracterizado por sua **Resposta em Impulsiva (RI)**, denotada por $h(t)$ no domínio do tempo, e sua respectiva **Função Resposta em Frequência (FRF)**¹, $H(f)$ no domínio da frequência. A FRF é a Transformada de Fourier da RI, expressa por

$$H(f) = \mathcal{F}\{h(t)\} = \int_{-\infty}^{\infty} h(t)e^{-j2\pi ft} dt. \quad (2.1)$$

A **RI** é a resposta do sistema quando excitado por um impulso de Dirac, $\delta(t)$. Já a **FRF** proporciona uma indicação direta de como o sistema amplifica ou atenua diferentes componentes de frequência de um sinal de entrada e introduz deslocamentos de fase. A magnitude da FRF representa o ganho do sistema em cada frequência, enquanto a fase de $H(f)$ representa o deslocamento de fase. A derivada do espectro de fase em relação à frequência define o **Atraso de Grupo (GDL)**, que corresponde ao atraso efetivo dos componentes espectrais no domínio temporal.

A partir da **RI** e da **FRF**, é possível derivar diversos parâmetros importantes do

¹ Vários autores tratam os termos Função de Transferência (FT) e Função de Resposta em Frequência (FRF) como sinônimos. Contudo, [Oppenheim \(2010\)](#) define a FRF como o resultado da Transformada de Fourier aplicada à Resposta ao Impulso (RI) de um sistema, enquanto a FT é definida como o resultado da Transformada de Laplace sobre a mesma RI. Essa nomenclatura é adotada neste trabalho.

sistema, como a resposta a sinais arbitrários através da convolução no domínio do tempo

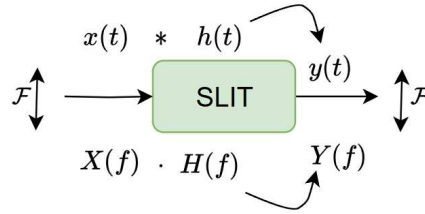
$$y(t) = x(t) * h(t), \quad (2.2)$$

ou da multiplicação no domínio da frequência

$$Y(f) = X(f)H(f), \quad (2.3)$$

conforme ilustrado na Figura 1, bem como medidas de linearidade de fase. No contexto da acústica de salas, o tempo de reverberação, Definição, Clareza dentre outros parâmetros também podem ser obtidos a partir da RIR (ISO, BSEN; STANDARD, BRITISH, 2009; Brandão, 2016; Kuttruff; Vorländer, 2024; Long, 2005).

Figura 1 – Ilustração de um SLIT com uma entrada e uma saída.



Fonte: Adaptado de Vorländer (2020).

A caracterização experimental de SLITs requer a medição de sua RI ou FRF. Diversos métodos foram desenvolvidos para essa finalidade. Excitações utilizando sinais de curta duração, como a utilização de disparos de pistola, buscam aproximar o impulso de Dirac, mas, além de serem suscetíveis a ruído. Também possuem limitações na energia espectral em baixas frequências, e podem excitar o sistema em regimes de amplitude não linear em alguns casos.

Uma abordagem robusta e amplamente utilizada emprega sinais determinísticos, como a Varredura Senoidal (*Sweep Sine*). Amplamente estudada e popularizada por Farina (2000; 2007; 2009), essa técnica utiliza um sinal senoidal cuja frequência varia com uma função predeterminada ao longo do tempo. A RI do sistema pode ser obtida através da deconvolução da resposta do sistema com o sinal de excitação original. O método de varreduras exponenciais oferece alta Relação Sinal Ruído (SNR)² e permite a medição simultânea da resposta linear e das distorções não lineares. A equação que descreve a varredura exponencial é expressa por

² Acrônimo do inglês para *Signal-to-Noise Ratio*.

$$x(t) = \text{sen} \left(\frac{2\pi f_0 T}{\ln \left(\frac{f_1}{f_0} \right)} \left[\left(\frac{f_1}{f_0} \right)^{t/T} - 1 \right] \right), \quad (2.4)$$

em que f_1 e f_2 são as frequências inicial e final do sinal, respectivamente e T é a sua duração.

Em uma revisão abrangente dos métodos de medição de Funções de Resposta em Frequência (FRF), Müller e Massarani (2001) também destacam as vantagens das varreduras exponenciais, enfatizando essa capacidade de proporcionar uma melhor SNR e a rejeição de possíveis não linearidades, que tendem a se manifestar na parte não causal da RI estimada a partir da desconvolução entre os sinais de excitação e resposta, expressa matematicamente no domínio discreto por

$$h[n] = \text{IDFT} \left\{ \frac{\text{DFT}\{y[n]\}}{\text{DFT}\{x[n]\}} \right\}, \quad (2.5)$$

em que $h[n]$ representa a resposta impulsiva do sistema, $y[n]$ é o sinal medido na saída e $x[n]$ o sinal de excitação aplicado na entrada. As operações DFT· e IDFT· correspondem, respectivamente, à transformada e à transformada inversa discreta de Fourier.

No mesmo contexto de caracterização de SLITs, Prawda et al. (2022) propõem um procedimento denominado *Rule of Two* (Ro2) para a seleção robusta de medições com varreduras exponenciais. A ideia central é realizar múltiplas medições sob as mesmas condições e avaliar a similaridade entre os sinais a partir do coeficiente de correlação³ de Pearson (PCC). Um par de medições, y_1 e y_2 , é considerado “limpo” quando sua correlação observada, ρ_{y_1, y_2} , excede um limiar adaptativo, $\hat{\rho}_{y_1, y_2}$.

Esse limiar é determinado considerando duas fontes de degradação: o ruído de fundo estacionário e a variação da FRF do sistema. Para garantir robustez, adota-se o cenário mais desfavorável, assumindo ruído de fundo totalmente anti-correlacionado⁴ entre as medições. O limiar resultante é então expresso como

$$\hat{\rho}_{y_1, y_2, \zeta=-1, \tau} = \frac{\hat{\rho}_{y_1, y_2, \zeta=-1}}{1 + \tau/2}, \quad (2.6)$$

³ O uso da correlação temporal, em vez da coerência espectral, segue a abordagem proposta por Prawda et al. (2022), pois permite avaliar a consistência global entre respostas temporais sem depender da estimativa da FRF, reduzindo a sensibilidade a ruído que pode ser mais proeminente em pontos específicos do espectro, por exemplo, em anti-picos da FRF.

⁴ O ruído anti-correlacionado, $\zeta = -1$, representa o cenário mais desfavorável, no qual o ruído de fundo das duas medições possui correlação negativa máxima, reduzindo o valor de correlação esperado mesmo para sinais idênticos.

em que τ representa o fator de variação da FRF e $\hat{\rho}_{y_1, y_2, \zeta=-1}$ corresponde ao PCC esperado para o cenário de ruído anti-correlacionado

$$\hat{\rho}_{y_1, y_2, \zeta=-1} = \frac{E[x] - E[n]}{E[x] + E[n]}. \quad (2.7)$$

Nessa equação, $E[x]$ representa a energia do sinal útil e $E[n]$ a energia do ruído. O par é aceito como válido se $\rho_{y_1, y_2} > \hat{\rho}_{y_1, y_2, \zeta=-1, \tau}$. Caso contrário, novas medições devem ser coletadas até que se obtenha uma condição satisfatória. Para aumentar a confiabilidade do critério, $E[n]$ é estimada de forma robusta pela mediana das amostras quadráticas do ruído, reduzindo a influência de ruídos transitórios. Dessa forma, o método Ro2 automatiza a seleção das melhores medições entre várias repetições, oferecendo maior robustez frente a ruídos não estacionários e simplificando o processo de validação experimental.

Outra técnica comum emprega Sequências de Comprimento Máximo (MLS⁵) (Rife; Vanderkooy, 1989). As MLS são sequências binárias pseudo-aleatórias com propriedades de auto-correlação ideais, aproximando um impulso quando correlacionadas com a resposta do sistema. Historicamente, sua facilidade de geração e a possibilidade de medições rápidas as tornaram atrativas em diversas aplicações. No entanto, com o avanço significativo da capacidade computacional moderna, essa vantagem relativa das MLS em termos de simplicidade e velocidade de medição tem se tornado menos distintiva em comparação com outras técnicas. As MLS apresentam desvantagens importantes, como um baixo fator de crista efetivo, o que pode limitar a dinâmica da medição. Além disso, a excitação de não linearidades do sistema pode ser problemática, gerando artefatos na resposta medida. Müller e Massarani (2001) também discutem essas limitações das MLS em sua revisão, apontando para a sua menor capacidade de rejeição de distorções não lineares em comparação com outros métodos.

Apesar de a MLS ser hoje considerada um método menos comum, a evolução da área de medição acústica levou ao desenvolvimento de técnicas mais robustas baseadas na correlação. Um exemplo é o trabalho de Farina (2009), que propôs a técnica do *silence sweep*. O método utiliza uma MLS com uma “imagem negativa” da varredura exponencial, cujo objetivo é gerar uma excitação para aprimorar a caracterização de sistemas não lineares.

Métodos baseados em sinais aleatórios como o ruído branco também podem ser utilizados. O ruído branco é caracterizado um espectro de potência constante em todas

⁵ Acrônimo do inglês para *Maximum length sequence*.

as frequências. Uma vantagem importante desses sinais é a sua natureza estacionária, que permite a aplicação de técnicas de análise espectral baseadas em médias, como o método de Welch (1967), para obter estimativas mais robustas da resposta do sistema. A RI do sistema é obtida por meio da correlação cruzada entre o sinal de entrada e a resposta do sistema (Bendat; Piersol, 2011). Adicionalmente, a utilização de sinais aleatórios geralmente utiliza a estimativa da função de coerência, que quantifica o grau de correlação linear entre a excitação e a resposta do sistema em função da frequência. A função de coerência fornece uma medida valiosa da qualidade da medição, indicando em quais faixas de frequência a resposta medida é predominantemente linear e influenciada pela excitação, e onde outros fatores, como ruído ou não linearidades, podem ter uma contribuição significativa.

A escolha da técnica de medição depende das características do sistema em análise, do ambiente de medição e dos requisitos de precisão, acurácia necessários para a aplicação específica.

2.1.1 Filtros lineares: definições e topologias convencionais

Filtros lineares são SLITs projetados para modificar o espectro de frequência de um sinal, atenuando ou amplificando seletivamente diferentes faixas de frequência. Sua aplicação é vasta, abrangendo desde o processamento de áudio e imagem até as telecomunicações e o controle de sistemas. A especificação de um filtro geralmente envolve a definição de uma ou mais faixas de frequência de passagem, em que os sinais devem ser transmitidos com mínima atenuação e faixas de rejeição, em que os sinais devem ser significativamente atenuados.

Diversas topologias de filtros são comumente utilizadas e amplamente discutidas na literatura (Oppenheim; Schaffer, 2009; Antoniou, 2005; Smith, 1997), cada uma com características específicas em termos de resposta em frequência e complexidade de implementação. Os filtros podem ser classificados, em sua forma mais básica, em quatro tipos principais, com base na faixa de frequência que eles permitem passar:

Filtro Passa-Baixas: atenua as frequências acima de uma determinada frequência de corte, f_c , e permite a passagem das frequências abaixo dela. São utilizados, por exemplo, para remover ruídos de alta frequência de sinais de áudio. Uma aplicação muito importante dos filtros passa-baixas analógicos é o anti-rebatimento, para evitar que frequências acima

da frequência de Nyquist contaminem o sinal digital (Oppenheim; Schafer, 2009). Em algoritmos de reamostragem de sinais digitais, os filtros passa-baixas digitais também devem ser utilizados para evitar o mesmo efeito de rebatimento.

Filtro Passa-Altas: atenua as frequências abaixo de uma determinada frequência de corte, f_c , e permite a passagem das frequências acima dela. Encontram aplicação na remoção de ruídos de baixa frequência ou *DC offset* de sinais;

Filtro Passa-Faixa: permite a passagem de sinais dentro de uma faixa de frequências definida entre duas frequências de corte (f_{c1} e f_{c2}) e atenua as frequências fora dessa faixa. São utilizados em equalizadores de áudio e em sistemas de comunicação para selecionar uma banda de frequência específica. No contexto da acústica, também desempenham um papel crucial na análise espectral, sendo amplamente empregados para representar sinais acústicos em bandas de oitava e terço de oitava, cujas especificações e critérios são definidos por normas internacionais como a ABNT NBR 61260-1 (2023). Além disso, no contexto deste trabalho, os filtros passa-faixa são particularmente úteis na preparação de sinais para a aplicação de análises psicoacústicas, como *fluctuation strength* e *roughness*. Essa representação facilita a análise da distribuição de energia sonora em diferentes faixas de frequência relevantes para a percepção humana e para a avaliação de ambientes acústicos.

Filtro Rejeita-Faixa (*notch*): atenua sinais dentro de uma faixa de frequências específica e permite a passagem das frequências fora dessa faixa. Filtros *notch* são particularmente úteis para remover ruídos de frequência estreita, como o “hum” da rede elétrica (50 ou 60 Hz). Além disso, esses filtros podem ser utilizados para remover componentes tonais de sinais em testes de acústica subjetiva e qualidade sonora, por exemplo, no trabalho de Di (2016), que propõe um modelo de incômodo sonoro para sinais de transformadores elétricos;

A implementação desses filtros pode ser realizada tanto de forma analógica, utilizando componentes como resistores, capacitores e indutores quanto de forma digital, por meio de algoritmos computacionais que operam sobre sequências de dados amostrados. A escolha entre a implementação analógica e digital depende de fatores como a aplicação, a exatidão requerida, a flexibilidade e o custo.

Em sistemas de processamento de sinais digitais, os filtros são frequentemente implementados utilizando estruturas como a de Resposta ao Impulso Finita (FIR⁶) e

⁶ Acrônimo do inglês para *Finite Impulse Response*.

Resposta ao Impulso Infinita (IIR⁷), cada uma com suas próprias vantagens e desvantagens em termos de resposta em fase, estabilidade e eficiência computacional (Oppenheim; Schaffer, 2009).

A equação recursiva geral para um filtro digital linear e invariante no tempo, que relaciona a saída $y[n]$ à entrada $x[n]$ e aos seus valores passados, é dada por uma equação de diferenças com coeficientes constantes. Essa equação é fundamental para a implementação de filtros digitais, especialmente os filtros IIR.

A equação geral é expressa por:

$$y[n] = \sum_{k=0}^M b_k x[n-k] - \sum_{k=1}^N a_k y[n-k], \quad (2.8)$$

em que $y[n]$ e $x[n]$ representam, respectivamente, as amostras de saída e de entrada no instante discreto n , b_k são os coeficientes do filtro que determinam a contribuição das amostras presentes e passadas da entrada para a saída e a_k são os coeficientes do filtro que determinam a contribuição das amostras passadas da saída para a saída (recursão).

Essa equação descreve um filtro digital geral que combina uma parte não recursiva, dependente apenas da entrada, e uma parte recursiva, dependente da saída passada. Assim, os filtros IIR são SLITs realimentados.

Filtros de Resposta ao Impulso Finito (FIR⁸) são um caso especial em que todos os coeficientes a_k (para $k \geq 1$) são zero, ou seja, sem realimentação, resultando em uma saída que depende apenas das amostras presentes e passadas da entrada. A escolha dos coeficientes b_k e a_k determina as características da resposta em frequência do filtro, como as frequências de corte, a atenuação na banda de rejeição e o ganho na banda de passagem. O projeto de filtros digitais envolve a determinação desses coeficientes para atender a requisitos específicos de filtragem, e existem vários métodos para obtê-los.

2.1.2 Filtros com espectro de magnitude arbitrária

Enquanto as topologias de filtros convencionais, como passa-baixas, passa-altas e passa-faixa, oferecem respostas em frequência comumente utilizadas e bem definidas, muitas aplicações em processamento de sinais requerem filtros com espectros de magnitude mais complexos e adaptados a características específicas do sinal ou do sistema em análise. Diferentemente das formas de resposta predeterminadas desses filtros básicos, o desenho

⁷ Acrônimo do inglês para *Infinite Impulse Response*.

⁸ Acrônimo do inglês para *Finite Impulse Response*.

de filtros com espectros de magnitude arbitrária busca moldar a resposta em frequência com elevado grau de detalhes sobre diversas faixas, permitindo a implementação de características espectrais personalizadas.

A importância desses filtros reside na sua capacidade de atender a demandas específicas, como a equalização de sistemas de áudio para compensar a resposta em frequência de alto-falantes e ambientes acústicos (Eargle, 1997; Karjalainen; Piirilä; Järvinen; Houpaniemi, 1997; Ramos; Lopez, 2006), a correção de respostas de sensores e transdutores para obter medições mais precisas (Plinge; Jacob; Haeb-Umbach; Fink, 2016), o projeto de bancos de filtros para análise espectral detalhada em aplicações de reconhecimento de padrões (Mallat, 1999), e a implementação de efeitos sonoros e musicais com características espectrais específicas (Zölzer, 2011; Murray, 2019; Watkinson, 2013). A flexibilidade no projeto da resposta em frequência desses filtros abre um vasto leque de possibilidades para o processamento e a manipulação de sinais em relação às suas características espectrais.

Diversas abordagens podem ser empregadas para o projeto de filtros com espectros de magnitude arbitrária, a depender da natureza do filtro (FIR ou IIR) e das características desejadas. No caso de filtros FIR, o método de janelamento é uma das técnicas mais utilizadas, permitindo moldar a resposta espectral a partir de uma função ideal truncada por uma janela. Já para filtros com fase mínima, uma alternativa comum envolve o uso da Transformada de Fourier Inversa (IFFT) aplicada a um espectro de magnitude alvo. Em filtros IIR, o método de Yule-Walker permite estimar diretamente os coeficientes a partir da modelagem autorregressiva de um espectro desejado.

2.1.2.1 Método de Janelamento para Projeto de Filtros FIR com Espectro de Magnitude Arbitrária

O método de janelamento é uma técnica clássica e amplamente utilizada para o projeto de filtros FIR com especificações de resposta em frequência arbitrárias (Oppenheim; Schafer, 2009; Smith, 1997). A ideia central consiste em multiplicar uma resposta ao impulso ideal desejada, $h_d[n]$, que teoricamente produziria a resposta em frequência especificada, por uma janela de comprimento finito, $w[n]$, resultando em uma resposta ao impulso implementável com duração finita, expressa por

$$h[n] = h_d[n] \cdot w[n]. \quad (2.9)$$

A resposta ao impulso ideal $h_d[n]$ é geralmente obtida pela Transformada de Fourier Inversa (IFFT) da resposta em frequência desejada, $H_d(e^{j\omega})$. No entanto, para respostas em frequência arbitrárias, essa IFFT pode resultar em uma resposta ao impulso de duração infinita, tornando a implementação direta impossível.

A aplicação da janela $w[n]$ limita a duração da resposta ao impulso para um comprimento finito $N + 1$, em que N é a ordem do filtro. A escolha da janela é crucial, pois suas características no domínio da frequência (lóbulo principal e lóbulos laterais) afetam diretamente a resposta em frequência do filtro resultante, influenciando a largura da banda de transição, a ondulação nas bandas de passagem e de rejeição, e a atenuação na banda de rejeição (Oppenheim; Schaffer, 2009). Janelas comuns incluem a retangular, Hamming, Hanning, Blackman, Kaiser, entre outras, cada uma oferecendo diferentes compromissos entre esses parâmetros.

No contexto da biblioteca `scipy.signal` em Python, a função `firwin2` (Weckesser, 2016) oferece uma implementação conveniente do método de janelamento para o projeto de filtros FIR com espectros de magnitude arbitrários definidos por partes lineares. Essa função permite especificar um conjunto de pares de frequência e magnitude desejada, e ela interpola linearmente entre esses pontos para definir a resposta em frequência alvo. A função, então, calcula os coeficientes do filtro FIR de ordem especificada, otimizando a resposta em frequência resultante por meio de uma janela apropriada, sendo a janela de Kaiser frequentemente utilizada internamente para fornecer um bom controle sobre as características do filtro.

O uso da função `firwin2` simplifica significativamente o processo de projeto, pois o usuário não precisa calcular explicitamente a resposta ao impulso ideal e aplicar a janela manualmente. A função lida com esses detalhes internamente, permitindo o projeto rápido de filtros FIR com formas de resposta em frequência complexas, adequados para aplicações como equalização espectral e modelagem de respostas em frequência específicas (Lyons, 2010).

Em resumo, o método de janelamento e, especificamente, a função `firwin2` do `scipy.signal`, fornecem uma ferramenta poderosa e flexível para o projeto de filtros FIR com espectros de magnitude arbitrária, superando as limitações dos filtros com topologias fixas e permitindo a adaptação precisa às necessidades de diversas aplicações de processamento de sinais.

Uma vantagem muito importante dessa técnica é a garantia de uma variação de fase

linear em relação à frequência, o que implica um espectro de atraso de grupo constante. Essa propriedade é crucial em aplicações sensíveis a distorções de fase, como o processamento de áudio para masterização e telecomunicações. Contudo, uma desvantagem significativa reside na potencial necessidade de um grande número de coeficientes (alta ordem do filtro) para obter uma convergência adequada entre a resposta em frequência desejada e a resposta real do filtro final. Isso pode resultar em um custo computacional elevado para o processamento do sinal, bem como em um atraso de grupo considerável, o que pode ser problemático em aplicações em tempo real ou com requisitos de baixa latência.

2.1.2.2 Obtenção da Representação de Fase Mínima via Transformada de Hilbert do Espectro de Magnitude

A propriedade de invertibilidade de um SLIT refere-se à existência de um sistema inverso que, quando posto em cascata com o sistema original, restaura o sinal de excitação original. Em outras palavras, um sistema invertível possui um “análogo inverso” capaz de neutralizar sua influência sobre o sinal de entrada. A representação de um sistema invertível em cascata com seu sistema inverso é ilustrada na Figura 2.

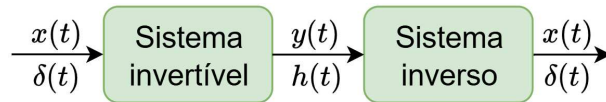


Figura 2 – Ilustração de um sistema em cascata com seu inverso.

Para determinar o sistema inverso de um SLIT, pode-se buscar a resposta ao impulso (RI) que, ao ser convoluída com a RI do sistema original, resulta em um sistema identidade, sendo a RI do sistema completo um impulso unitário, expressa por

$$h(t) * h_{\text{inv}}(t) = \delta(t). \quad (2.10)$$

De acordo com Manolakis (2005), um SLIT é classificado como de fase mínima se tanto ele quanto seu sistema inverso forem simultaneamente causais e estáveis.⁹ Este conceito assume particular importância para o presente trabalho, pois a calibração de cadeias de medição demanda a identificação da sua resposta ao impulso e, crucialmente, a obtenção de uma RI que efetivamente cancele seu efeito, conforme expresso na Equação (2.10).

⁹ A definição de estabilidade adotada neste trabalho segue o critério BIBO (*Bounded Input, Bounded Output*), segundo o qual um sistema é considerado estável quando toda entrada limitada gera uma saída igualmente limitada.

Filtros com resposta de fase mínima são de grande importância em diversas aplicações de processamento de sinais, especialmente quando um atraso de grupo não constante pode introduzir distorções indesejadas. Uma abordagem comum para obter a resposta de fase mínima a partir de um espectro de magnitude é a partir da relação entre magnitude e fase, utilizando a Transformada de Hilbert aplicada ao logaritmo do espectro de magnitude (Oppenheim; Schafer, 2009).

Essa técnica se baseia na propriedade fundamental de sistemas de fase mínima, em que a fase do espectro está univocamente relacionada ao logaritmo da magnitude do espectro através da Transformada de Hilbert. Conforme detalhado na literatura (Oppenheim, 2010; Ingle; Kogon; Manolakis, 2005; He, 1983; Hayes, 1996; Wikipedia, 2025), para um sistema de fase mínima $H(j\omega) = |H(j\omega)|e^{j\phi(\omega)}$, a fase $\phi(\omega)$ pode ser obtida a partir da magnitude $|H(j\omega)|$ usando a Transformada de Hilbert ($\mathcal{H}$)

$$\phi(\omega) = -\mathcal{H}\{\ln(|H(j\omega)|)\}. \quad (2.11)$$

E, reciprocamente, a magnitude pode ser reconstruída a partir da fase:

$$\ln(|H(j\omega)|) = \ln(|H(j0)|) + \mathcal{H}\{\phi(\omega)\}. \quad (2.12)$$

Neste trabalho, explora-se uma implementação específica dessa técnica, baseada em funcionalidades presentes nos pacotes Monkey Forest¹⁰ e ITA-Toolbox (Berzborn; Bomhardt; Klein; Richter; Vorländer, 2017), com o objetivo de otimizar a representação de fase mínima para o projeto de filtros compactos com um número reduzido de coeficientes e um espalhamento temporal minimizado.

A principal vantagem dessa abordagem reside na sua capacidade de gerar representações de fase mínima compactas, otimizando potencialmente o espalhamento temporal da resposta ao impulso e resultando em filtros eficientes com um número reduzido de coeficientes. Para alcançar essa compactação e otimização temporal, manipulações específicas são realizadas no domínio do tempo do sinal de Hilbert visando refinar a resposta de fase mínima para aplicações em que a latência e a complexidade computacional são fatores críticos.

A escolha de uma janela modificada também contribui para moldar as características temporais da resposta ao impulso resultante. Diferentemente de implementações genéricas

¹⁰ Uma descrição dessa aplicação é apresentada por Müller e Soares (2013) em seu trabalho sobre caracterização de microfones.

de fase mínima, esta abordagem busca uma otimização direcionada para o projeto de filtros com características temporais e espectrais específicas.

2.1.2.3 Método de Yule-Walker para Projeto de Filtros IIR com Espectro de Magnitude Arbitrária

O método de Yule-Walker oferece uma abordagem eficiente para o projeto de filtros IIR com espectros de magnitude arbitrária, a partir da modelagem autorregressiva de sinais. Ao contrário dos métodos de janelamento, que produzem filtros FIR com variação linear de fase em função da frequência, o método de Yule-Walker busca ajustar o espectro de magnitude desejado modelando-o como um processo autorregressivo (AR), cuja densidade espectral é determinada pelos coeficientes do denominador do filtro IIR (Hayes, 1996; Walker, 1931).

Esse procedimento envolve a estimação da função de autocorrelação associada ao espectro desejado, seguida da resolução das equações de Yule-Walker para obter os coeficientes a_k que definem os polos do filtro. Em implementações simplificadas, costuma-se assumir o numerador como constante, visando unicamente a modelagem da envoltória espectral.¹¹

A principal vantagem desse método é a capacidade de representar espectros complexos com filtros de ordem relativamente baixa, o que resulta em baixo custo computacional e menor atraso de processamento — características desejáveis em aplicações em tempo real. No entanto, como o método ajusta a resposta no sentido de mínimos quadrados, a aproximação pode não ser precisa em todas as faixas de frequência, especialmente em espectros com variações abruptas. Além disso, há o risco de instabilidade quando os polos resultantes ficam fora do círculo unitário.

No contexto prático, a biblioteca `yulewalker` para Python (Friedlander; Porat, 2007) permite aplicar o método de forma direta, recebendo o espectro alvo e a ordem do filtro como entrada, e retornando os coeficientes do denominador que melhor aproxima a resposta desejada.

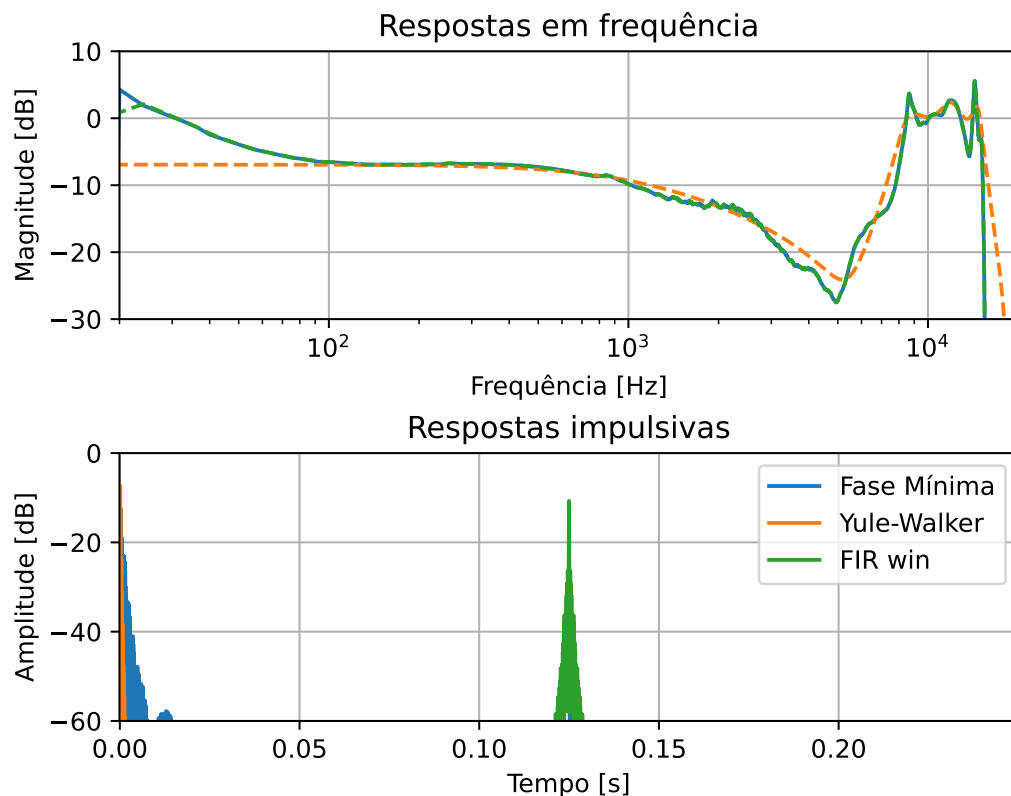
Na Figura 3, referente a um filtro de correção de fones de ouvido¹², observa-se que

¹¹ Embora os polos (coeficientes do denominador) influenciem tanto a magnitude quanto a fase da resposta em frequência, neste contexto considera-se apenas a modelagem do envelope espectral, razão pela qual o numerador pode ser constante.

¹² Nesse caso, há uma curva de FRF desejada obtida a partir de algum método de caracterização do fone, conforme será descrito na Seção 2.2.2. Essa curva foi omitida na Figura 3 por não ter diferença significativa do espectro de magnitude das curvas obtidas pelo filtro FIR e de fase mínima.

os métodos de fase mínima e FIR por janelamento reproduzem a resposta em frequência desejada com elevada concordância espectral, capturando detalhes finos do espectro. Em contrapartida, o FIR por janelamento apresenta resposta impulsiva simétrica, implicando atraso de grupo elevado (pico em torno de ~ 125 ms), o que pode ser limitante em aplicações de baixa latência. O filtro IIR obtido pelo método de Yule–Walker exibe maior discrepância nas baixas frequências, comportamento típico da modelagem AR em espectros com variações acentuadas. Porém resulta numa resposta impulsiva mais compacta, com decaimento mais rápido e custo computacional inferior devido ao número reduzido de coeficientes. Em contextos de correção de fones, a escolha entre essas abordagens reflete um compromisso entre fidelidade espectral (FIR/fase mínima) e eficiência temporal/latência (IIR Yule–Walker ou fase mínima).

Figura 3 – Exemplo: filtro de correção de fones de ouvido. Comparação de três métodos de projeto aplicados ao mesmo espectro-alvo de correção: fase mínima (a partir do módulo alvo), IIR por Yule–Walker e FIR por janelamento (`firwin`).



Fonte: Autor.

2.2 TECNOLOGIA BIAURICULAR

A tecnologia biauricular consiste em técnicas de gravação e reprodução sonora que buscam replicar a maneira como o sistema auditivo humano percebe naturalmente os sons no espaço tridimensional. A premissa fundamental é que, se a mesma pressão sonora for recriada na membrana timpânica do ouvinte, o estímulo sonoro, as mesmas sensações serão provocadas, incluindo a espacialidade e a sensação de volume sonoro. Diferentemente do áudio estéreo convencional, que geralmente utiliza dois canais para criar uma imagem sonora lateralizada, as técnicas biauriculares empregam métodos de captura e síntese que preservam as sutis informações espaciais inerentes à modificação dos sinais sonoros devido aos efeitos de difração provocadas pelo tronco, cabeça e orelha externa.

Ao apresentar sinais de áudio especificamente formatados para cada orelha, o biauricular possibilita a criação de uma experiência auditiva imersiva, com a percepção de altura, frente-trás e lateralização das fontes sonoras, simulando a audição no mundo real (Blauert, 1996; Møller; Sørensen; Jensen; Hammershøi, 1996; Hammershøi; Møller, 2002; Rafaely; Tourbabin; Habets; Ben-Hur; Lee; Gamper; Arbel; Birnie; Abhayapala; Samarasinghe, 2022; Burkhard; Sachs, 1975; Middlebrooks; Makous; Green, 1989; Hammershøi; Møller, 1996; Algazi; Avendano; Thompson, 1999; Masiero, 2012).

A chave para a eficácia de um áudio biauricular reside na consideração de como a cabeça, o tronco e, principalmente, os pavilhões auriculares (pinas) modificam as ondas sonoras que chegam aos nossos tímpanos. Essas modificações, que variam com a direção e a distância da fonte sonora, são geralmente modeladas pelas chamadas Funções de Transferência Relacionadas à Cabeça (HRTFs¹³). O áudio biauricular busca capturar ou sintetizar essas modificações, de forma que, ao serem reproduzidas através de fones de ouvido, as mesmas pistas espaciais estejam presentes na membrana timpânica do ouvinte, induzindo a percepção de que os sons se originam de diferentes pontos no espaço ao seu redor. No entanto, é importante ressaltar que nenhum sistema de gravação e reprodução biauricular consegue replicar perfeitamente o estímulo natural na membrana timpânica devido às incertezas na medição e reprodução das HRTFs, e às diferenças antropomórficas entre os indivíduos. Isso que motiva pesquisas sobre a individualização das HRTFs (Masiero, 2012), técnicas de rastreamento e adaptação do sinal biauricular em função do movimento da

¹³ Acrônimo do inglês para *Head Related Transfer Function*. Apesar desse conceito se referir a uma FRF, a nomenclatura de função de transferência é adotada na maior parte da literatura, portanto será mantida neste trabalho.

cabeça do ouvinte com o uso de *head trackers* (sistema de rastreamento do movimento da cabeça do ouvinte).

A capacidade de criar uma percepção espacial tridimensional rica e precisa torna o áudio biauricular uma tecnologia fundamental para diversas aplicações que buscam imersão e realismo sonoro. Desde a criação de experiências envolventes em realidade virtual e jogos até a produção de áudio para cinema e sistemas de comunicação mais naturais, o biauricular explora a sofisticada capacidade do sistema auditivo humano de localizar e contextualizar sons no ambiente, oferecendo um potencial significativo para aprimorar a interação e a experiência do usuário com o mundo sonoro virtual e real (Rumsey, 2012).

A relevância da tecnologia biauricular se manifesta em uma ampla gama de aplicações, cada uma com seus próprios objetivos e requisitos específicos. No entretenimento, como jogos e cinema, a prioridade reside na criação de espacialidade, imersão, coloração espectral e inteligibilidade para o ouvinte, intensificando a experiência narrativa e sensorial. Em acústica de salas, o foco se volta para a avaliação da espacialidade da reprodução sonora, a correlação interaural dos sinais que chegam às orelhas, e a fração lateral do som, buscando otimizar a qualidade da audição em ambientes específicos (Brandão, 2016; Long, 2005).

No âmbito da acústica de edificações, a tecnologia biauricular pode ser empregada para analisar a coloração espectral do som dentro de um espaço, as FRFs entre diferentes pontos, o nível de pressão sonora, e, crucialmente, requer calibração e ajuste do sistema de reprodução para garantir medições e simulações precisas (Thaden, 2005). A compreensão desses diferentes focos e requisitos é fundamental para a aplicação eficaz da tecnologia biauricular em cada contexto específico.

2.2.1 Resposta em frequência relacionada à cabeça

A Função de Resposta em Frequência Relacionada à Cabeça (HRTF) representa a influência da cabeça, do tronco e das orelhas no campo acústico. Matematicamente, define-se como a relação entre a pressão sonora em um ponto dentro do canal auditivo, $P_2(f, r, \theta, \phi)$, e a pressão sonora no centro da cabeça na ausência dessas estruturas, $P_1(f)$, (Blauert; Jekosch, 1997; Wightman; Kistler, 1989; Hartmann, 2004; Masiero, 2012).

Assim, a HRTF é definida matematicamente pela expressão

$$\text{HRTF}(f, r, \theta, \phi) = \frac{P_2(f, r, \theta, \phi)}{P_1(f, r)}, \quad (2.13)$$

em que f é a frequência, r é a distância da fonte sonora, e θ e ϕ são os ângulos de incidência em coordenadas polares (azimuth e elevação, respectivamente), conforme ilustrado em trabalhos como (Algazi; Duda; Thompson; Avendano, 2001; Masiero, 2012). Para ondas planas, observa-se que a HRTF não varia significativamente com a distância fonte-receptor.

A HRIR para uma posição fonte-receptor fixa pode ser obtida aplicando a **Transformada Inversa Discreta de Fourier (IDFT)** à HRTF. A pressão sonora no centro da cabeça (P_1) é independente da direção de incidência, supondo que os efeitos de direcionalidade do microfone sejam negligenciáveis e que efeitos de campo próximo não sejam relevantes, condições que são satisfeitas por ondas planas no campo distante.

As HRTFs codificam informações cruciais para a percepção espacial auditiva, atuando como filtros lineares e variantes com a direção que modificam a RI e o espectro sonoro de maneira única para cada ponto no espaço ao redor do ouvinte. As principais pistas espaciais codificadas pelas HRTFs incluem as seguintes

- **Diferenças Interauriculares de Nível (ILDs).** A atenuação dependente da frequência causada pela cabeça e pelo pavilhão auricular resulta em diferentes níveis de pressão sonora nas duas orelhas para fontes sonoras laterais, especialmente em altas frequências.
- **Diferenças Interauriculares de Tempo (ITDs).** A diferença no tempo de chegada do som às duas orelhas, resultante da separação física da cabeça, fornece informações sobre a localização lateral da fonte sonora. O sistema auditivo é mais sensível a essa pista em baixas frequências, região na qual as ILDs são consideravelmente baixas, não proporcionando informação suficiente para a localização sonora.
- **Pistas Espectrais (*Spectral Cues*).** As reflexões e difrações causadas pela cabeça, tronco e, principalmente, pelas complexas geometrias dos pavilhões auriculares introduzem modificações espectrais no sinal sonoro. Essas modificações, como picos e vales na resposta em frequência, são fundamentais para a percepção da elevação e da distinção entre fontes sonoras frontais e traseiras, especialmente em altas frequências e para posições de simetria, em que as diferenças interauriculares são baixas. Além disso, essas pistas são importantes para avaliação de qualidade sonora, pois as modificações espectrais afetam as características percebidas dos sinais.

Um aspecto importante das HRTFs é que sua natureza é altamente individual, dependendo da anatomia específica da cabeça, do tronco e das orelhas de cada ouvinte ([Wightman; Kistler, 1989](#); [Masiero, 2012](#)). Variações significativas nas dimensões da cabeça, no formato do pavilhão auricular e na distância entre as orelhas levam a HRTFs únicas para cada indivíduo. O uso de HRTFs não individualizadas pode resultar em uma espacialização sonora mais pobre e em fenômenos de “internalização” do som dentro da cabeça do ouvinte. A medição de HRTFs pode ser realizada por meio de medições diretas em indivíduos ou em [HATS](#). Nesse caso, são utilizados os microfones embutidos nas orelhas do [HATS](#) que geralmente são projetados para ter características antropométricas médias. Uma ampla revisão sobre a evolução dos modelos de [HATS](#) é feita por Paul ([2009](#)).

Já para obter HRTFs de sujeitos, as medições diretas envolvem a colocação de microfones minúsculos dentro dos canais auditivos dos participantes enquanto fontes sonoras são apresentadas em diversas posições angulares ao redor deles. A montagem das bancadas pode ser desafiadora e uma abordagem muito inovadora para medição de múltiplas HRTFs individualizadas em um curto período de tempo é proposta por Masiero ([2012](#)).

Além disso, existe a abordagem de modelos computacionais para sintetizar HRTFs com base em parâmetros geométricos, embora a precisão ainda seja um desafio mesmo com o avanço da capacidade computacional. Brinkmann *et. al.* ([2023](#)) descrevem uma biblioteca de código aberto para calcular HRTFs utilizando o [Método de Elementos de Contorno \(BEM\)](#).¹⁴

A compreensão das HRTFs e de como elas codificam a informação espacial é fundamental para o desenvolvimento de sistemas de áudio biauricular eficazes, visando a criação de experiências sonoras imersivas e realistas em diversas aplicações.

2.2.2 Equalização de Sistemas de Reprodução Biauricular

A equalização de sistemas de reprodução é uma etapa crucial para garantir a fidelidade na apresentação de áudio biauricular. Conforme discutido na Seção [2.1.2](#) sobre o desenho de filtros digitais com espectros de magnitude arbitrária, é possível moldar a resposta em frequência de um sistema de reprodução para compensar as características espectrais indesejadas introduzidas pelos componentes do sistema, como fones de ouvido e

¹⁴ Acrônimo do inglês para *Boundary Elements Method*.

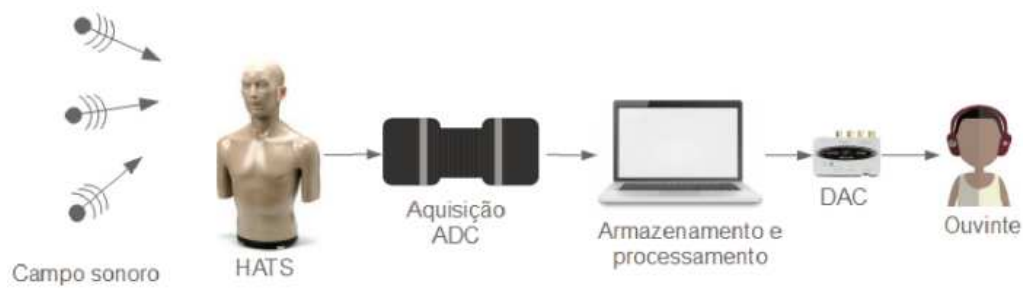
interfaces de áudio. O objetivo da equalização é compensar a resposta em frequência do sistema de reprodução (fone–ouvido), de modo que o sinal sonoro reproduzido preserve as características espectrais originais da condição de escuta. Dessa forma, as informações espaciais codificadas nas HRTFs podem ser percebidas sem distorções significativas introduzidas pelo sistema de reprodução (Møller, 1992).

A Figura 4 ilustra de forma genérica um sistema de gravação e reprodução biauricular utilizando HATS e fones de ouvido. De acordo com Moller (1992), caso o fone de ouvido seja de acoplamento equivalente ao campo livre (FEC¹⁵), a pressão sonora pode ser gravada em qualquer ponto do canal auditivo, o que não afetará o desempenho na reprodução.

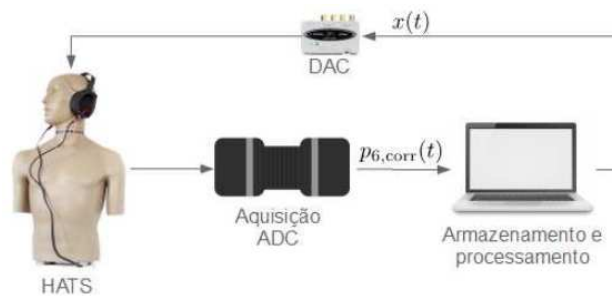
Além disso, Brum (2020) avaliou o processo de equalização de um fone de ouvido FEC utilizando um HATS para equalizar e gravar sinais estacionários e impulsivos com o próprio fone. Uma das conclusões importantes da pesquisa foi a aparente irrelevância da fase na correção do sistema de reprodução para a reprodução de sinais estacionários e impulsivos, considerando o *loudness* variante no tempo como parâmetro de avaliação. Os experimentos indicaram que a correção da magnitude da resposta em frequência foi o fator dominante para a obtenção de uma reprodução biauricular precisa, ao se utilizar parâmetros psicoacústicos como referência para comparar diferentes métodos de filtragem. Essa observação pode simplificar o processo de equalização, permitindo concentrar o projeto dos filtros na resposta de magnitude do sistema de reprodução.

¹⁵ Acrônimo do inglês para *Free Air Coupled*.

Figura 4 – Ilustração genérica de um sistema de gravação biauricular e sua calibração.



(a) Sistema de gravação e reprodução biauricular.



(b) Calibração do sistema de gravação e reprodução.

Fonte: Brum (2020).

Porém, quando outros tipos de fone são utilizados, esse comportamento pode não ser observado. Li *et. al.* (2019) avaliaram o processo de equalização de fones de ouvido utilizando correção somente de magnitude, de fase ou de ambas. Os autores concluíram que a correção de fase tem um impacto considerável no decaimento da resposta transiente dos fones testados. Nesse caso, a equalização de fase melhorou a clareza dos áudios reproduzidos, segundo os autores.

Dessa forma, além dos aspectos de equalização em magnitude e fase, a literatura também aponta para a importância de critérios que avaliem o grau de abertura (*openness*) dos fones de ouvido utilizados na reprodução biauricular. Esses critérios têm como objetivo verificar em que medida os fones preservam a carga acústica natural¹⁶ da orelha externa, condição fundamental para uma reprodução perceptualmente realista.

Entre os principais critérios encontrados na literatura, destacam-se os seguintes

- **PDR¹⁷**: introduzido por Møller e colaboradores (1995), o PDR verifica se a pressão sonora no tímpano durante a reprodução por fones é equivalente àquela registrada em condição de campo livre. Valores de PDR próximos de 1 indicam que o fone não

¹⁶ A carga acústica da orelha externa refere-se à impedância acústica apresentada pelo conduto auditivo e pelo pavilhão auricular, que determina como a pressão sonora é transferida para o tímpano.

¹⁷ Acrônimo do inglês para *Pressure division ratio*.

altera significativamente a impedância de irradiação da orelha, característica dos fones classificados como [Acoplamento Equivalente de Campo Livre \(FEC\)](#)¹⁸.

- [HSC](#)¹⁹: proposto por Völk como extensão do PDR, o HSC compara diretamente a relação entre pressões no canal auditivo bloqueado e no tímpano para condições de campo livre e de reprodução via fones. Valores próximos a 1 indicam que o fone mantém uma carga acústica equivalente à condição natural ([Völk, 2012](#); [Masiero; Fels, 2011](#)).
- Impedância acústica: a medição direta da impedância acústica dos fones de ouvido, realizada por meio de tubo de impedância, fornece uma caracterização robusta principalmente em baixas frequências, em que os efeitos de oclusão podem ser mais relevantes ([Schlieper; Li; Preihs; Peissig, 2018](#)).
- [Índice de Oclusão \(OI\)](#): parâmetro de desenvolvimento mais recente, derivado da curva de impedância acústica, que quantifica o reforço de baixas frequências associado ao efeito de oclusão. Esse índice serve como indicador prático da percepção subjetiva de “abertura” ou “fechamento” do fone ([Schlieper; Li; Preihs; Peissig, 2019](#)).

Os parâmetros utilizados neste trabalho foram o [PDR](#) e [HSC](#), cujas equações são

$$\text{PDR}(f) = \frac{P_3(f)/P_2(f)}{P_6(f)/P_5(f)}, \quad (2.14)$$

em que $P_2(f)$ e $P_3(f)$ são os espectros da pressão sonora com excitação externa do canal bloqueado e aberto, respectivamente, e $P_5(f)$, $P_7(f)$ são os espectros da pressão sonora com excitação por fones de ouvido do canal bloqueado e aberto, respectivamente,

$$\text{HSC}(f) = \frac{P_4(f)/P_2(f)}{P_7(f)/P_5(f)}, \quad (2.15)$$

em que $P_3(f)$ e $P_6(f)$, são as pressões do canal aberto com excitação por fonte externa e fone de ouvido, respectivamente. A medição de impedância acústica envolve aparato de tubo de impedância e não foi avaliada neste trabalho.

A incorporação desses critérios auxilia na seleção de fones adequados para experimentos de reprodução binaural, permitindo maior controle sobre a fidelidade acústica e a validade perceptual dos resultados.

¹⁸ Acrônimo do inglês para *Free Air Equivalent Coupling*.

¹⁹ Acrônimo do inglês para *Headphone Selection Criterion*.

Hammershøi e Møller (1996) destacaram que a gravação na entrada bloqueada do canal auditivo é preferível, pois preserva de forma mais fiel as pistas espaciais e apresenta menor variabilidade entre sujeitos, sendo portanto mais adequada para a obtenção das curvas de equalização de fones de ouvido. Quando essa abordagem é combinada ao uso de fones classificados como *Free Air Equivalent Coupling* (FEC), as variações antropomórficas dos canais auditivos individuais podem ser em grande parte negligenciadas, aumentando a consistência dos resultados entre diferentes ouvintes. Além disso, calibrações individualizadas em sujeitos são procedimentos onerosos, que exigem infraestrutura específica e pessoal qualificado para a inserção de microfones em canais auditivos. Nesse sentido, a utilização de HATS com microfones posicionados na entrada bloqueada do canal auditivo para calibração de fones de ouvido mostra-se muito mais prática e economicamente viável, favorecendo a padronização e a reprodutibilidade em aplicações experimentais.

Posteriormente, Møller et al. (1999) realizaram uma avaliação comparativa de diferentes HATS em testes de localização auditiva, confrontando-os com a escuta em campo real. Os autores observaram que, embora os manequins permitam uma reprodução biauricular padronizada, o desempenho de localização obtido com suas gravações foi consistentemente inferior ao da condição real, principalmente devido a confusões no plano mediano e inversões frente-trás. Ainda assim, ressaltaram que os HATS constituem uma alternativa prática e superior a outras formas de reprodução não biauricular, embora demandem aprimoramentos para representar de maneira mais fiel a anatomia média do ouvinte humano.

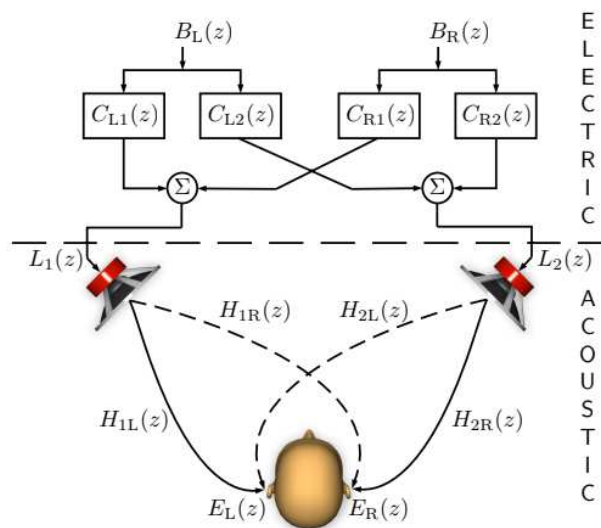
2.2.3 Apresentação de sinais utilizando alto-falantes

Embora a reprodução biauricular seja tipicamente associada ao uso de fones de ouvido, também é possível implementá-la utilizando alto-falantes. No entanto, a reprodução com alto-falantes introduz o problema do *crosstalk*, no qual o sinal destinado a uma orelha também é incide sobre a outra, comprometendo a separação biauricular e a percepção espacial. Para equalizar sistemas de reprodução biauricular com alto-falantes, é frequentemente empregada a técnica de cancelamento de *crosstalk*. Essa técnica envolve o projeto de filtros que, quando aplicados aos sinais direcionados a cada alto-falante, geram ondas sonoras que cancelam o sinal indesejado que chega à orelha oposta (Choueiri, 2008; Masiero, 2012). A equalização do sistema com cancelamento de *crosstalk* visa garantir que

o sinal correto, com as pistas espaciais desejadas, chegue a cada orelha do ouvinte. Um diagrama esquemático desse sistema é ilustrado na Figura 5.

Dessa forma, a equalização de sistemas de reprodução biauricular, seja através de fones de ouvido FEC calibrados com um HATS correspondente ou através de sistemas de alto-falantes com cancelamento de *crosstalk*, utilizando filtros com espectro de magnitude arbitrária, representa uma estratégia eficaz para minimizar as distorções espectrais e os problemas de reprodução. A correção adequada da magnitude da resposta em frequência, juntamente com o tratamento do *crosstalk* em sistemas de alto-falantes, é fundamental para garantir que as pistas espaciais codificadas nas HRTFs sejam apresentadas ao ouvinte de forma íntegra, contribuindo para uma experiência biauricular imersiva e espacialmente precisa.

Figura 5 – Diagrama ilustrando a reprodução biauricular com filtros de cancelamento de *crosstalk*.



Fonte: Masiero (2012).

2.2.4 Normas e Padrões para Simuladores de Cabeça e Torso

Diversas normas internacionais foram desenvolvidas para padronizar a construção e o uso de HATSs, com foco em garantir a reprodutibilidade e a comparabilidade de medições acústicas em áreas como a eletroacústica, a audiolgia e as telecomunicações. Essas normas definem requisitos geométricos e de desempenho acústico, assegurando que os simuladores representem de forma consistente as características médias da anatomia humana.

2.2.4.1 Padrões para Audiologia e Eletroacústica (Série IEC 60318)

A série de normas **IEC 60318** descreve diferentes acopladores e simuladores voltados principalmente para a calibração de fones de ouvido e aparelhos auditivos. Entre os documentos, destacam-se os seguintes

- **IEC 60318-1** (2009): especifica métodos para a medição de fones supra-auriculares e circumauriculares;
- **IEC 60318-4** (2010): define um simulador de orelha ocluída para medições com fones de inserção;
- **IEC 60318-7** (2022): descreve um manequim de cabeça e torso (HATS) para medições de fontes próximas ao ouvido.

Essas normas têm como objetivo central garantir a padronização técnica em medições de resposta em frequência, impedância acústica e calibração de dispositivos. O foco está na avaliação e conformidade de equipamentos, e não necessariamente na reprodução fiel de experiências auditivas complexas.

2.2.4.2 Especificações para Medições Acústicas In-Situ (ANSI S3.36)

A norma **ANSI S3.36** (2012) especifica um manequim destinado a simulações de medições acústicas aéreas em condições *in-situ*. O modelo descrito compreende cabeça, orelhas externas, canais auditivos e um torso representativo da média anatômica adulta. Assim como as normas IEC, a ANSI S3.36 estabelece parâmetros geométricos e de desempenho acústico, voltados principalmente para a avaliação de dispositivos eletroacústicos em situações práticas.

2.2.4.3 Outros referenciais normativos

Além da IEC e da ANSI, há especificações adicionais voltadas para o cálculo de parâmetros psicoacústicos, como recomendações da **ECMA** e da **ISO**. Essas incluem a utilização de funções de transferência até o tímpano (*transfer function to the eardrum*) como referência, aplicáveis tanto a microfones quanto a HATS, reforçando a padronização em contextos de telecomunicações e inteligibilidade da fala.

2.2.4.4 Lacunas para Qualidade Sonora

Apesar dos avanços normativos em audiologia e telecomunicações, observa-se uma lacuna significativa no que diz respeito ao uso de [HATS](#) em estudos de qualidade sonora de produtos. Embora as normas existentes forneçam parâmetros para a calibração e caracterização de dispositivos, ainda não há diretrizes consolidadas que estabeleçam procedimentos padronizados para:

- gravação biauricular de campos sonoros complexos;
- reprodução perceptualmente válida em ensaios subjetivos de qualidade sonora;
- definição de parâmetros específicos que correlacionem a captação com a percepção de atributos sonoros.

Essa ausência de padronização resulta em variações entre laboratórios e dificulta a comparação entre diferentes estudos. O presente trabalho busca justamente contribuir nesse ponto, discutindo adaptações e recomendações para a aplicação de [HATS](#) em contextos de avaliação de qualidade sonora.

2.3 FUNDAMENTOS PSICOACÚSTICOS DAS SENSações SONORAS E PARÂMETROS DE AVALIAÇÃO

De forma ampla e generalista, a psicoacústica pode ser definida como o estudo da relação entre as sensações auditivas e os estímulos sonoros. Compreender a resposta sensorial dos ouvintes em relação às características acústicas dos sinais sonoros é fundamental para avaliar a qualidade sonora em diversas aplicações, incluindo a reprodução de áudio.

Além disso, a avaliação subjetiva, baseada na experiência perceptiva humana, é um critério muito importante para determinar a adequação e a qualidade de um sinal ou sistema de áudio. Para auxiliar a avaliação subjetiva, os modelos psicoacústicos²⁰ fornecem estimativas objetivas de atributos relacionados às sensações auditivas. Embora não realizem uma medição direta da percepção, esses métodos buscam modelar a magnitude de fenômenos como *loudness*, *sharpness* e *roughness*, que são amplamente utilizados na

²⁰ Na literatura encontram-se tanto os termos parâmetro psicoacústico quanto modelo psicoacústico. O primeiro refere-se à grandeza derivada de um modelo que quantifica uma sensação auditiva, enquanto o segundo enfatiza o aspecto computacional da modelagem, que busca estimar essa grandeza a partir do sinal físico e de hipóteses sobre o funcionamento do sistema auditivo.

avaliação da qualidade sonora (Zwicker; Fastl, 2013; Blauert; Jekosch, 1997; Vorländer, 2020).

A *loudness*, quantificada em sone, representa a sensação de volume sonoro. A literatura apresenta diversos modelos para calcular a *loudness*, em que se pode destacar duas grandes escolas: a de Munique e a de Cambridge. Na primeira, o modelo é proposto por Zwicker (1965). Já na escola de Cambridge, há o modelo proposto por Moore e Glasberg (1996). A norma mais atual que descreve os modelos de *loudness* é a ISO 532-1:2017 (2017), sendo a Parte 1 para o modelo de Zwicker e a Parte 2 para o modelo Moore-Glasberg.

A *sharpness* (sensação de agudeza), que depende do conteúdo espectral do sinal, pode ser estimada utilizando a norma DIN 45692:2009-08 (2009). Já a *roughness*, expressa em asper, estima a sensação de flutuação rápida (Zwicker; Fastl, 2013). Esses modelos estão implementados em ferramentas computacionais como a biblioteca MOSQUITO para Python (San Millán-Castillo; Latorre-Iglesias; Glessner; Wanty; Jiménez-Caminero; Álvarez-Jimeno, 2021), que é de código aberto, gratuita e permite automatizar análises de forma reprodutível, sem restrições para aplicações científicas ou comerciais.

Para as análises subsequentes realizadas neste trabalho, essa biblioteca foi utilizada para calcular o *loudness*, cujo modelo implementado é o de Zwicker e descrito na norma ISO 532-1 (2017). Já a *sharpness* foi calculada a partir das recomendações da norma DIN 45692 (2009) e o modelo de *roughness* implementado é o de Daniel e Weber (1997).

Embora os modelos psicoacústicos ofereçam informações valiosas sobre atributos específicos do som, a avaliação subjetiva por ouvintes humanos permanece essencial. Testes subjetivos permitem avaliar aspectos da qualidade sonora que podem não ser totalmente capturados por parâmetros isolados, como a naturalidade, o conforto auditivo e a ausência de artefatos. A combinação de análises objetivas e subjetivas fornece uma compreensão mais completa da qualidade sonora de um sistema ou sinal. Para o desenvolvimento futuro deste trabalho de doutorado, pretende-se explorar e incorporar outros parâmetros psicoacústicos relevantes para a avaliação de diferentes aspectos da percepção sonora, complementando as análises realizadas.

2.4 MÉTRICAS DE COMPARAÇÃO DE SINAIS

As métricas de comparação têm por objetivo avaliar a diferença entre sinais ou espectros a partir de um parâmetro único. Tais medidas são amplamente utilizadas em processamento de fala e áudio para análise de distorções, avaliação objetiva de qualidade e como aproximações para diferenças perceptivas (Loizou, 2007). De forma geral, podem ser classificadas em intrusivas ou não intrusivas.

Os modelos intrusivos requerem um sinal de referência para comparação. Exemplos incluem a MCD, a distância Itakura–Saito e a distância log-espectral, descritas nas seções subsequentes.

Por outro lado, os modelos não intrusivos operam apenas sobre o sinal degradado ou processado, sem a necessidade de uma referência, como ocorre nos métodos P.563 e NISQA. No presente trabalho, são consideradas apenas modelos do primeiro tipo, dado que os sinais de referência estão disponíveis. A seguir, apresentam-se algumas das medidas mais empregadas na literatura.

2.4.1 Distância espectral euclidiana

A distância espectral euclidiana consiste em uma métrica direta que avalia a diferença ponto a ponto entre dois espectros. Embora seja de implementação simples e tenha servido de base para técnicas de correção espectral, não possui fundamentação psicométrica consolidada. Sua utilidade está em quantificar diferenças espectrais absolutas, sem necessariamente refletir a percepção auditiva humana.

De forma geral, pode ser expressa como

$$D_{\text{EUC}} = \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} |P(\omega) - \hat{P}(\omega)|^p d\omega \right\}^{1/p}, \quad (2.16)$$

em que $P(\omega)$ representa a densidade espectral de potência do sinal de referência, $\hat{P}(\omega)$ corresponde ao espectro estimado ou processado, ω é a frequência angular e p define a norma L^p utilizada, sendo $p = 2$ o caso mais comum na literatura.

2.4.2 Distância Itakura–Saito

A divergência Itakura–Saito foi proposta originalmente por Fumitada Itakura e Shuzo Saito no contexto de análise e síntese de fala por máxima verossimilhança (Itakura; Saito, 1968). Trata-se de uma métrica assimétrica, não satisfazendo a desigualdade trian-

gular, e por isso não constitui uma distância no sentido estrito. Apesar dessa limitação, é amplamente empregada em modelagem espectral de sinais de fala e música, pois captura diferenças relevantes no domínio log-espectral e se conecta a conceitos estatísticos de entropia relativa. Estudos posteriores discutiram suas propriedades e aplicações em processamento de sinais multivariados e interpolação espectral (Jiang; Ning; Georgiou, 2012; Enqvist; Karlsson, 2008).

Matematicamente, é definida como

$$D_{\text{IS}}(P(\omega), \hat{P}(\omega)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\frac{P(\omega)}{\hat{P}(\omega)} - \log \frac{P(\omega)}{\hat{P}(\omega)} - 1 \right] d\omega, \quad (2.17)$$

em que $P(\omega)$ representa a densidade espectral de potência do sinal original, $\hat{P}(\omega)$ corresponde à estimativa espectral obtida pelo modelo e ω é a frequência angular.

2.4.3 Distância Log-Espectral

A distância log-espectral, também conhecida como *log-spectral distance*, é frequentemente utilizada para avaliação objetiva de qualidade de fala sintetizada ou transmitida (Gray; Markel, 2003; Barnwell III; Clements; Quackenbush, 1988). Diferentemente da medida de Itakura–Saito, esse parâmetro é simétrico, satisfazendo $D(x, y) = D(y, x)$, o que o caracteriza como uma distância válida no sentido matemático. Esse modelo é sensível a diferenças no logaritmo do espectro, aproximando-se mais da forma como o sistema auditivo humano percebe variações espectrais.

É definida por

$$D_{\text{LS}} = \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} [\log P(\omega) - \log \hat{P}(\omega)]^p d\omega \right\}^{1/p}, \quad (2.18)$$

em que $P(\omega)$ e $\hat{P}(\omega)$ são, respectivamente, os espectros de referência e estimado, ω é a frequência angular e p é o parâmetro da norma L^p empregada, sendo $p = 2$ o caso mais usual.

2.4.4 Mel-Cepstrum Distance

A *Mel-Cepstrum Distance* **MCD**²¹ é uma métrica amplamente empregada para quantificar a diferença espectral entre dois sinais de fala ou áudio no domínio dos coeficientes

²¹ Uma tradução possível para o português seria Distância Cepstral em Escala Mel. Entretanto, esse também é um termo não consolidado na literatura em português, logo, optou-se por manter a nomenclatura original em inglês.

cepstrais em escala Mel (MFCCs). Originalmente proposta por Kubichek (1993) como medida objetiva de similaridade em sistemas de síntese e reconhecimento de fala, a MCD tornou-se referência em diversos estudos de avaliação de qualidade de fala e técnicas de conversão de voz.

Matematicamente, a MCD entre dois sinais é definida como:

$$\text{MCD}(k) = \alpha_e \cdot \sqrt{\sum_{d=1}^D (\text{MFCC}_x(d, k) - \text{MFCC}_y(d, k))^2}, \quad (2.19)$$

em que $\text{MFCC}_x(d, k)$ e $\text{MFCC}_y(d, k)$ representam o d -ésimo coeficiente cepstral em escala Mel para o k -ésimo quadro dos sinais x e y , respectivamente, D é o número de coeficientes considerados, e α_e é um fator de escala, comumente definido por

$$\alpha_e = \frac{10}{\ln 10} \sqrt{2} \approx 6, 14. \quad (2.20)$$

O valor final da métrica é obtido pela média ao longo dos quadros é calculado por

$$\text{MCD} = \frac{1}{N} \sum_{k=1}^N \text{MCD}(k), \quad (2.21)$$

em que N é o número total de quadros analisados, definido pela duração do sinal e de cada segmento.

Além da formulação original, algumas variações incorporam alinhamento temporal entre as sequências de MFCCs, empregando *Dynamic Time Warping*, o que permite comparações robustas entre sinais de diferentes durações ou com pequenas variações temporais (Kubichek, 1993; Kominek; Black, 2004; Morise; Yokomori; Ozawa, 2016).

Parâmetros usuais

A implementação disponível na biblioteca `mel-cepstral-distance`²² utiliza como valores padrão: taxa de amostragem de 16 kHz, janela de 25 ms, avanço de 10 ms, FFT de 512 pontos, uso de 24 filtros Mel, limite de frequência entre 0 e 8 kHz e cálculo sobre 12 coeficientes cepstrais ($D = 12$), com $\alpha_e = \frac{10}{\ln 10} \sqrt{2}$. Esses parâmetros estão alinhados ao que é mais frequentemente encontrado na literatura, conforme sistematizado em (Kubichek, 1993; Kominek; Black, 2004; Morise; Yokomori; Ozawa, 2016).

²² Link para a biblioteca: <https://pypi.org/project/mel-cepstral-distance/>

Aplicações

A MCD é amplamente utilizada em tarefas de síntese e conversão de fala, avaliação de algoritmos de codificação e em análises de correspondência espectral em estudos psicoacústicos. Sua principal vantagem é capturar diferenças perceptualmente relevantes no espectro de fala em uma métrica compacta, embora não represente diretamente medidas subjetivas de inteligibilidade ou qualidade percebida.

2.5 QUALIDADE SONORA DE SISTEMAS DE REFRIGERAÇÃO

Conforme mencionado anteriormente, a qualidade sonora de eletrodomésticos, em particular de refrigeradores, tem se tornado um fator decisivo na percepção de valor e na satisfação do consumidor. Sendo um equipamento que opera de forma contínua em ambientes domésticos, o ruído gerado por um refrigerador pode ser fonte constante de incômodo. Dessa forma, a avaliação da qualidade sonora desses equipamentos passou a ser uma área de estudo relevante na engenharia e na acústica.

A pesquisa de Jeon et al. (2007) é fundamental para estabelecer as características do ruído de refrigeradores. Os autores analisaram a radiação sonora de diferentes tipos de refrigeradores em ambientes de laboratório e em ambientes de sala de estar reais. Um dos principais achados foi a diferença no nível de pressão sonora: o nível medido em uma sala de estar era cerca de 10 dB mais alto que o medido em uma câmara anecoica, o que destaca a importância de avaliar o ruído em condições de uso real ao avaliar a qualidade sonora do produto final. Porém, nas etapas de pesquisa e desenvolvimento, também é importante realizar medições em ambientes padronizados como câmara reverberante ou semi-anecoica. Isso assegura uma padronização de testes para realizar comparações entre amostras e configurações de montagem.

Complementando os estudos na área, You e Jeon (2009) aprofundaram a análise, com o objetivo de determinar a DLO para os principais parâmetros de qualidade sonora de refrigeradores. Utilizando sinais de refrigeradores reais e sintetizados em testes subjetivos, os autores determinaram os limiares como sendo 0,5 sone para *loudness*, 0,08 acum para *sharpness* e 0,04 asper para *roughness*. Nesse trabalho, os autores calcularam os parâmetros utilizando o software B&K Sound Quality System, o qual, segundo os autores, possui implementados os modelos de Zwicker para *loudness*, *sharpness* e *roughness*²³.

²³ Os autores referem-se a esses parâmetros como “modelos de Zwicker”. Entretanto, não há documentação

A percepção do ruído também é fortemente influenciada pelas condições de instalação. Em sua dissertação de mestrado, Dal Moro (2019) estudou a influência de diferentes ambientes de cozinha e de diversas condições de montagem na percepção sonora de refrigeradores. A pesquisa comparou a avaliação de ruídos em uma câmara semi-anecoica com a simulação de ambientes reais, mostrando como o contexto de instalação pode alterar a percepção do som. De maneira similar, a dissertação de Demoliner (2017) explorou um modelo de predição de incômodo, correlacionando parâmetros psicoacústicos com a avaliação subjetiva do ruído gerado por diferentes compressores instalados em um mesmo refrigerador e em um ambiente típico de uso. Nesse contexto, o compressor é frequentemente considerado a fonte de ruído predominante, sendo objeto de avaliação acústica isolada nas fases de projeto e desenvolvimento do produto (Demoliner, 2017). Tanto Dal Moro (2019) como Demoliner (2017) utilizaram HATS sem simulador de pina e com canal bloqueado para obter os sinais biauriculares para suas análises.

Ainda sobre a percepção de ruídos específicos, o trabalho de Silveira (2022) avaliou a detecção de sinais impulsivos em refrigeradores de uso doméstico. O estudo comparou o desempenho de dois métodos psicofísicos, o método da escada (*staircase method*) e o PEST. Os resultados indicaram que o método PEST apresentou um desempenho superior para a detecção desses sinais, fornecendo uma ferramenta mais eficaz para a avaliação e aprimoramento da qualidade sonora.

Por fim, a dissertação de Brum (2020) focou na influência de diferentes técnicas de aurilização para a avaliação de qualidade sonora. O autor explorou a aplicação de **HRTFs** em sinais gravados com microfones simples, encontrando boa concordância com as gravações feitas com **HATS**. Essa abordagem sugere que é possível obter resultados confiáveis de percepção espacial do som, embora com a limitação de que as pistas espaciais do ambiente de gravação original são perdidas no processo. A pesquisa também demonstrou uma discrepância na resposta da sala a sinais anecoicos caso a direcionalidade da fonte não seja considerada.

Em resumo, a literatura sobre a qualidade sonora de refrigeradores evoluiu da análise de propriedades básicas do som para a incorporação de parâmetros psicoacústicos e a consideração do ambiente de utilização e de configurações de instalação. Trabalhos mais

pública confirmando quais algoritmos são efetivamente implementados no software B&K Sound Quality System (Type 7698), utilizado no estudo. É possível que as implementações sejam baseadas ou derivadas dos modelos de Zwicker, mas com adaptações proprietárias.

recentes buscam otimizar as técnicas de gravação e reprodução para avaliações subjetivas e aprimorar os métodos de teste psicofísicos para sons específicos, garantindo que o desenvolvimento de produtos seja guiado por dados perceptuais precisos.

2.6 MÉTODOS SUBJETIVOS PARA AVALIAÇÃO DE DIFERENÇA ENTRE SINAIS

A avaliação subjetiva desempenha um papel muito importante ao determinar as diferenças de percepção entre sinais sonoros. Diversos métodos podem ser empregados para quantificar a capacidade dos ouvintes de discriminar entre diferentes estímulos. A seguir, serão descritos os aspectos gerais de alguns desses métodos.

2.6.1 Teste AB

O teste AB é uma técnica simples e direta para avaliação subjetiva, na qual o participante escuta dois estímulos sonoros, denominados A e B, apresentados sequencial ou simultaneamente. Após a escuta, o participante deve julgar se os estímulos são diferentes ou, em testes de preferência, indicar qual dos dois é preferível em algum critério previamente definido, por exemplo, qualidade sonora ou naturalidade.

Por solicitar uma resposta binária do ouvinte, o teste AB exige baixo esforço cognitivo e é de fácil aplicação, sendo amplamente utilizado em experimentos perceptivos com grupos variados de ouvintes. No entanto, sua simplicidade também representa uma limitação, pois o método é um teste de escolha forçada de duas alternativas em que o participante deve fazer uma escolha mesmo quando a diferença entre os estímulos não é perceptível, o que pode introduzir um viés de adivinhação nos resultados.

Aplicações típicas do teste AB incluem comparações entre versões de um mesmo sinal processadas por diferentes algoritmos ou sistemas. Alguns estudos ([Kanetada; Yamamoto; Mizumachi, 2013](#); [Pras; Zimmerman; Levitin; Guastavino, 2009](#)) empregaram o teste AB para avaliar diferenças perceptivas entre sinais comprimidos em diferentes taxas de codificação MP3, demonstrando sua utilidade na análise de degradações sutis e preferências subjetivas em cenários de compressão de áudio.

2.6.2 Teste ABX

Descrita nos trabalhos de Munson e Gardner ([1950](#)) e Clark ([1982](#)), consiste em uma técnica robusta e amplamente aceita para avaliação de discriminação. Nesse método,

o participante escuta dois sinais de referência distintos, designados como A e B. Em seguida, um terceiro sinal, rotulado como X, é apresentado. Este sinal X é sempre idêntico a um dos sinais de referência, ou seja, X é uma cópia exata de A ou de B, escolhida aleatoriamente em cada tentativa. A tarefa do participante consiste em identificar se o sinal X é perceptualmente mais similar ou idêntico ao sinal A ou ao sinal B. Dessa forma, o teste ABX é um teste de escolha forçada de duas alternativas e 3 estímulos sonoros.

A análise dos resultados do teste ABX se baseia na taxa de acertos dos participantes. Se a proporção de identificações corretas de X como A ou B for estatisticamente significativa acima do nível de chance (aproximadamente 50% em um cenário de escolha binária), conclui-se que os sinais A e B são perceptivelmente distinguíveis para o grupo de ouvintes testado. Por outro lado, se a taxa de acertos não se desviar significativamente do acaso, sugere-se que as diferenças entre os sinais A e B não são perceptíveis ou são indistinguíveis para os ouvintes sob as condições do teste. A interpretação dos resultados, no entanto, depende criticamente de uma análise estatística apropriada (Boley; Lester, 2009), pois uma simples taxa de acertos pode ser enganosa.

Uma das principais vantagens do método ABX reside na sua robustez metodológica e na sua ampla aceitação na comunidade de pesquisa em áudio. Ele fornece uma medida direta da capacidade de discriminação perceptível entre dois estímulos sonoros. No entanto, uma limitação importante dos testes AB e ABX é que eles apenas informam se uma diferença é perceptível ou não, sem fornecer informações sobre a magnitude dessa diferença ou sobre a preferência dos ouvintes por um sinal em relação ao outro. Para obter informações sobre o grau de diferença ou preferências, outros métodos de avaliação subjetiva são necessários.

Exemplos de aplicação do método ABX são os trabalhos de Rogowska (2015) e Lee et al. (2013), que utilizaram a técnica para comparar diferentes compressões de áudio em instrumentos musicais. Rogowska (2015) investigou a audibilidade de artefatos introduzidos por compressão com perdas (*lossy*), especialmente em tons instrumentais sustentados, revelando que, em baixas taxas de bits, as diferenças perceptivas tornam-se significativamente mais detectáveis, sendo o teste ABX fundamental para comprovar a discriminabilidade entre os sinais comprimidos e os originais.

Lee et al. (2013) adotaram uma abordagem complementar, baseada em comparações triádicas de similaridade, uma variação conceitualmente próxima ao teste ABX, para quantificar a alteração no espaço de timbre (*timbre space*) provocada pela compressão MP3 em diferentes taxas de bits. Os resultados demonstraram que, enquanto taxas mais

altas (128 kbps) mantêm a estrutura perceptual do espaço de timbre, compressões maiores (32 kbps) causam distorções significativas, evidenciadas tanto pela redução nas correlações entre os pares de dissimilaridade quanto pela reorganização das posições dos instrumentos em análises de escalonamento multidimensional. Esses efeitos tornam-se perceptíveis, por exemplo, quando instrumentos originalmente semelhantes, como clarinete e oboé, passam a ser percebidos como mais distintos após a compressão.

Esses estudos ilustram o papel central desse tipo de teste como uma ferramenta robusta para investigar mudanças sutis ou acentuadas na qualidade perceptiva de sinais comprimidos, sendo particularmente útil em contextos onde o julgamento humano é essencial para validar algoritmos de codificação ou modelos de percepção sonora.

2.6.3 Escolha Forçada de 3 Alternativas

O teste de Escolha Forçada de 3 Alternativas (3AFC²⁴) é um método projetado para obter uma medida mais robusta do limiar de detecção ou discriminação. Enquanto o teste ABX apresenta dois estímulos de referência (A e B) e um terceiro (X) para comparação, o teste 3AFC simplifica a tarefa para o ouvinte. Nesse método, três estímulos sonoros são apresentados, sendo dois idênticos e um diferente. A tarefa do participante é identificar qual dos três estímulos é o diferente (“*odd-one-out*”). Essa configuração reduz a chance de acerto por acaso de 50% (como no teste AB) para 33,3%.

Frequentemente, o teste 3AFC é combinado com procedimentos adaptativos, como o método da escada, para determinar a DLO. A partir de um nível de diferença inicial, o procedimento ajusta o parâmetro de diferença entre os estímulos. Nos estudos de Fritz et al. (2007) e Woodhouse et al. (2012), que investigaram os limiares de percepção para sons de violino e violão, os parâmetros manipulados incluíram a frequência, o amortecimento das ressonâncias do corpo do instrumento e a rigidez da corda, obtendo os limiares para cada parâmetro sob estudo. O teste é finalizado quando um número pré-determinado de reversões (*turnpoints*) é atingido, indicando que a DLO foi estimada com precisão.

A utilização do 3AFC é particularmente útil em aplicações cujo objetivo é quantificar a sensibilidade do sistema auditivo a pequenas variações em parâmetros físicos. Em contraste com o teste ABX, que se concentra em uma única comparação de similaridade, o 3AFC exige que o ouvinte se concentre em detectar a anomalia, tornando-o ideal para a

²⁴ Acrônimo do inglês para *Three-Alternative Forced-Choice*.

medição precisa de limiares de discriminação.

2.6.4 Norma ITU-R BS.1116-3 (2015)

Essa norma descreve um protocolo para avaliação subjetiva de pequenas diferenças em sistemas de áudio de alta fidelidade. O método recomendado é denominado Teste de Estímulo Triplo Duplo-Cego com Referência Oculta²⁵, também conhecido como "ABC-h". Nesse teste, três estímulos são apresentados ao participante: o sinal de referência conhecido (A), a versão processada e a referência oculta, atribuídas aleatoriamente às posições B e C. A tarefa do ouvinte é avaliar o grau de degradação percebido de B e C em relação ao A, utilizando uma escala discreta de cinco pontos baseada na percepção de distorção (de 5 = imperceptível até 1 = muito incômodo).

Este método se destaca por sua alta sensibilidade para detecção de diferenças sutis, exigindo ouvintes treinados e controle estrito das condições experimentais. Ele é recomendado para contextos nos quais os sinais comparados apresentam diferenças tão pequenas que poderiam ser negligenciadas em testes mais simples como o ABX ou AB. Além disso, a norma fornece diretrizes detalhadas sobre o ambiente de teste, seleção de ouvintes, tratamento estatístico e atributos auditivos a serem avaliados, como qualidade básica do áudio e imagem estéreo.

2.6.5 Teste MUSHRA

O MUSHRA é um método de avaliação subjetiva amplamente utilizado, detalhado na norma ITU-R BS.1534-3 (2014) e referenciado em diversos trabalhos da literatura. Esse teste é particularmente indicado para a avaliação de sistemas de qualidade intermediária, sendo esperado que o Dispositivo sob Teste (DUT)²⁶ introduza modificações consideráveis nos sinais testados.

Em cada rodada do teste, múltiplos estímulos são apresentados aos participantes para avaliação comparativa. O número de estímulos sonoros, incluindo a referência oculta e as âncoras, não deve exceder 12 sinais. A avaliação é realizada utilizando uma escala de classificação contínua de 0 a 100, ancorada em descritores verbais que representam diferentes níveis de qualidade, por exemplo, de “imperceptível” a “muito ruim”. A interface de avaliação é projetada para otimizar a consistência dos resultados, com regras como o

²⁵ Do inglês: *double-blind triple-stimulus with hidden reference*.

²⁶ Acrônimo do inglês para *device under test*.

congelamento do *slider* de classificação enquanto o sinal é ouvido e a exigência de que todos os sinais de uma rodada sejam reproduzidos antes de o resultado ser salvo. Um exemplo de interface sugerido pela norma é ilustrado na Figura 6.

Para normalizar a escala e fornecer pontos de referência, o teste inclui uma referência oculta e âncoras. A referência oculta é uma cópia idêntica da referência principal, cuja presença não é informada aos participantes. Ela serve como um controle de qualidade para a consistência das respostas, sendo que a norma estabelece os seguintes critérios para a avaliação da referência oculta:

- Idealmente, ela deve ser classificada com a pontuação máxima (100%);
- Caso a referência oculta receba uma classificação abaixo de 90% em mais de 15% das avaliações, os resultados do teste podem ser considerados inválidos, indicando inconsistência ou falta de atenção por parte do avaliador.

As âncoras são sinais de qualidade inferior à referência e ajudam os avaliadores a calibrar suas percepções. A norma, por exemplo, sugere o uso de âncoras criadas por filtros passa-baixa em 3,5 kHz e 7 kHz. Na aplicação implementada para este trabalho, foi utilizada somente uma âncora (explícita e oculta), conforme será descrito no Capítulo 3.

O teste MUSHRA consiste em um método bastante robusto e tem sido amplamente aplicado em diversas áreas, incluindo a avaliação da qualidade de codecs de áudio, a qualidade de reprodução de fones de ouvido e a eficácia de sistemas de controle ativo de ruído. Como exemplo, pode-se citar o estudo de Soulodre e Lavoie (2004), que utilizou o protocolo para avaliar a qualidade subjetiva de codecs MPEG Layer II com e sem replicação espectral de banda (SBR).

2.6.6 Método de Ajuste

O método de ajuste oferece uma abordagem interativa para quantificar a diferença subjetiva entre dois sinais sonoros, permitindo que o próprio participante manipule um dos sinais até que ele seja percebido como equivalente ao outro. Nesta metodologia, o ouvinte recebe um sinal de referência (A) e um segundo sinal ajustável (B).

De acordo com a estrutura do teste, o participante pode ter a capacidade de modificar diversos parâmetros do sinal B, como por exemplo, sua equalização espectral, nível de intensidade ou outros atributos relevantes para a percepção da similaridade com o sinal A.

Figura 6 – Exemplo de interface gráfica sugerido pela norma.



Fonte: ITU R BS.1534-3 (2014).

Por meio de uma interface de controle, o ouvinte realiza ajustes iterativos até atingir um ponto em que o sinal B se torna perceptualmente indistinguível ou o mais similar possível ao sinal de referência A, de acordo com seu julgamento subjetivo.

O valor final dos ajustes realizados pelo participante pode ser interpretado como uma medida indireta da magnitude da diferença subjetiva inicial entre os sinais A e B. Grandes ajustes sugerem uma diferença perceptual maior, enquanto ajustes mínimos indicam uma similaridade inicial mais elevada.

O controle do estímulo pode ser implementado por meio de um seletor contínuo e deve ser desprovido de marcações e livre de qualquer tipo de trava mecânica, a fim de evitar pistas táteis que possam influenciar ou enviesar a resposta dos participantes (Gelfand, 2017).

Além disso, é comum a inserção de um segundo controle entre o seletor do participante e o sistema de instrumentação, possibilitando ao pesquisador alterar o ponto de partida de cada série de testes de forma que permaneça desconhecida ao participante. Essa abordagem busca minimizar vieses associados à posição inicial do controle e ao uso inconsciente de configurações anteriores como referência entre tentativas consecutivas. Atualmente, as técnicas de processamento digital de sinais proporcionam uma enorme gama de possibilidades para implementação do método do ajuste.

Recentemente, variações do método de ajuste têm sido exploradas em contextos

aplicados, especialmente em tarefas que simulam ambientes reais de tomada de decisão técnica. Leonard (2015), por exemplo, propôs uma adaptação do método baseada em tarefas de mixagem (*mix tasks*), nas quais engenheiros de áudio ajustam parâmetros técnicos, como nível de reverberação ou balanceamento estéreo, com o objetivo de igualar ou otimizar uma mixagem de referência.

Essa abordagem permite avaliar a percepção auditiva de forma contextualizada, aproximando o teste da prática profissional. De forma similar, Martin et al. (2016) utilizaram tarefas de mixagem para investigar a audibilidade de artefatos em arquivos MP3 de alta taxa de bits. Nesse caso, os participantes eram instruídos a igualar sua mixagem a uma versão de referência sem conhecer previamente o formato de codificação do sinal, o que permitia inferir diferenças perceptivas a partir das decisões técnicas tomadas. Segundo os autores, essa metodologia foi inspirada nos exercícios do curso *Advanced Technical Ear Training*, ministrado por Martin e Massenburg (2015). Tais variações demonstram o potencial do método de ajuste quando integrado a contextos mais realistas e interativos, oferecendo uma alternativa sensível e prática para a avaliação subjetiva de diferenças sutis entre sinais.

No entanto, uma desvantagem significativa é que o processo pode ser demorado e a precisão dos resultados pode depender da familiaridade e da habilidade do participante em realizar ajustes sonoros finos. Além disso, a natureza iterativa e subjetiva do processo pode introduzir variabilidade entre diferentes participantes.

2.7 ESTIMATIVA DE PARÂMETROS POR TESTE SEQUENCIAL

A estimativa de parâmetros por teste sequencial (PEST) foi proposta por Taylor, Forbes e Creelman (1967) como um procedimento adaptativo destinado à estimativa rápida e eficiente de limiares perceptíveis. Diferente de métodos fixos, em que os parâmetros do estímulo são definidos previamente e aplicados em toda a sessão experimental, a PEST seleciona dinamicamente seus valores com base no histórico das respostas do participante. Essa característica permite concentrar as medições na região próxima ao limiar, reduzindo o número de tentativas necessárias para obter estimativas precisas.

O princípio fundamental da PEST é aplicar incrementos relativamente grandes no início do teste e, a cada resposta do observador, ajustar tanto a direção quanto o tamanho dos passos. Quando ocorre uma reversão de tendência (por exemplo, o estímulo deixa

de ser detectado após ter sido percebido), o tamanho do passo é reduzido, geralmente pela metade, de forma que a sequência converge gradualmente para o ponto crítico da função psicométrica. Esse processo continua até que seja atingido um critério de parada pré-definido, normalmente estabelecido pelo tamanho mínimo do passo aceitável para a resolução desejada do limiar.

Do ponto de vista estatístico, a [PEST](#) incorpora o teste sequencial de Wald como regra de decisão, assegurando que, em cada nível, seja possível avaliar com eficiência se o desempenho do observador está acima ou abaixo da probabilidade alvo, geralmente 50% em tarefas de detecção ou valores específicos em tarefas de escolha forçada. O nível estimado ao final do rastreamento é tomado como o limiar perceptivo para a condição em análise.

Uma das vantagens do método é sua flexibilidade: pode ser aplicado em diferentes modalidades sensoriais (audição, visão, tato) e em tarefas de detecção, discriminação ou julgamento de igualdade. Além disso, é adequado tanto para participantes experientes quanto para ouvintes leigos, sendo amplamente empregado em psicofísica pela sua robustez e simplicidade operacional. Estudos de simulação e aplicações práticas mostraram que, para limiares associados a probabilidades de acerto típicas como 75% em tarefas 2AFC, a [PEST](#) atinge convergência em menor número de tentativas, oferecendo uma relação vantajosa entre precisão e esforço experimental ([Wetherill; Levitt, 1965](#); [Kaernbach *et al.*, 1991](#)).

3 COMPARAÇÃO DE SINAIS DE SISTEMAS DE REFRIGERAÇÃO OBTIDOS COM HATS E MICROFONE EM CÂMARA SEMI-ANECOICA

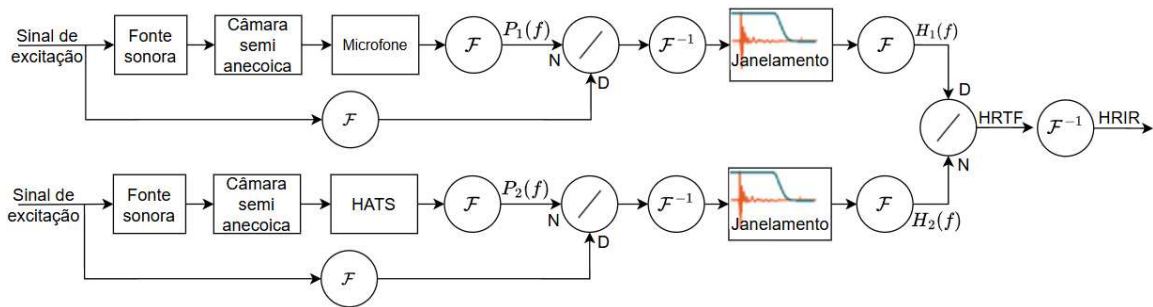
Neste capítulo é apresentado um estudo comparativo entre medições biauriculares realizadas com **HATS** e com microfone em câmara semi-anecoica. O objetivo central é analisar as diferenças objetivas e subjetivas entre os sinais obtidos pelas duas abordagens, de modo a verificar em que medida um microfone convencional pode substituir ou complementar o uso de manequins na gravação biauricular voltada à avaliação da qualidade sonora de sistemas de refrigeração.

3.1 MEDIÇÃO DA HRTF

A primeira etapa deste estudo foi a medição da **HRTF** para que sua respectiva **HRIR** fosse aplicada aos sinais gravados com microfone de campo livre HBK modelo 4189. Todas as medições (**HRTF** e sinais sonoros) foram realizadas com o HATS HSU III.2 da Head Acoustics voltado para a fonte sonora, ou seja, com ângulos polar e azimutal ($\theta = \phi = 0$, ver Figuras 9 e 12).

O processo de medição é ilustrado na Figura 7. Um sinal de excitação foi enviado para a fonte sonora que excita o campo acústico do ambiente (câmara semi-anecoica). As FRFs são calculadas a partir da pressão sonora medida no microfone e na entrada do canal auditivo bloqueado do torso, respectivamente.

Figura 7 – Diagrama esquemático da medição da HRTF.



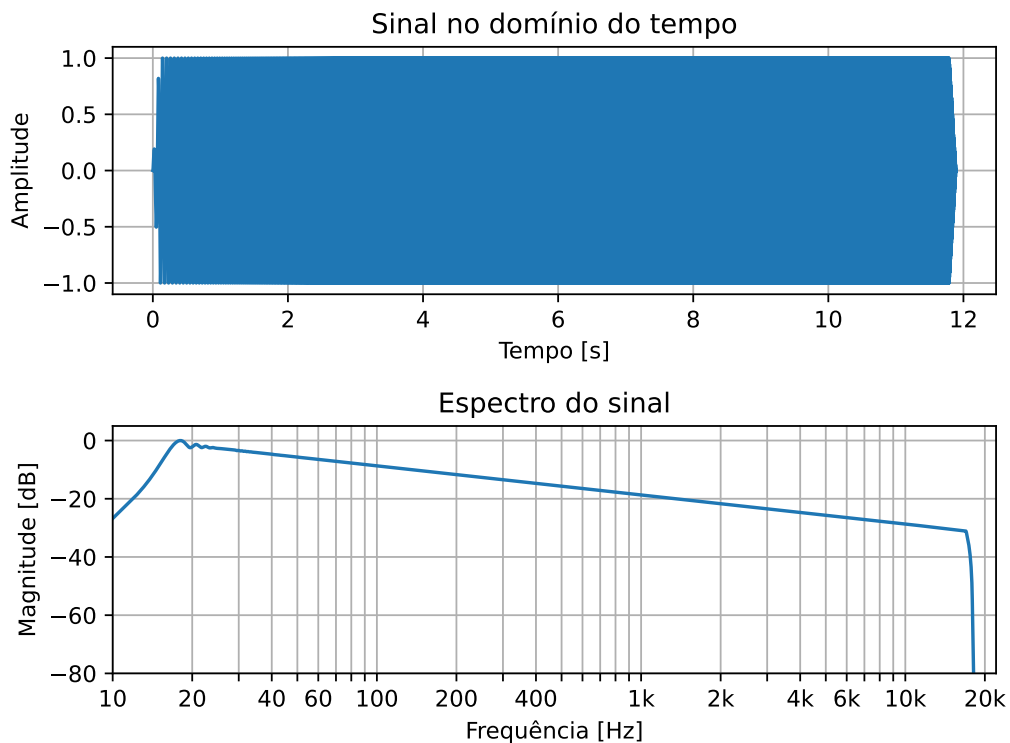
Fonte: Autor.

Para mitigar a influência da reflexão do piso da câmara, as FRFs são convertidas para o domínio do tempo, janeladas e, então, convertidas para o domínio da frequência novamente. Finalmente, a HRTF é calculada pela razão entre os espectros $H_1(f)$ e $H_2(f)$ e a HRIR a partir de sua transformada de Fourier. As FRFs $H_1(f)$ e $H_2(f)$, bem como a RI

e janela, são mostradas na Figura 10.¹ As diferenças entre a curva do HATS e microfone é a HRTF apresentada na Figura 11.

O sinal de excitação utilizado foi uma varredura exponencial, com frequência de amostragem de 44,1 kHz e 2^{19} amostras, resultando em uma duração total $T \approx 11,9$ s (ver Equação 2.4). As frequências inicial (f_1) e final (f_2) foram definidas como 15 Hz e 18 kHz, respectivamente, cobrindo a faixa de interesse do estudo. Para evitar descontinuidades no início e no final do sinal, aplicaram-se funções de suavização do tipo *fade-in* e *fade-out*. A Figura 8 apresenta o sinal no domínio do tempo e seu respectivo espectro de magnitude obtido a partir da FFT com janela retangular em um único bloco.

Figura 8 – Sinal de excitação utilizado nas medições.

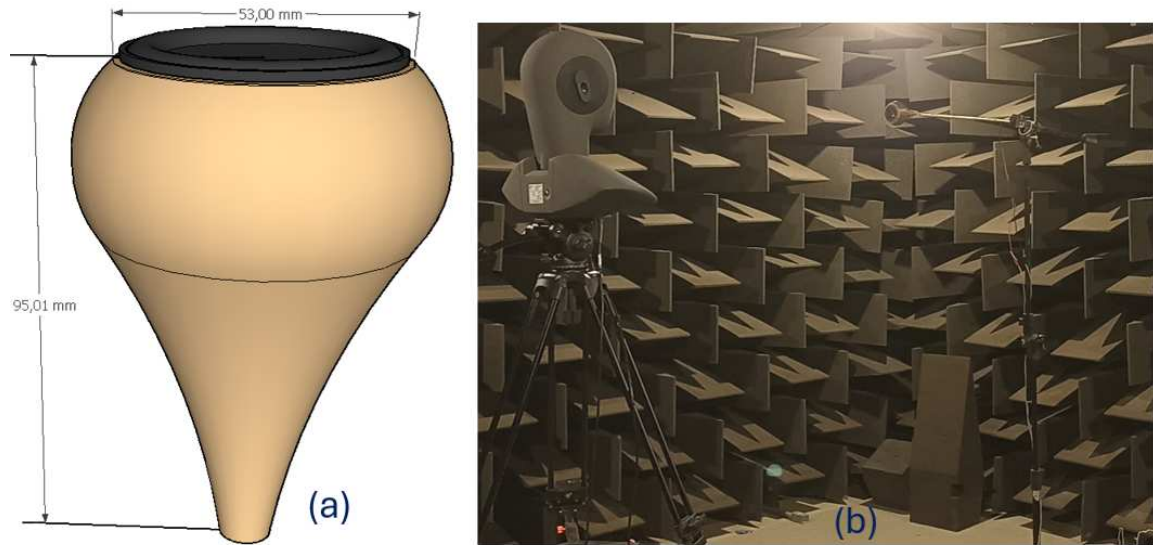


Fonte: Autor.

Para garantir ondas planas no receptor, foi prototipada uma fonte sonora de pequenas dimensões, conforme mostrado na Figura 9(a). Além disso, foi utilizado um monitor de áudio Genelec 8020D para fazer a mesma medição e comparar os resultados.

¹ Na Figura 10 é apresentada uma curva de microfone sem a aplicação da HRTF. A partir desse resultado, todos os demais dados de microfone apresentados correspondem ao sinal convoluido com a HRIR, de modo a incorporar as características espaciais do sistema de reprodução.

Figura 9 – Bancada de medição da HRTF: (a) Desenho da fonte sonora utilizada para excitação do campo acústico, e (b) Bancada montada na câmara semi-anecoica.

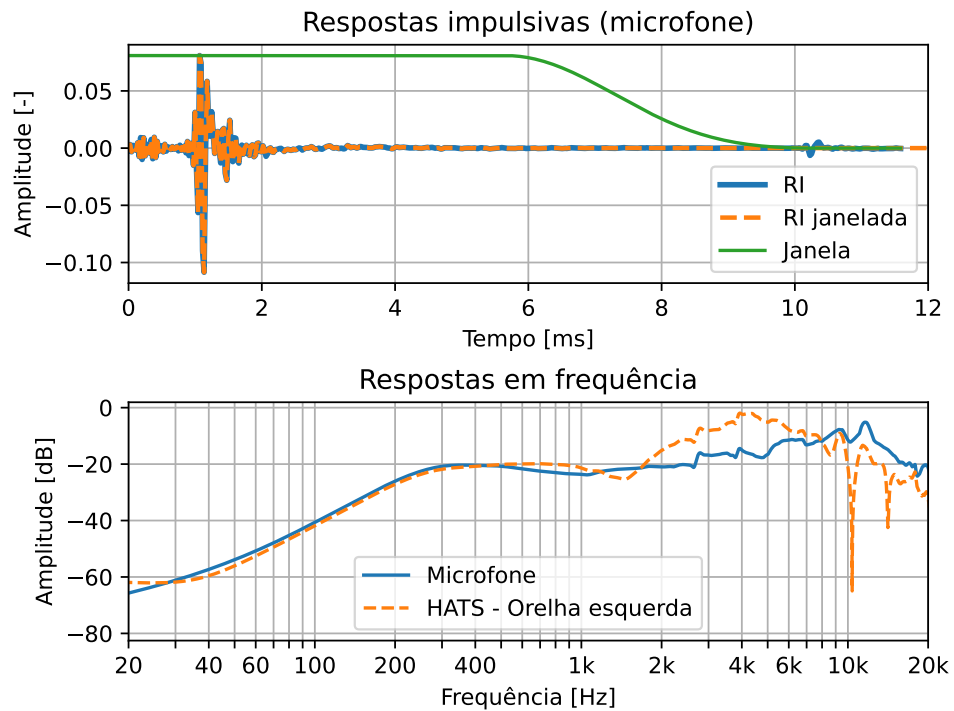


Fonte: Autor.

A Figura 11 mostra os espectros de magnitude da HRTF obtida com as duas fontes sonoras excitando o campo acústico.² Ao comparar as curvas obtidas com as duas fontes sonoras, observa-se boa concordância nos dados, em que as discrepâncias maiores ocorrem somente a partir de 1 kHz. Além disso, as discrepâncias observadas abaixo de 60 Hz ocorrem devido à baixa energia sonora da fonte prototipada em função de suas dimensões reduzidas. Porém, nessa faixa de frequências sabe-se que o HATS é praticamente transparente ao campo acústico, permitindo que os valores da HRTF sejam configurados com 0 dB de ganho (Masiero, 2012). Ademais, as discrepâncias entre altas frequências podem ser causadas pelo reposicionamento do HATS e fontes sonora entre as medições.

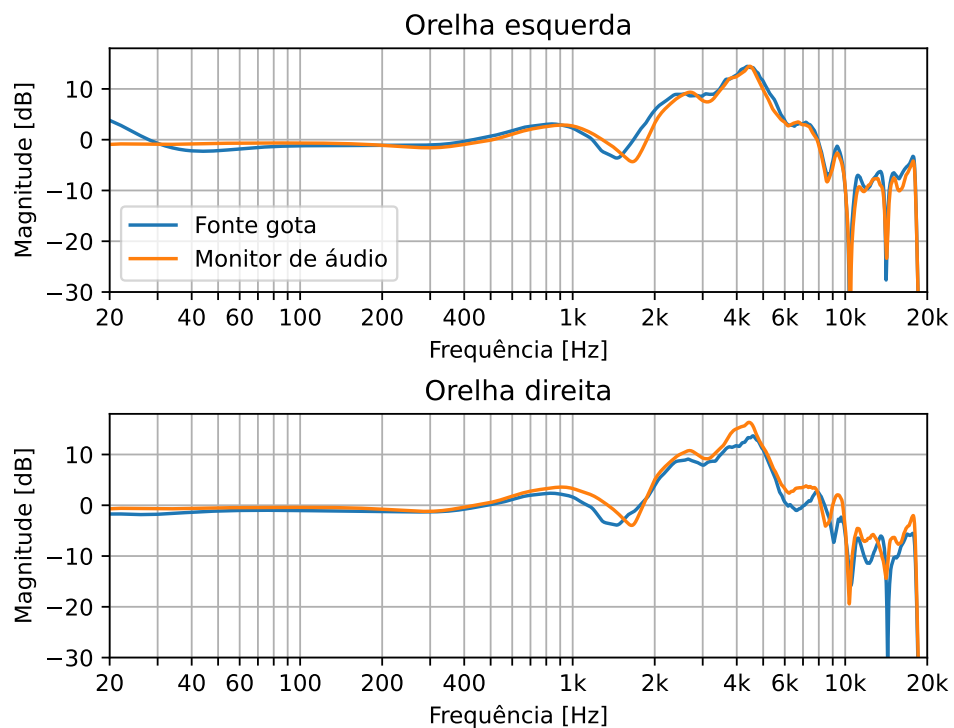
² Devido à similaridade observada entre orelha direita e esquerda, para os dados subsequentes serão apresentados somente os resultados para orelha esquerda ou biauriculares equivalentes.

Figura 10 – Respostas impulsivas e FRF entre fonte gota e receptor (microfone e HATS) com aplicação do janelamento. A janela é representada em escala para melhor visualização.



Fonte: Autor.

Figura 11 – Comparação das HRTFs utilizando as duas fontes sonoras.

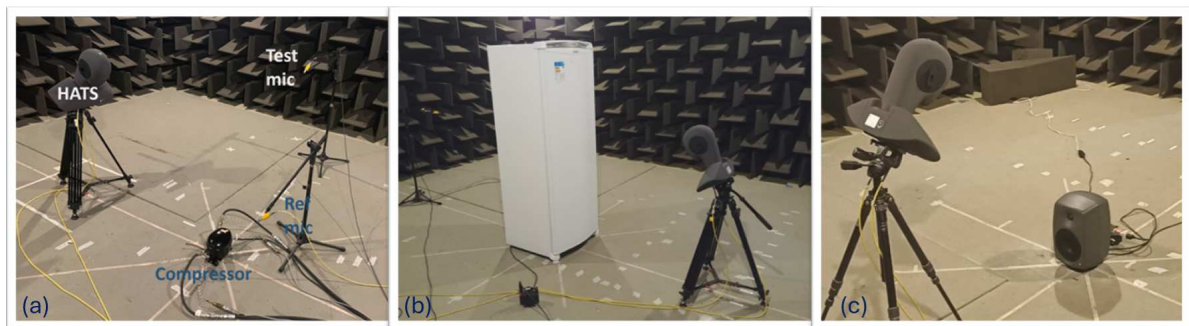


Fonte: Autor.

3.2 FONTES SONORAS

Para comparar sinais obtidos com **HATS** e microfone, três fontes sonoras foram utilizadas: um compressor hermético, um refrigerador e o monitor de áudio Genelec 8020D. As fontes sonoras foram posicionadas no centro câmara semi-anecoica e os receptores (microfone e HATS) foram posicionados a aproximadamente a um metro da fonte sonora e um metro do piso, conforme mostrado na Figura 12, o que garante campo afastado para frequências acima de 1 kHz. Para assegurar que o microfone medisse no centro da cabeça com ela ausente, pelo menos duas rodadas de medição foram necessárias. Assim, os detalhes de procedimento para cada fonte sonora será descrito a seguir.

Figura 12 – Fontes sonoras posicionadas na câmara semi-anecoica: (a) compressor, (b) refrigerador, (c) monitor de áudio.



Fonte: Autor.

3.2.1 Compressor

A primeira fonte utilizada foi um compressor de velocidade única operando a 3600 rpm conectado a um refrigerador posicionado fora da câmara. Essa conexão foi feita utilizando mangueiras longas, o que permitiu a medição acústica somente do compressor em condições nominais de operação. O posicionamento dos receptores é mostrado na Figura 12(a). Para monitorar as condições de operação, sensores de pressão foram instalados nas linhas de sucção e descarga do compressor. Além disso, um microfone de referência de campo livre foi posicionado próximo à fonte sonora para avaliar a sua estabilidade operacional entre as medições (Ref. Mic. na Figura 12(a)).

Na primeira rodada de medição, HATS e microfone foram posicionados conforme mostrado na Figura 12(a). Na segunda rodada, a posição do HATS e microfone teste foram intercambiadas com bastante cuidado para assegurar a posição do microfone no centro da cabeça.

3.2.2 Refrigerador

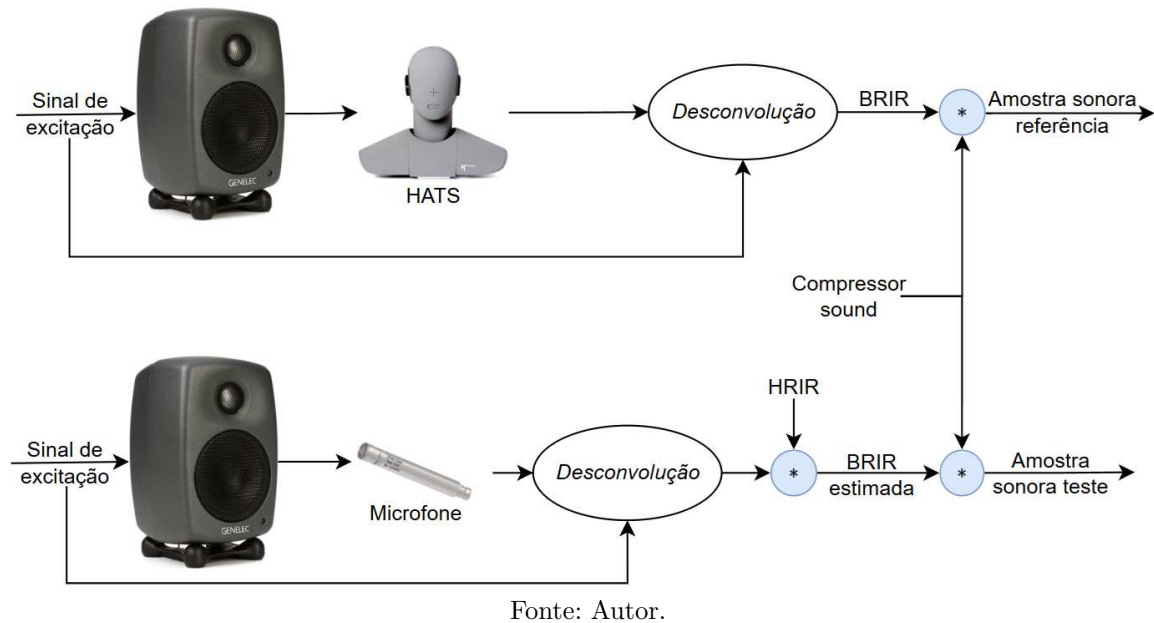
A segunda fonte sonora utilizada foi um refrigerador de porta única em sua configuração original (Figura 12(b)). Nesse caso, também foi utilizado o mesmo microfone de referência para monitorar a estabilidade operacional entre as rodadas. HATS e microfone teste foram posicionados na parte frontal e traseira do refrigerador, respectivamente, e intercambiados na segunda rodada.

3.2.3 Monitor de áudio

As fontes sonoras descritas anteriormente são sistemas de refrigeração, ou seja, fontes mecânicas. Nesse caso, a estabilidade operacional pode ser um desafio, visto que as condições de operação podem variar entre as rodadas de medição. Para contornar esse problema, uma abordagem é utilizar uma fonte sonora eletroacústica para excitar o ambiente.

Logo, o monitor de áudio foi utilizado para fazer as medições e a repetibilidade do comportamento da fonte sonora pode ser desconsiderada (Figura 12c). Além disso, essa abordagem viabiliza a avaliação de repetibilidade das medições em função do reposicionamento dos receptores. Para avaliar esse efeito, 10 reposicionamentos de HATS e microfone foram realizados, sendo realizadas 3 medições sequenciais para cada reposicionamento. Logo, foram obtidos 30 sinais para cada receptor. A fonte sonora não foi inclinada para que fosse representada uma montagem típica de medição de potência sonora.

Figura 13 – Diagrama ilustrando o processo de comparação entre as medições com HATS e microfone.



Nesse caso, a FRF entre sinal de excitação (varredura exponencial) e sinal gravado foi obtida. Além da possibilidade de realizar essa análise, outra vantagem em utilizar fonte eletroacústica é a possibilidade de reproduzir qualquer sinal desejado. Para isso, as respostas impulsivas biauriculares foram obtidas a partir das FRFs mostradas na Figura 16. Esse processo é ilustrado mais detalhadamente na Figura 13, em que os painéis superior e inferior mostram a aquisição das respostas impulsivas biauriculares (BRIR) com HATS e microfone, respectivamente. A BRIR estimada com microfone é obtida a partir da convolução da RI com esse receptor e a HRIR obtida a partir da medição da HRTF (Figura 9).

Todas as análises deste estudo foram restritas a sinais estacionários, de forma a simplificar o estudo e evitar a necessidade de tratar transientes e variações temporais.

3.3 RESULTADOS OBTIDOS

Esta seção apresenta os resultados objetivos deste estudo. Inicialmente, são mostradas as comparações espectrais entre os sinais medidos com HATS e com microfone, seguidas da análise das métricas psicoacústicas extraídas desses sinais.

3.3.1 Comparação espectral entre as gravações obtidas com HATS e microfone.

As Figuras 14 e 15 mostram os espectros dos sinais biauriculares do refrigerador e do compressor, respectivamente, medidos com HATS e microfone. Pequenas diferenças espectrais podem ser observadas entre os sinais obtidos para ambas as fontes sonoras em frequências de até 1 kHz. Essas diferenças são provavelmente atribuídas à baixa SNR durante as medições, bem como à variabilidade operacional inerente às fontes. Para o compressor, a menor pressão sonora da fonte contribui para uma SNR ainda mais limitada, o que reforça essas discrepâncias em baixas frequências.

As curvas identificadas como ruído do HATS e ruído do microfone representam os níveis de ruído residual medidos com o HATS e o microfone, respectivamente, na ausência da fonte sonora primária. Da mesma forma, nos gráficos inferiores, a curva rotulada como ruído do microfone de referência refere-se ao ruído ambiental capturado pelo microfone de referência.

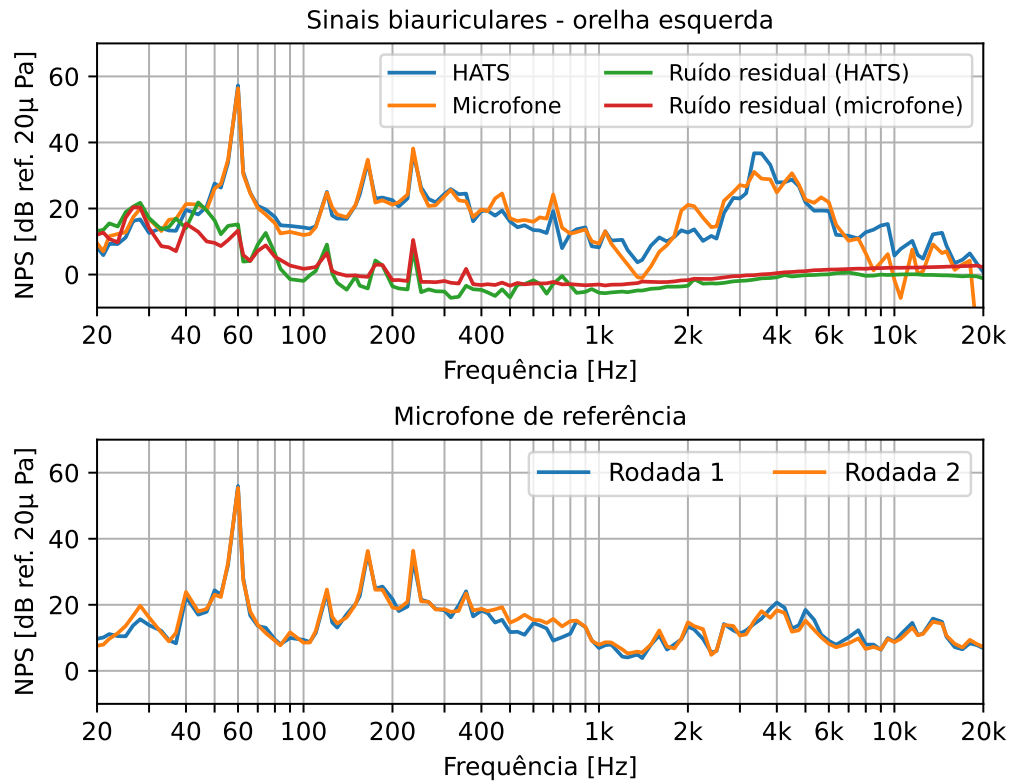
Acima de 1 kHz, observam-se diferenças de magnitude de até 10 dB entre os sinais medidos com HATS e microfone. Essas diferenças não podem ser explicadas apenas pela SNR ou pela variabilidade operacional. Nesse caso, a HRTF utilizada para caracterizar o sistema fonte–receptor apresenta limitações, principalmente porque não se pode presumir que o campo sonoro incidente sobre o ouvinte seja composto unicamente por ondas planas. Essa limitação decorre das características de direcionalidade de cada radiador e da interação das ondas sonoras com o piso da câmara, que modificam a distribuição espacial do campo acústico em relação às condições ideais de campo livre. Isso compromete a caracterização precisa do sistema e explica as discrepâncias observadas nas faixas de frequência mais altas.

Por fim, o sinal do compressor apresenta maiores variações em comparação com o do refrigerador, o que pode ser explicado por dois fatores principais. O primeiro é a menor SNR, devido à menor potência acústica da fonte. A potência sonora radiada é menor que a do refrigerador, principalmente devido à menor superfície e às características inerentes do compressor.

O segundo fator é a maior variabilidade operacional, que impacta diretamente a estabilidade do sinal medido. Isso é influenciado pela configuração experimental, que inclui tubulações longas e múltiplas conexões. Esses aspectos tornam os dados do compressor

mais suscetíveis à variabilidade espectral tanto em baixas quanto em altas frequências.³

Figura 14 – Espectros de nível de pressão sonora (SPL) em bandas de 1/12 de oitava dos sinais registrados por HATS e por microfone convoluído com a HRIR, utilizando a geladeira como fonte sonora. Os espectros foram obtidos a partir da FFT de uma amostra de 3 s do sinal, correspondente ao trecho apresentado aos jurados nos testes subjetivos.

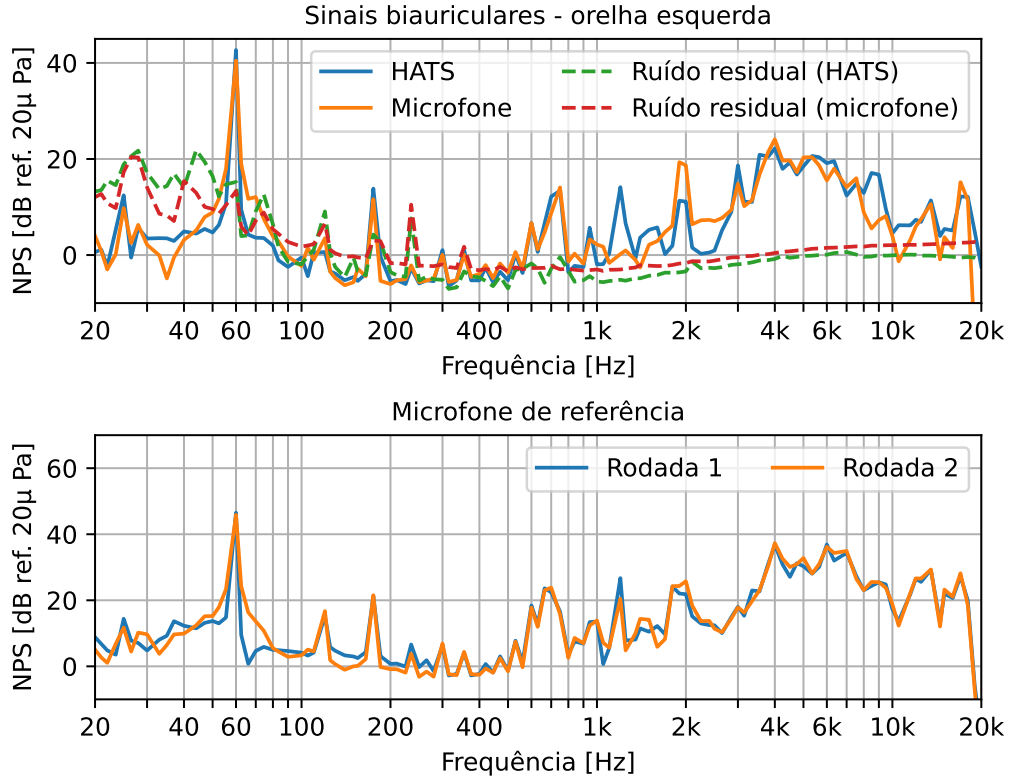


Fonte: Autor.

A Figura 16 mostra a FRF média e o respectivo desvio padrão e SNR das FRFs biauriculares obtidas utilizando HATS e microfone. A faixa de variação da SNR é de ± 2 desvios padrão, sendo omitidos os resultados do HATS devido à forte similaridade com os dados do microfone.

³ A representação em 1/12 de oitava foi adotada para aumentar a resolução espectral e evidenciar componentes tonais de banda estreita além de manter coerência com o método de correção espectral que será discutido na Seção 3.5.2.

Figura 15 – Espectros SPL de banda de 1/12 de oitava dos sinais de HATS e microfone convoluído com HRIR utilizando o compressor como fonte sonora. Os espectros foram obtidos a partir da FFT de uma amostra de 3 s do sinal, da mesma forma que na Figura 14.

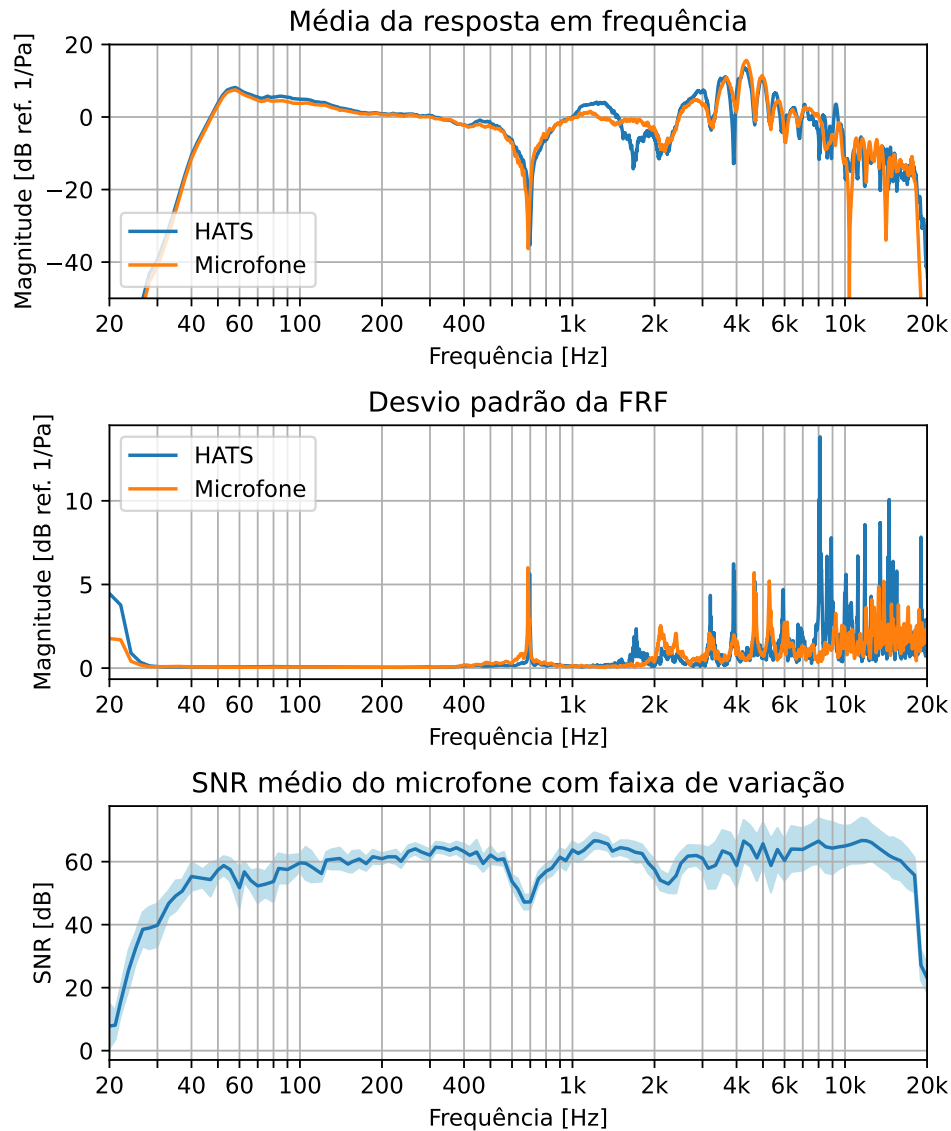


Fonte: Autor.

As FRFs médias do HATS e do microfone apresentam boa concordância em termos de espectros de magnitude. No entanto, observa-se uma antirressonância em torno de 700 Hz, atribuída ao efeito dipolo decorrente da interação entre a fonte e o piso. Esse fenômeno pode ser modelado pela introdução de uma fonte virtual espelhada no piso, levando à interferência destrutiva em certas frequências, conforme detalhado em Allen e Berkley (1979). A interferência se torna significativa quando o comprimento de onda é menor que a distância entre o piso e o centro acústico da fonte, resultando no cancelamento parcial de componentes espectrais.

Entre 1 kHz e 2 kHz, nota-se ainda uma queda no espectro de magnitude do HATS que não aparece no microfone. Essa discrepância provavelmente decorre da influência combinada da cabeça e da fonte interagindo com o piso, gerando ondas não planas que chegam ao receptor. Como a HRTF assume a propagação de ondas planas, tais efeitos não são capturados, o que ressalta uma limitação na modelagem precisa da interação acústica entre HATS e campo sonoro incidente.

Figura 16 – Média, desvio padrão e SNR das FRFs entre fonte e receptor (HATS e microfone convoluído com HRIR) obtidas com as medições do monitor de áudio. O desvio padrão foi calculado a partir de 30 amostras: 10 reposicionamentos e 3 repetições por reposicionamento.



Fonte: Autor.

A análise do desvio padrão das [FRFs](#) mostra que a variabilidade permanece baixa para a maior parte da faixa de frequências de interesse, aumentando de forma significativa apenas nas regiões associadas às interferências destrutivas, isso é, nos anti-picos da [FRF](#), por exemplo em 700 Hz. Nessas frequências, pequenas diferenças de posicionamento ou de acoplamento geométrico entre fonte, receptor e piso resultam em variações expressivas, reforçando que a instabilidade espectral está diretamente vinculada a fenômenos de cancelamento. Porém, estudos de inteligibilidade da fala apontam que ouvintes humanos tendem a ser mais sensíveis a picos espectrais do que a vales equivalentes ([Bucklein,](#)

1981; Masiero; Fels, 2011), o que implica que ressonâncias indesejadas podem ter impacto perceptual mais relevante do que atenuações.

3.4 ANÁLISE DE VARIABILIDADE DAS MÉTRICAS PSICOACÚSTICAS

A variabilidade dos parâmetros psicoacústicos foi analisada considerando as três fontes sonoras deste estudo. Para o compressor e o refrigerador (fontes mecânicas), os sinais gravados foram divididos em blocos não sobrepostos de 3 segundos, com janela retangular, e as métricas calculadas em cada bloco. Como não houve reposicionamento do **HATS** ou do microfone, a dispersão observada reflete apenas a instabilidade operacional dessas fontes.⁴

No caso do monitor de áudio, a análise buscou isolar o efeito do reposicionamento dos receptores. Para isso, foram selecionados três sinais de compressores de rotação variável (2400, 3000 e 4500 rpm), previamente gravados em câmara semi-anecoica. Sinais de refrigerador não foram utilizados nesta etapa, pois considerou-se acusticamente incoerente convoluir o som de uma fonte de grandes dimensões com a resposta impulsiva da sala excitada por uma fonte significativamente menor (o monitor de áudio). Trechos de 3 segundos de cada gravação do compressor foram então convoluídos com as 30 respostas impulsivas obtidas com **HATS** e microfone, permitindo quantificar exclusivamente a variabilidade decorrente do reposicionamento.

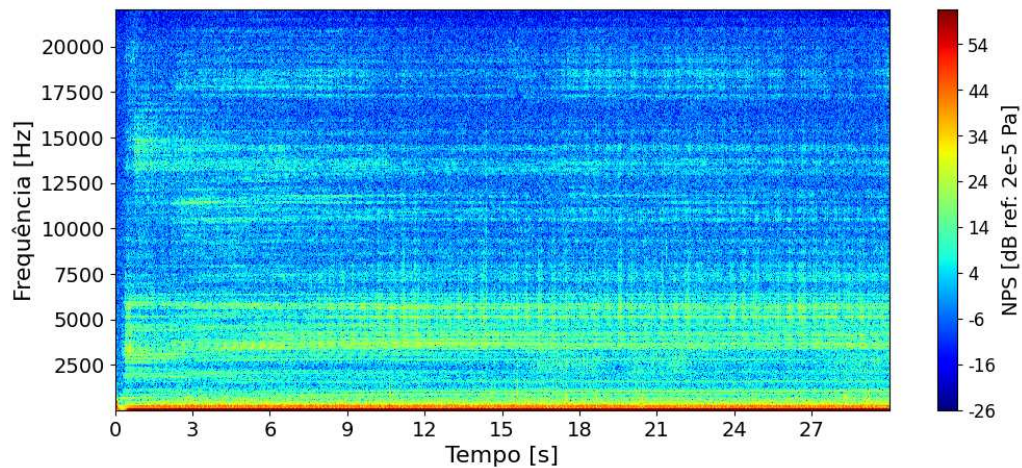
As Figuras 17 a 21 apresentam os espectrogramas dos sinais analisados (somente gravações com microfone sem convolução com HRIR), sendo considerados 30 segundos para as fontes mecânicas (Figuras 17 e 18) e 3 segundos para a avaliação do reposicionamento (Figuras 19, 20 e 21). Somente os espectrogramas para os sinais de microfone convoluído com HRIR e para o lado esquerdo são apresentados devido à semelhança observada entre os demais dados.

O refrigerador (Figura 17) apresenta um transiente inicial associado à partida do compressor, além de maior variabilidade espectral ao longo do tempo, configurando-se como o sinal menos estacionário. Também se observa a presença de componentes de banda larga sobrepostas às tonais, atribuídas principalmente ao escoamento do fluido refrigerante.

⁴ O intervalo de 3 s foi adotado por corresponder à duração típica dos trechos apresentados aos participantes durante os testes subjetivos. Análises complementares com diferentes tamanhos de bloco (3, 6, 12 e 18 s) mostraram que os valores médios e a tendência temporal das métricas convergem a partir de 3 s, conforme apresentado no Apêndice A.

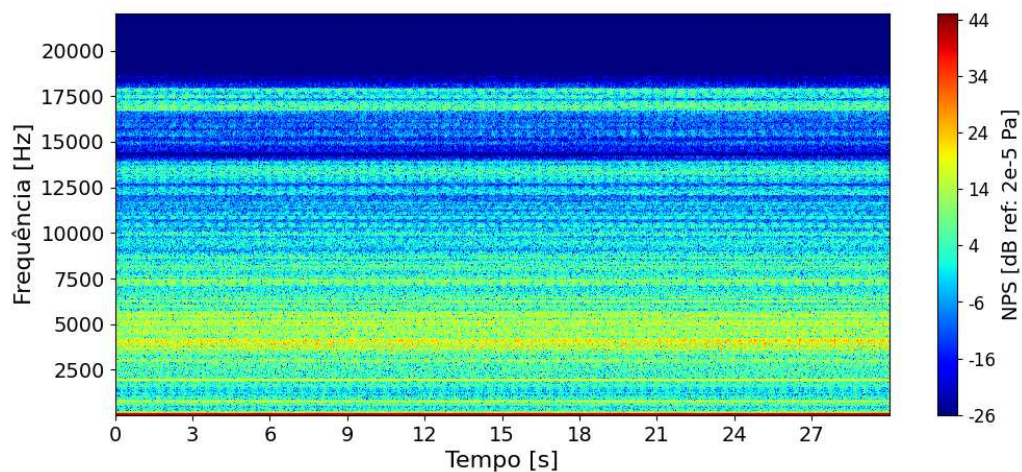
Já os sinais do compressor (Figuras 18, 19, 20 e 21) mostram espectrogramas visualmente estáveis, sem alterações significativas ao longo do tempo. Essa consistência permite tratá-los como estacionários, justificando o uso de métricas espectrais e psicoacústicas médias sem necessidade de análises variantes no tempo.

Figura 17 – Espectrograma do sinal de refrigerador utilizado na análise de variabilidade operacional. Blocos de 1024 amostras com janela *hanning* e sobreposição de 50%.



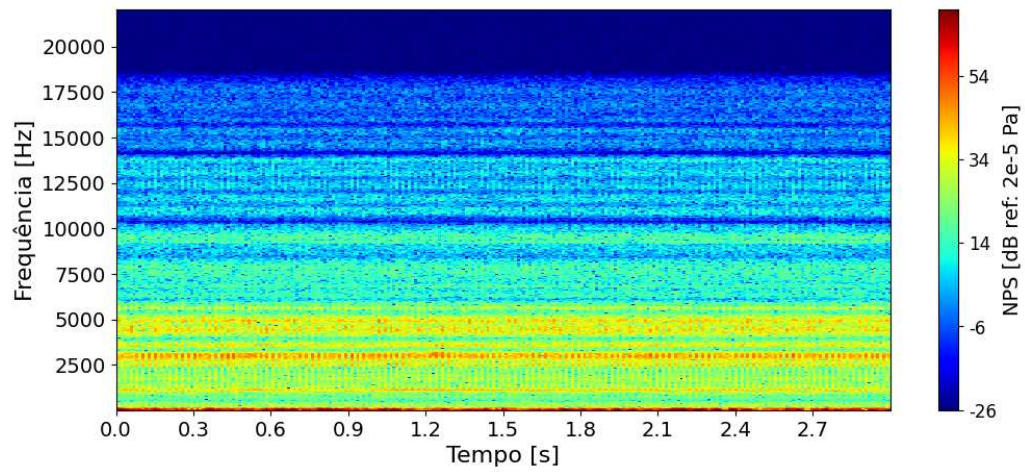
Fonte: Autor.

Figura 18 – Espectrograma do sinal de compressor utilizado na análise de variabilidade operacional. Blocos de 1024 amostras com janela *hanning* e sobreposição de 50%.



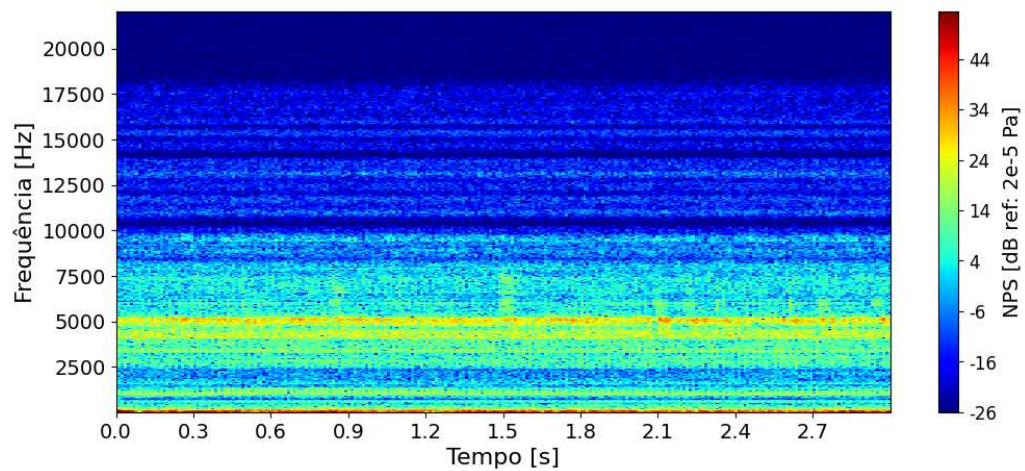
Fonte: Autor.

Figura 19 – Espectrograma do sinal de compressor (2400 rpm) utilizado na análise de variabilidade de reposicionamento dos receptores. Blocos de 1024 amostras com janela *hanning* e sobreposição de 50%.



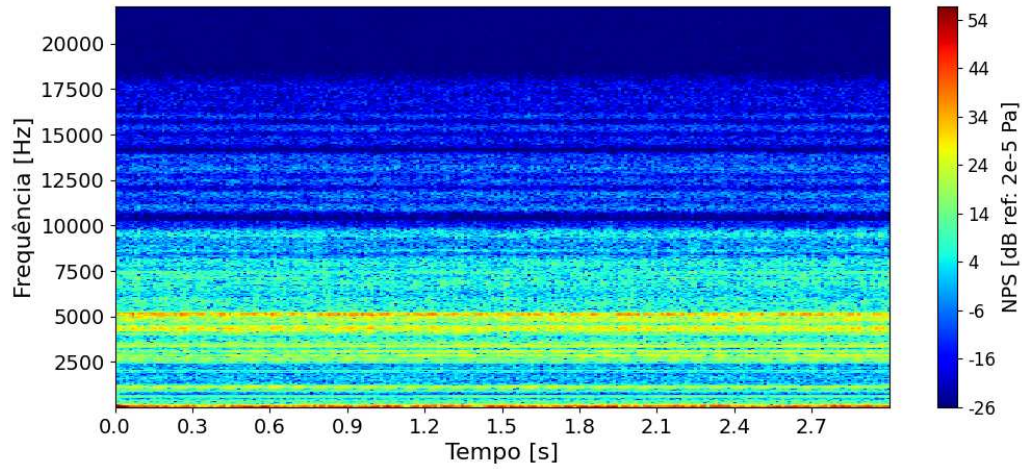
Fonte: Autor.

Figura 20 – Espectrograma do sinal de compressor (3000 rpm) utilizado na análise de variabilidade de reposicionamento dos receptores. Blocos de 1024 amostras com janela *hanning* e sobreposição de 50%.



Fonte: Autor.

Figura 21 – Espectrograma do sinal de compressor (4500 rpm) utilizado na análise de variabilidade de reposicionamento dos receptores. Blocos de 1024 amostras com janela *hanning* e sobreposição de 50%.



Fonte: Autor.

3.4.1 Resultados para monitor de áudio (fonte eletroacústica)

As Tabelas 1 a 3 apresentam os resultados dos testes estatísticos para os sinais obtidos com o monitor de áudio, focando no efeito do reposicionamento dos receptores (HATS e microfone convoluído com HRIR). Os testes de Shapiro-Wilk confirmaram que, na maioria dos casos, as distribuições das métricas não seguem uma distribuição normal (valores de $p < 0,05$). Portanto, optou-se pelo uso do teste não-paramétrico de Mann-Whitney para comparar as distribuições entre os sinais de HATS e de microfone convoluído com HRIR. Cada áudio analisado teve duração de 3 segundos, correspondente à utilizada nos testes de júri subsequentes, assegurando consistência entre as análises objetivas e subjetivas.

Os testes de Mann-Whitney indicaram que as diferenças para *loudness* e *roughness* foram estatisticamente significativas ($p \ll 0,05$) em todas as rotações do compressor. Para *sharpness*, a diferença não foi estatisticamente significativa a 2400 rpm ($p > 0,05$), mas tornou-se significativa nas rotações mais altas.

No entanto, a análise da significância perceptual, comparando as diferenças medianas com as DLOs, revela uma tendência mais complexa. Para a métrica de *sharpness*, a diferença entre as medianas permaneceu abaixo do DLO de 0,08 acum para todas as três rotações. Em contraste, para *loudness* e *roughness*, o impacto perceptual mostrou-se dependente da rotação do compressor. A 2400 rpm, as diferenças de 1,37 sone para *loudness* e 0,15 asper para *roughness* excederam seus respectivos DLOs (0,5 sone e 0,04 asper)

estabelecidas por You e Jin (2008), sugerindo uma diferença audível. Contudo, a 3000 rpm e 4500 rpm, as diferenças para estes mesmos parâmetros diminuíram, ficando abaixo dos limiares.

Em síntese, os resultados sugerem que, para a análise de *sharpness*, o microfone convoluído com a HRIR adequada pode substituir o HATS de forma adequada, pois as diferenças se mantêm muito próximas ou abaixo das DLOs utilizadas como referência. Para *loudness* e *roughness*, contudo, a intercambialidade entre os receptores é condicional: as discrepâncias se mostraram relevantes para o sinal de baixa rotação (2400 rpm), mas tendem a se tornar imperceptíveis à medida que a velocidade do compressor aumenta e o espectro do ruído se altera.

Tabela 1 – Resultados dos testes de Shapiro-Wilk e Mann-Whitney para *loudness*, *sharpness* e *roughness* entre sinais de HATS e microfone convoluído com HRIR, considerando 10 substituições do receptor, para o compressor a 2400 rpm. A última linha apresenta as diferenças medianas para cada parâmetro.

	Métrica	Loudness		Sharpness		Roughness	
	Receptor	HATS	Mic.	HATS	Mic.	HATS	Mic.
Teste de Shapiro Wilk	Valor p	1.6E-02	2.3E-03	3.6E-04	7.2E-04	3.8E-05	1.8E-04
	Estatística	0.91	0.87	0.83	0.83	0.78	0.78
Teste de Mann-Witney	Valor p	6.4E-11		1.5E-01		6.5E-11	
	Estatística	841		813		841	
Diferença entre medianas		1.37 sone		0.01 acum		0.15 asper	

Tabela 2 – Resultados dos testes de Shapiro-Wilk e Mann-Whitney para *loudness*, *sharpness* e *roughness* entre sinais de HATS e microfone convoluído com HRIR, considerando 10 substituições do receptor, para o compressor a 3000 rpm. A última linha apresenta as diferenças medianas para cada parâmetro.

	Métrica	Loudness		Sharpness		Roughness	
	Receptor	HATS	Mic.	HATS	Mic.	HATS	Mic.
Teste de Shapiro Wilk	Valor p	3.0E-02	8.1E-03	2.9E-02	7.6E-05	8.1E-01	6.5E-05
	Estatística	0.92	0.90	0.92	0.92	0.98	0.98
Teste de Mann-Witney	Valor p	8.4E-06		6.5E-11		6.5E-11	
	Estatística	841		813		841	
Diferença entre medianas		0.13 sone		-0.07 acum		0.04 asper	

Tabela 3 – Resultados dos testes de Shapiro-Wilk e Mann-Whitney para *loudness*, *sharpness* e *roughness* entre sinais de HATS e microfone convoluído com HRIR para o compressor a 4500 rpm. A última linha apresenta as diferenças medianas para cada parâmetro.

	Métrica	Loudness		Sharpness		Roughness	
	Receptor	HATS	Mic.	HATS	Mic.	HATS	Mic.
Teste de Shapiro Wilk	Valor p	1.2E-02	3.9E-04	3.9E-03	6.1E-03	1.4E-02	8.0E-02
	Estatística	0.90	0.84	0.88	0.88	0.91	0.91
Teste de Mann-Witney	Valor p	1.2E-05		6.5E-11		1.6E-05	
	Estatística	702		0		698	
Diferença entre medianas		0.2 sone		-0.06 acum		0.01 asper	

A análise gráfica das distribuições dos parâmetros (Figuras 22 a 24) reforça e detalha os achados estatísticos previamente apresentados.

Para o parâmetro de *loudness*, a dependência da rotação do compressor é evidente. A 2400 rpm (Figura 22), as distribuições para HATS e microfone são claramente separadas, e a diferença entre as medianas (1,37 sone) excede significativamente a DLO de 0,5 sone. Contudo, para 3000 e 4500 rpm (Figuras 23 e 24), as distribuições se aproximam e apresentam considerável sobreposição, com as diferenças de medianas (0,13 e 0,20 sone, respectivamente) permanecendo abaixo das DLOs estabelecidas na literatura.

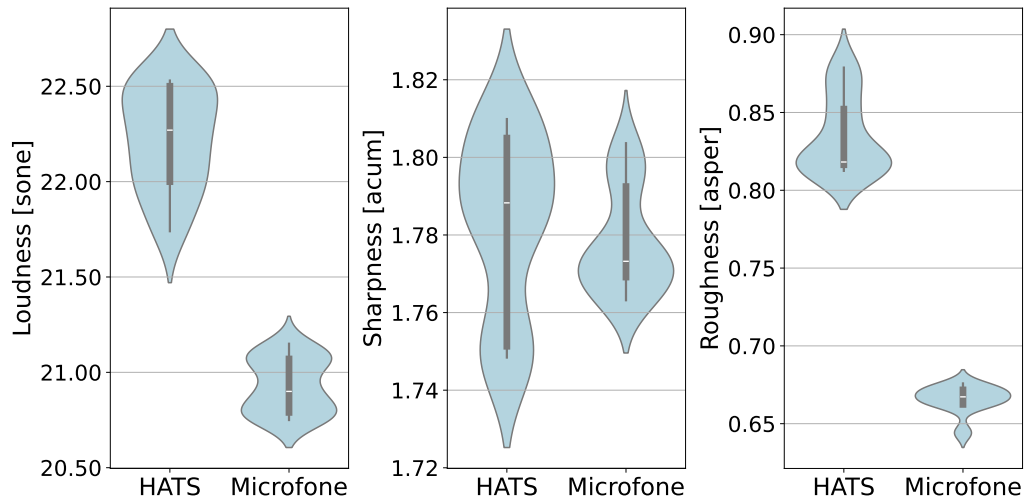
No caso da *sharpness*, os gráficos de violino confirmam a baixa sensibilidade do parâmetro para pequenas variações de posicionamento. Embora existam pequenas discrepâncias nas medianas, as distribuições entre os dois receptores se sobrepõem significativamente em todas as três rotações. As diferenças mantêm-se abaixo da DLO de 0,08 acum, reforçando a conclusão de que, para esse parâmetro, HATS e microfone convoluído com HRIR fornecem resultados equivalentes.

Para *roughness*, observa-se uma tendência similar à do *loudness*. A 2400 rpm, as distribuições estão bem separadas e a diferença de 0,15 asper excede em muito a DLO de 0,04 asper. Para o compressor a 3000 rpm, a diferença torna-se limítrofe (0,04 asper), e a 4500 rpm a diferença de 0,01 asper é imperceptível, o que é corroborado pela aproximação e sobreposição dos gráficos de violino.

Em conjunto, os resultados gráficos e estatísticos indicam uma conclusão mais sutil. Enquanto o microfone pode ser considerado um substituto confiável do HATS para análises de *sharpness*, a equivalência para *loudness* e *roughness* é condicional. As diferenças entre os receptores são potencialmente audíveis para o sinal de baixa rotação, mas tornam-se imperceptíveis à medida que a velocidade do compressor aumenta e o espectro do ruído

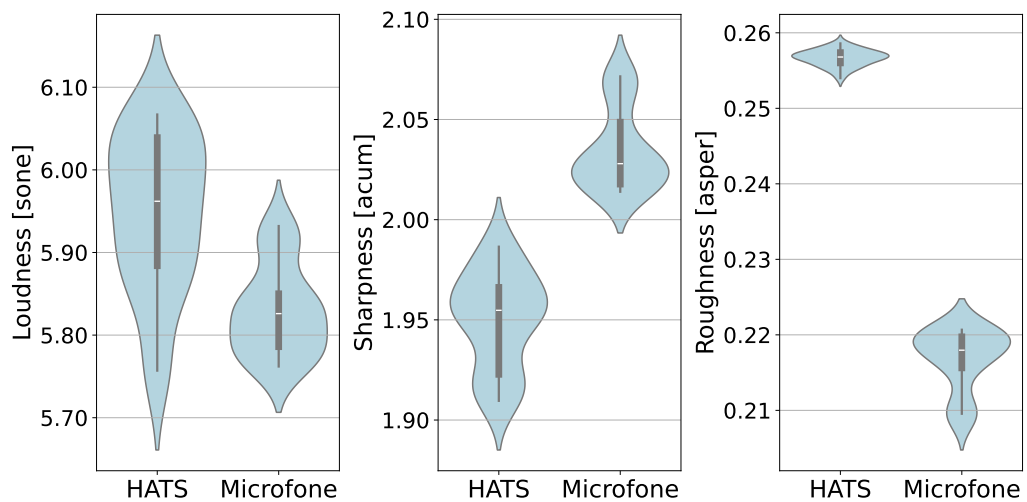
se altera, demonstrando que a intercambialidade dos métodos de medição depende do conteúdo espectral do sinal analisado.

Figura 22 – Gráficos de violino de parâmetros no reposicionamento dos receptores (HATS e microfone convoluído com HRIR) utilizando o sinal de compressor a 2400 rpm.



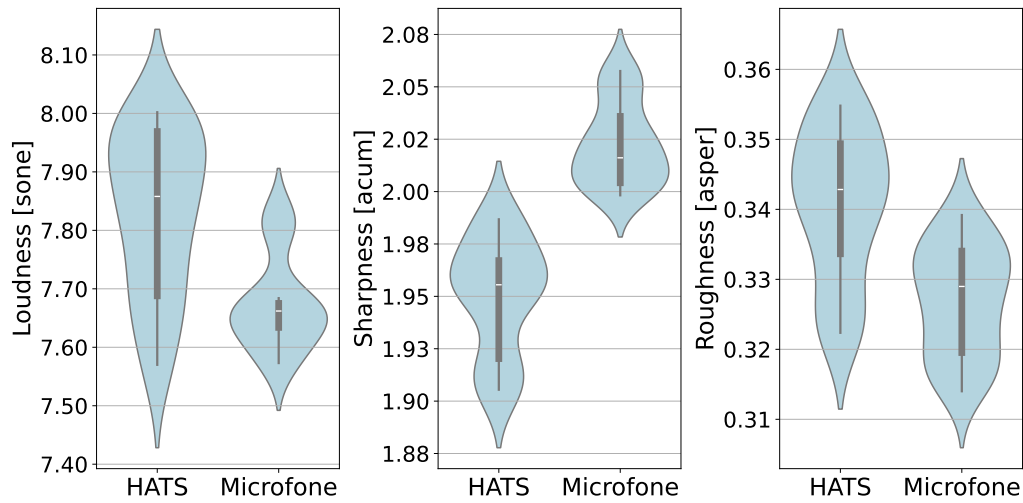
Fonte: Autor.

Figura 23 – Gráficos de violino de parâmetros no reposicionamento dos receptores (HATS e microfone convoluído com HRIR) utilizando o sinal de compressor a 3000 rpm.



Fonte: Autor.

Figura 24 – Gráficos de violino de parâmetros no reposicionamento dos receptores (HATS e microfone convoluído com HRIR) utilizando o sinal de compressor a 4500 rpm.



Fonte: Autor.

3.4.2 Resultados para fontes mecânicas

Os resultados estatísticos para as fontes mecânicas são apresentados nas Tabelas 4 e 5, enquanto os gráficos de violino das Figuras 25 e 26 ilustram a distribuição dos parâmetros. Para garantir a consistência da análise, e dado que algumas das distribuições não passaram no teste de normalidade de Shapiro-Wilk, optou-se novamente pelo uso do teste não-paramétrico de Mann-Whitney para todas as comparações.

Tabela 4 – Resultados dos testes de Shapiro-Wilk e Mann-Whitney para *loudness*, *sharpness* e *roughness* entre sinais de HATS e microfone convoluído com HRIR, considerando 10 substituições do receptor, para o sinal de refrigerador. A última linha apresenta as diferenças medianas para cada parâmetro.

	Métrica	Loudness		Sharpness		Roughness	
	Receptor	HATS	Mic.	HATS	Mic.	HATS	Mic.
Teste de Shapiro Wilk	Valor p	4.4E-01	5.7E-02	7.8E-01	2.9E-02	2.1E-01	1.3E-01
	Estatística	0.97	0.93	0.98	0.92	0.95	0.95
Teste de Mann-Witney	Valor p	7.4E-10		1.2E-04		4.1E-07	
	Estatística	867		190		793	
Diferença entre medianas		0.31 sone		-0.01 acum		0.01 asper	

Tabela 5 – Resultados dos testes de Shapiro-Wilk e Mann-Whitney para *loudness*, *sharpness* e *roughness* entre sinais de HATS e microfone convoluído com HRIR, considerando 10 substituições do receptor, para o sinal de compressor. A última linha apresenta as diferenças medianas para cada parâmetro.

	Métrica	Loudness		Sharpness		Roughness	
	Receptor	HATS	Mic.	HATS	Mic.	HATS	Mic.
Teste de Shapiro Wilk	Valor p	2.3E-01	1.7E-02	1.6E-01	3.3E-01	7.8E-03	5.1E-01
	Estatística	0.95	0.91	0.95	0.96	0.90	0.97
Teste de Mann-Witney	Valor p	1.3E-02		3.0E-11		1.4E-07	
	Estatística	619		900		93	
Diferença entre medianas		0.03 sone		0.14 acum		-0.02 asper	

No caso do refrigerador (Tabela 4 e Figura 25), embora o teste de Mann-Whitney tenha apontado diferenças estatisticamente significativas, a análise das diferenças entre medianas mostra que elas não atingem magnitude para serem perceptualmente relevantes. As diferenças para *loudness* (0,31 sone), *sharpness* (0,01 acum) e *roughness* (0,01 asper) permaneceram abaixo de seus respectivos DLOs.

Figura 25 – Gráficos de violino de parâmetros para avaliação de estabilidade operacional do do refrigerador para os dois receptores (HATS e microfone convoluído com HRIR).

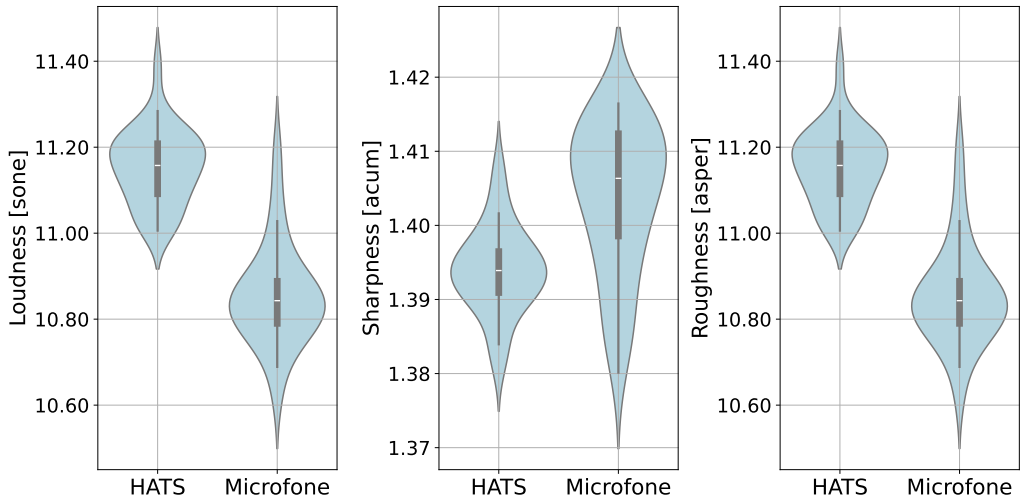
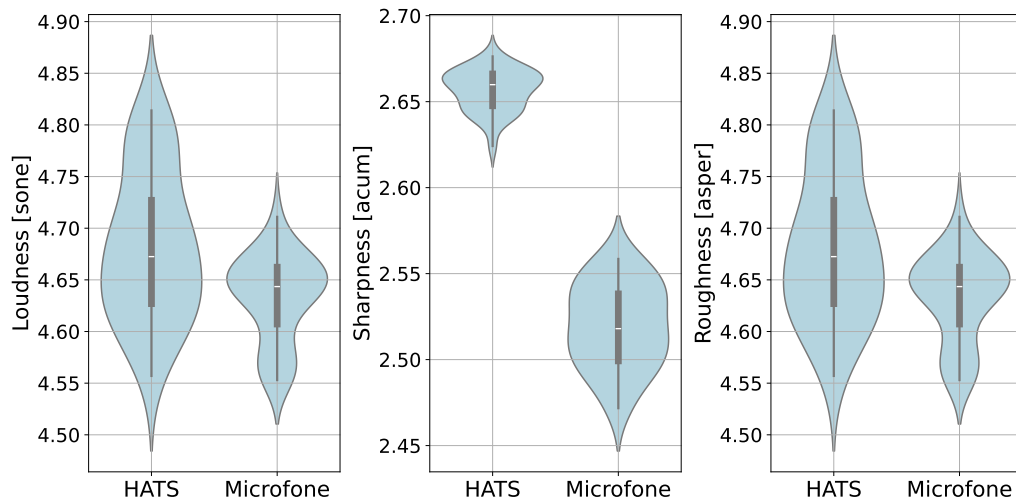


Figura 26 – Gráficos de violino de parâmetros para avaliação de estabilidade operacional do compressor para os dois receptores (HATS e microfone convoluído com HRIR).



Para o compressor (Tabela 5 e Figura 26), o comportamento é distinto. Enquanto as diferenças para *loudness* (0,03 sone) e *roughness* (0,02 asper) também se mantiveram abaixo do limiar perceptível, a diferença para *sharpness* foi de 0,14 acum. Este valor excede o DLO de 0,08 acum, sugerindo que, para esta fonte, a diferença na percepção de agudeza entre o HATS e o microfone é potencialmente audível.

Os gráficos de violino corroboram essa interpretação. Para o refrigerador (Figura 25), as distribuições para todos os parâmetros se sobrepõem consideravelmente. Já para o compressor (Figura 26), nota-se uma clara separação entre as distribuições de *sharpness*, confirmando visualmente a diferença significativa apontada pelos dados.

Em síntese, os resultados indicam que a equivalência perceptual entre os receptores depende da fonte sonora. Para o ruído do refrigerador, HATS e microfone convoluído com HRIR forneceram resultados equivalentes nos três parâmetros avaliadas. Contudo, para o ruído do compressor, a diferença em *sharpness* mostrou-se relevante, indicando que, para esta condição, o uso do microfone em substituição ao HATS deve ser interpretado com cautela.

3.5 AVALIAÇÃO SUBJETIVA

A partir dos resultados objetivos apresentados na seção anterior, foi conduzida uma série de avaliações subjetivas com o intuito de verificar em que medida as diferenças observadas entre os sinais gravados com HATS e microfone convoluído com HRIR são perceptíveis. Essa etapa foi organizada em duas fases:

- **PEST**: utilizado para estimar fatores de correção espectral críticos (α_{crit}) que minimizam a percepção de diferenças entre os sinais gravados com HATS e microfone convoluído com HRIR.
- **MUSHRA**: aplicado para quantificar as diferenças perceptivas com diferentes fatores de correção α , permitindo relacionar as respostas do júri com os parâmetros psicoacústicos.

3.5.1 Amostra sonora

As amostras utilizadas nos testes subjetivos foram retiradas do mesmo conjunto de dados da seção anterior, garantindo alinhamento entre a análise objetiva e a perceptiva. Em todos os casos foram selecionados trechos estacionários e pareados dos sinais gravados com HATS e microfone.

A Tabela 6 lista os cinco pares avaliados pelo júri: pares (1) refrigerador e (2) compressor, ou seja, as gravações diretas das fontes mecânicas, e (3 a 5), ou seja, os estímulos obtidos por convolução das BRIRs medidas com o monitor de áudio com gravações de compressor a 2400, 3000 e 4500 rpm.

Tabela 6 – Pares de sinal HATS-microfones utilizados no teste de júri.

Par	Descrição
1	Refrigerador (fonte mecânica)
2	Compressor (fonte mecânica)
3	Monitor de áudio - compressor a 2400 rpm
4	Monitor de áudio - compressor a 3000 rpm
5	Monitor de áudio - compressor a 4500 rpm

3.5.2 Correção espectral ponderada

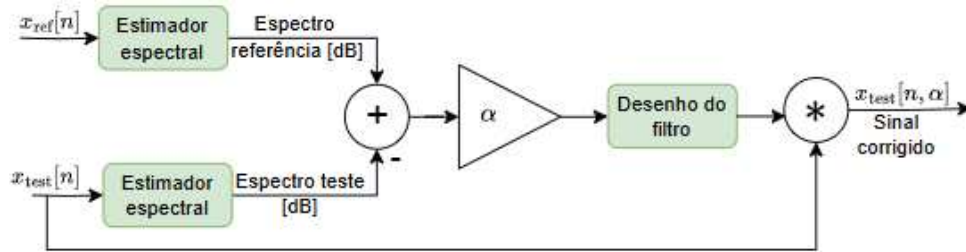
Para manipular o sinal de teste foi implementado um algoritmo de correção espectral ponderada, cujo objetivo é transformar gradativamente esse sinal de forma seu espectro de

magnitude se aproxime ou se afaste do sinal de referência. Para isso, foi definido o fator de correção α , que atua como parâmetro de ponderação no processo de ajuste espectral.

O procedimento está ilustrado no fluxograma da Figura 27. Inicialmente, são calculados os níveis sonoros espectrais em bandas de fração de oitava (em escala dB) para os sinais de referência (HATS) e de teste (microfone convoluído com HRIR). Esse processamento em bandas de oitava atua como uma etapa de suavização espectral, reduzindo flutuações locais acentuadas, como anti picos, e fornecendo uma representação melhor das diferenças entre os espectros dos sinais.

Em seguida, obtém-se a correção em função da frequência a partir da diferença entre ambos, o qual é ponderado pelo fator α . Assim, $\alpha = 0$ mantém o sinal de teste inalterado, enquanto $\alpha = 1$ o transforma completamente para coincidir com o espectro do sinal de referência. Valores intermediários de α produzem uma transição contínua entre essas duas condições. Além disso, ao se adotar valores negativos de α , o erro espectral pode ser extrapolado, possibilitando avaliar se é necessário amplificar a diferença para que o ouvinte perceba distinções entre sinais muito semelhantes.

Figura 27 – Fluxograma do processo de correção do sinal de teste em função do sinal de referência e do fator de correção α .



Fonte: Autor.

Esse mecanismo pode ser aproximado pela seguinte expressão:

$$X_{\text{test,corr}}(f, \alpha) \approx X_{\text{test}}(f) \left[\frac{X_{\text{ref}}(f)}{X_{\text{test}}(f)} \right]^\alpha, \quad (3.1)$$

em que $X_{\text{test}}(f)$ e $X_{\text{ref}}(f)$ são os espectros de teste e de referência, respectivamente. A igualdade não é exata, pois a implementação prática via filtro FIR introduz limitações relacionadas à resolução espectral e à aproximação de fase mínima. Ainda assim, a Equação (3.1) descreve de forma adequada a ideia central do processo: a criação de um contínuo unidimensional controlado por α , no qual o sinal de teste é gradualmente transformado até coincidir com o sinal de referência.

O valor de α no expoente equivale a uma ponderação linear ao representar os espectros em escala dB,

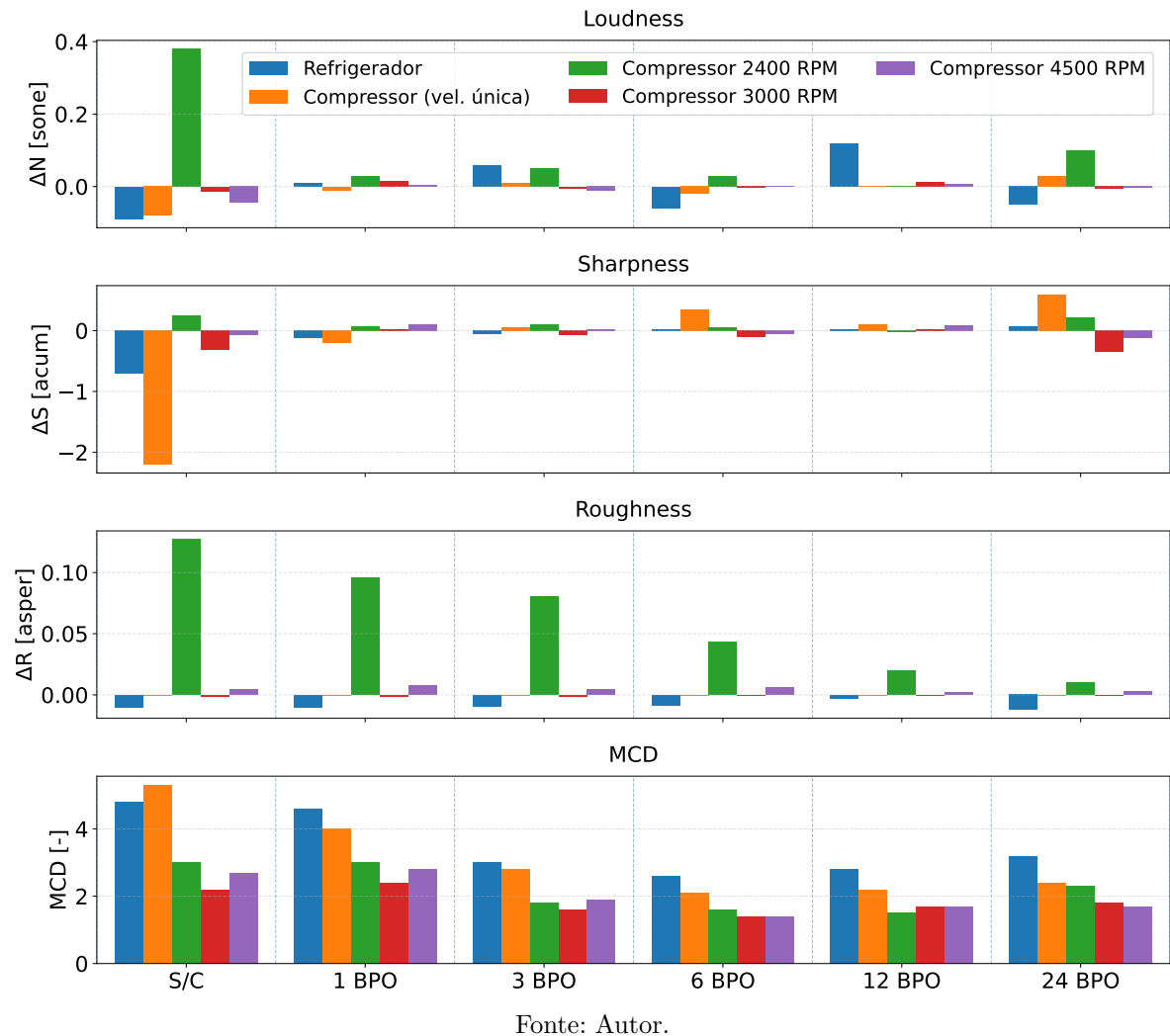
$$L_{X_{\text{test,corr}}}(f, \alpha) \approx L_{X_{\text{test}}}(f) + \alpha[L_{X_{\text{ref}}}(f) - L_{X_{\text{test}}}(f)], \quad (3.2)$$

em que $L_{X_{\text{test,corr}}}(f, \alpha)$, $L_{X_{\text{test}}}(f)$ e $L_{X_{\text{ref}}}(f)$ são os espectros de NPS do sinal corrigido, sinal de referência (HATS) e sinal teste, respectivamente.

Para garantir resolução adequada no projeto do filtro, os níveis de correção são interpolados em uma escala de frequência linear. O filtro é projetado como FIR de fase mínima, a partir da diferença espectral ponderada, utilizando a transformada de Hilbert para reconstrução (ver Seção 2.1.2.2). Essa abordagem assegura que a correção seja aplicada de forma estável e com mínima distorção de fase. O sinal corrigido é então obtido pela convolução entre o sinal de teste e o filtro projetado.

Essa formulação é particularmente interessante para comparar sinais que atravessaram cadeias acústicas distintas, como no caso de HATS e microfone. O parâmetro α atua como interpolador entre dois extremos acústicos, permitindo gerar versões intermediárias do sinal com diferenças graduais e controladas. Essas versões constituem justamente os estímulos necessários para os testes subjetivos (PEST e MUSHRA), viabilizando a avaliação perceptual da equivalência ou discrepância entre os dois métodos de captura.

Figura 28 – Análise de convergência para o fator de correção α e o número de bandas por oitava utilizadas na discretização espectral.



Para definir a resolução espectral do algoritmo de correção, foi realizada uma análise de convergência (Figura 28), avaliando-se as diferenças residuais com correção total ($\alpha = 1$). Nessa etapa, aplicou-se o algoritmo de correção espectral considerando diferentes discretizações em **Bandas por oitava (BPO)** (1, 3, 6, 12 e 24), além da condição sem correção (S/C), correspondente ao sinal de teste original captado por microfone. Para cada configuração, calcularam-se as diferenças entre o sinal de referência (HATS) e o sinal corrigido para os parâmetros de *loudness* (ΔN), *sharpness* (ΔS), *roughness* (ΔR) e para a **MCD**. Essa varredura permitiu avaliar a influência da resolução espectral na convergência dos parâmetros, identificando a partir de qual discretização as diferenças residuais se tornam mínimas.

A análise demonstra que todos os quatro parâmetros (*loudness*, *sharpness*, *roughness*

e MCD) convergem para baixas diferenças residuais à medida que a resolução espectral, definida em BPO, aumenta. Para *sharpness* e *roughness*, observa-se uma tendência clara de redução do erro com o aumento do número de bandas. De forma similar, a MCD apresenta uma diminuição consistente até a faixa de 6 a 12 BPO, ponto em que os resultados se estabilizam. A utilização de uma resolução mais fina (24 BPO) não trouxe ganhos adicionais, introduzindo pequenas divergências atribuídas a um sobreajuste (*overfitting*) das flutuações espectrais, isto é, à amplificação de irregularidades locais do espectro que não correspondem necessariamente a diferenças perceptivas reais, mas resultam do excesso de resolução aplicado durante o processo de correção.

Com base nessa análise, a discretização em 12 BPO foi adotada para todos os sinais. Esse valor apresentou um equilíbrio adequado entre resolução espectral e robustez perceptual, garantindo a convergência dos parâmetros sem introduzir instabilidades no processo de correção.

3.5.3 Implementação da PEST para controlar o fator de correção espectral

Uma adaptação do método PEST foi implementada de forma a controlar adaptativamente o fator de correção α durante o teste para estimar o fator de correção crítico de discriminação do participante α_{crit} . O processo é ilustrado no fluxograma da Figura 29. Em cada iteração são apresentados os sinais de referência $x_{\text{ref}}(t)$ e teste $x_{\text{test}}(t, \alpha)$, ou seja, microfone convoluído com HRIR e corrigido com determinado valor de α , e o participante deve avaliar os sinais como “iguais” ou “diferentes”.

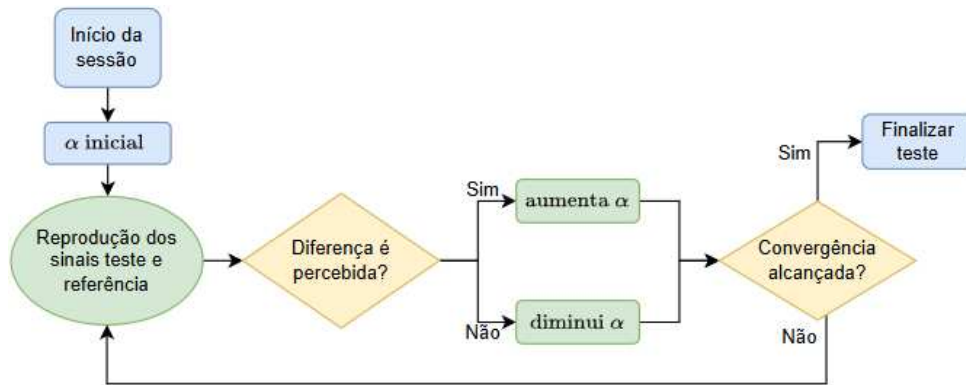
As duas principais regras do algoritmo são:

- Caso o participante julgue iguais, o algoritmo diminui α (menor correção), diferenciando gradualmente os espectros dos sinais.
- Caso julgue diferentes, o algoritmo aumenta α (maior correção), aproximando os espectros sinais.

Sempre que ocorre uma reversão de resposta (mudança de “igual” para “diferente” ou vice-versa), o passo de α é reduzido pela metade, garantindo a convergência para o respectivo α_{crit} . O valor final é tomado como a estimativa individual do α_{crit} . Este valor pode ser interpretado como uma medida análoga à DLO, ou seja, ele estima o ponto em que a diferença entre os dois sons se torna imperceptível para o ouvinte.

Para cada par de sinais, o teste iniciou em $\alpha_0 = -1$ com passo inicial $\Delta\alpha_0 = 2$, de modo que a primeira apresentação tivesse diferença claramente audível e a segunda, tipicamente, coincidissem com $\alpha \approx 1$. Adotaram-se os seguintes limites e critérios de parada: convergência quando $\Delta\alpha \leq 0,05$ ou máximo de 25 iterações.

Figura 29 – Fluxograma das etapas do método PEST para estimativa de limiar.



Fonte: Autor.

3.5.4 Implementação da plataforma digital para o teste e participantes

A plataforma digital para aplicação dos testes foi desenvolvida utilizando a linguagem Python. A interface foi projetada para garantir a facilidade de uso e agilizar o processo de teste (ver Figura 30⁵). A primeira etapa do teste envolveu o cadastro dos participantes, coletando dados como nome, gênero, idade, saúde auditiva e conhecimento acústico, conforme mostrado na Figura 30a.

Um total de 28 participantes participou do teste, e os dados demográficos mostraram que a maioria dos participantes eram homens (60,7%), com faixa etária predominante entre 20 e 29 anos (53,6%). Além disso, a maioria dos participantes se identificou como especialista em acústica (85,7%) e relatou boa saúde auditiva (60,7%). Uma observação interessante foi que 28,6% dos participantes desconheciam sua saúde auditiva, destacando a necessidade de maior conscientização sobre o tema. Por fim, os dados demográficos não revelaram tendências significativas nos resultados dos testes e, portanto, nenhuma segregação de grupo foi aplicada na análise.

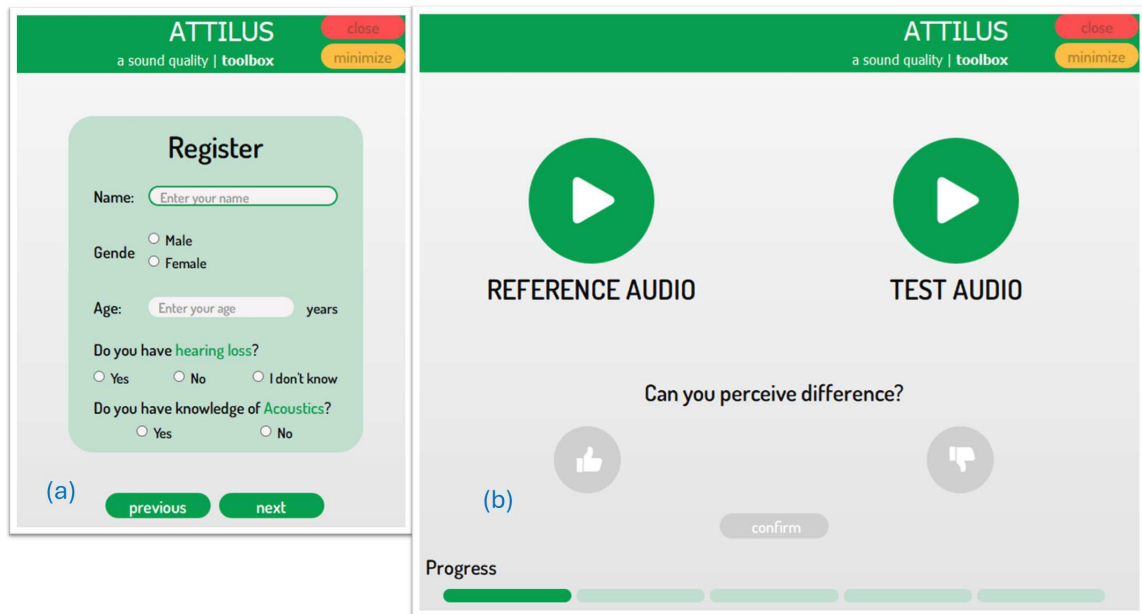
Após a conclusão do registro, o teste prosseguiu para a segunda tela (Figura 30b).

O participante ouviu os áudios de referência e de teste, cada um reproduzido uma vez, e

⁵ A interface dos testes foi mantida em inglês para acomodar a eventual participação de voluntários internacionais e facilitar a reutilização do protocolo em colaborações externas. Antes do início do teste, todos os participantes receberam instruções em português e um glossário dos termos exibidos na tela; verificou-se a compreensão mínima do vocabulário para evitar equívocos de entendimento.

indicou se percebeu alguma diferença. O algoritmo PEST então definiu um novo fator de correção que foi aplicado ao sinal gravado com microfone. O participante realizou outra avaliação auditiva, e esse processo continuou iterativamente até que a convergência fosse alcançada. O tempo médio do teste foi de aproximadamente 15 minutos.

Figura 30 – Interface gráfica utilizada para o teste de juri: (a) tela de registro; (b) tela de teste.



Fonte: Autor.

3.5.5 Ambiente de teste

Os testes de audição foram realizados em uma sala de reuniões do [Laboratório de Vibrações e Acústica \(LVA\)](#) da [Universidade Federal de Santa Catarina \(UFSC\)](#), entre setembro e outubro de 2024, em horários de acordo com a disponibilidade dos voluntários. O local apresentava baixo tráfego de pedestres e um nível de ruído residual favorável à execução dos testes psicométricos ([ISO 8253-2, 2009](#); [ITU R BS1116-3, 2015](#); [Fastl; Zwicker, 2007](#)). O ruído residual global com filtro de ponderação A medido na sala foi de cerca de 33 dB, um nível suficientemente baixo para evitar interferência na percepção auditiva dos participantes.

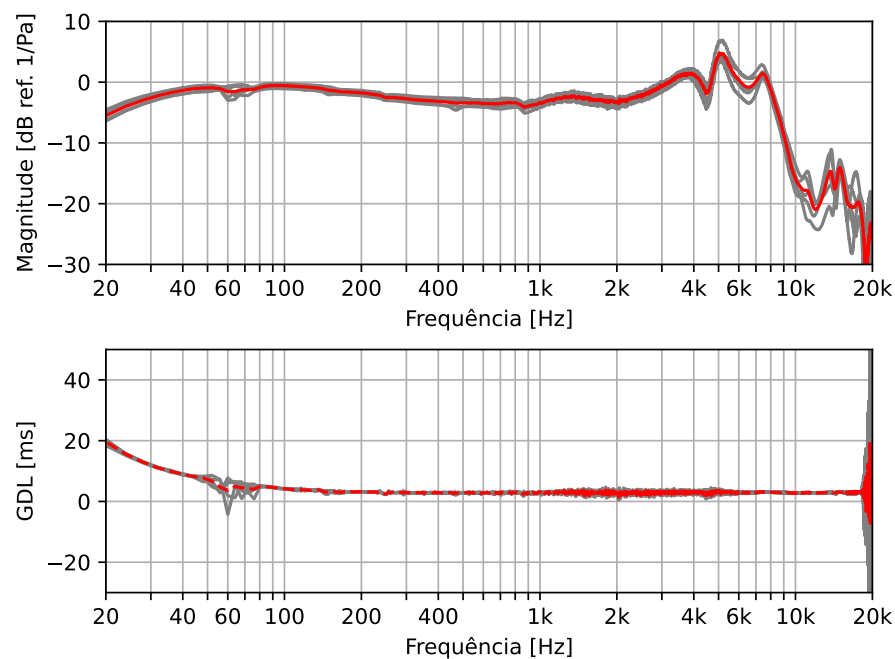
3.5.6 Calibração do sistema de reprodução dos sinais

Para assegurar a reprodução fiel dos sinais nos testes subjetivos, o sistema de reprodução foi calibrado utilizando o HATS HSU III.2 como referência. Esse procedimento,

descrito em detalhes na Seção 2.2.2, permite compensar a influência da cadeia de gravação e reprodução por meio de um processo de equalização, garantindo que o sinal apresentado ao ouvinte corresponda de forma fidedigna ao sinal de referência biauricular.

Os fones de ouvido Sennheiser HD 58X foram posicionados no HATS, e a FRF do sistema foi medida em 10 reposicionamentos, de modo a capturar a variabilidade introduzida pelo acoplamento físico. Os resultados são apresentados na Figura 31, na qual as curvas em cinza representam as medições individuais e a curva vermelha indica a média. Observa-se desempenho consistente entre as repetições, com apenas pequenas variações em faixas específicas de frequência, típicas de ajustes de acoplamento entre fone e orelha artificial.

Figura 31 – Espectro de magnitude e atraso de grupo (GDL) da FRF do sistema de reprodução (canal esquerdo). As curvas em cinza representam as 10 repetições e a curva vermelha representa a média.



Fonte: Autor.

Além da magnitude, foi também analisado o atraso de grupo. Observa-se que, acima de aproximadamente 100 Hz, o atraso permanece praticamente constante em toda a banda, não indicando distorções significativas de fase. Logo, o espectro de fase da FRF foi desconsiderado. Nota-se ainda uma leve variabilidade em torno de 60 Hz, atribuída à interferência do ruído da rede elétrica, que reduz a SNR nessa faixa de frequência.

Com base na FRF média obtida, projetou-se um filtro de correção pela inversão

espectral. Para otimizar o desempenho temporal e minimizar distorções de fase, derivou-se uma versão de fase mínima a partir do espectro de magnitude do filtro. A resposta final foi limitada à faixa de 15 Hz a 18 kHz, evitando ganhos excessivos em regiões fora da banda de interesse e assegurando uma reprodução confiável nos testes subjetivos.

3.5.7 Resultados do teste de júri

A análise inicial realizada nos dados do teste do júri teve como objetivo verificar a normalidade dos fatores de correção críticos (α_{crit}) que convergiram para os 28 participantes. A Tabela 7 apresenta os resultados do teste de Shapiro-Wilk para os fatores de correção correspondentes aos cinco pares de sinais avaliados pelo júri (Tabela 6). Os resultados indicam que, apenas para o par 5, a hipótese nula de normalidade não pode ser rejeitada, considerando um valor-p crítico de 0,05. Portanto, análises não paramétricas foram escolhidas para posterior avaliação dos dados.

Assim, o teste de Kruskal-Wallis foi realizado para avaliar se existiam diferenças estatisticamente significativas entre os fatores de correção (α_{crit}) para os cinco pares de sinais apresentados ao júri. Os resultados da estatística do teste foi de 48,5 e o valor-p de $7,13 \cdot 10^{-10}$, indicando diferenças significativas entre os grupos analisados a um nível de significância de 5%. Com base nesses resultados, foram realizadas análises para identificar quais pares apresentaram essas diferenças.

A Figura 32 apresenta a matriz de valores-p obtida a partir do teste de Dunn, que foi utilizado para comparações pareadas os valores de α_{crit} dos cinco sinais avaliados pelos participantes. A diagonal da matriz representa a auto comparação de cada sinal, em que $p = 1$, enquanto os valores fora da diagonal indicam a significância estatística das diferenças entre os sinais.

Os resultados revelam que os sinais das fontes compressor e refrigerador apresentaram similaridade perceptual significativa, com valores-p acima de 0,05 para comparações entre esses sinais. Da mesma forma, os sinais gerados pelo monitor de áudio também apresentaram comportamento consistente entre si.

No entanto, as comparações entre sinais de fontes eletroacústicas (monitor de áudio) e fontes mecânicas (compressor e refrigerador) demonstraram diferenças estatisticamente significativas, com valores-p abaixo de 5%. Esses resultados sugerem que a percepção das diferenças entre os sinais está intimamente relacionada às características espectrais e à

variabilidade operacional ou de reposicionamento específicas de cada tipo de fonte sonora.

Tabela 7 – Teste de Shapiro-Wilk para a convergência dos valores de α para os dados de todos os sujeitos.

Sinais	Estatística	Valor-p
Par 1	0.882	0.004
Par 2	0.886	0.005
Par 3	0.907	0.017
Par 4	0.889	0.006
Par 5	0.945	0.146

Fonte: Autor.

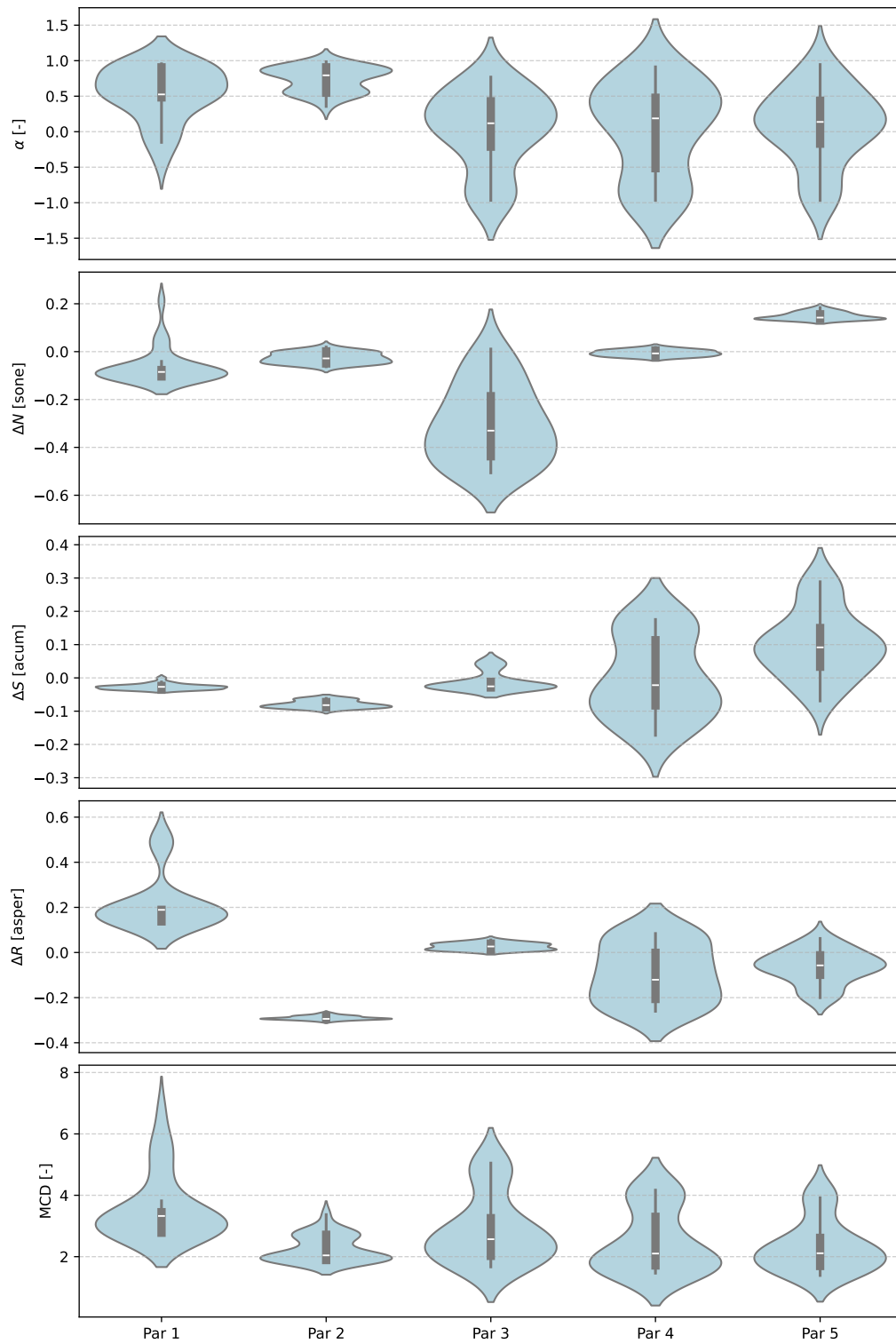
Figura 32 – Matriz de valores- p do teste de Dunn para comparações dos fatores de correção críticos (α_{crit}) entre os cinco pares de sinais avaliados.

Par 1	1.000	0.887	0.002	0.011	0.005
Par 2	0.887	1.000	0.000	0.000	0.000
Par 3	0.002	0.000	1.000	1.000	1.000
Par 4	0.011	0.000	1.000	1.000	1.000
Par 5	0.005	0.000	1.000	1.000	1.000
	Par 1	Par 2	Par 3	Par 4	Par 5

Fonte: Autor.

Após as análises estatísticas, a distribuição dos valores de α_{crit} foi examinada para explorar as tendências nos sinais avaliados. A Figura 33 ilustra, por meio de gráficos de violino, tanto a distribuição de (α_{crit}), quanto das respectivas diferenças dos parâmetros de *loudness* (ΔN), *sharpness* (ΔS), e *roughness* (ΔR), além da *MCD*, calculadas entre o sinal de referência e o sinal de teste corrigido pelos respectivos valores de α_{crit} . A Tabela 8 apresenta as respectivas medianas de cada distribuição para uma análise quantitativa.

Figura 33 – Distribuições do fator de correção crítico (α_{crit}) e das diferenças residuais dos parâmetros (ΔN , ΔS , ΔR e MCD) calculados no limiar de discriminação para os cinco pares de sinais avaliados.



Fonte: Autor.

Tabela 8 – Medianas dos parâmetros mostrados na Figura 33.

	α_{crit}	ΔN	ΔS	ΔR	MCD
Par 1	0.53	-0.08	-0.03	0.19	3.33
Par 2	0.79	-0.03	-0.08	-0.29	2.04
Par 3	0.12	-0.33	-0.03	0.03	2.57
Par 4	0.19	-0.01	-0.02	-0.12	2.10
Par 5	0.14	0.14	0.09	-0.06	2.11

Fonte: Autor.

O painel superior da Figura 33, referente aos valores de α_{crit} , revela que os sinais das fontes mecânicas (Par 1: Refrigerador e Par 2: Compressor) necessitaram de fatores de correção mais altos (medianas de 0,53 e 0,79, respectivamente) para se tornarem indistinguíveis do sinal de referência. Em contraste, os sinais gerados pelo monitor de áudio (Pares 3, 4 e 5) apresentaram medianas de α_{crit} próximas de zero e com maior variabilidade. Essa tendência sugere que as diferenças nos sinais mecânicos foram inerentemente mais evidentes para os participantes, enquanto os sinais obtidos a partir da fonte eletroacústica eram muito mais similares entre si, exigindo pouca ou nenhuma correção para atingir o α_{crit} .

Os painéis inferiores da Figura 33 e os dados da Tabela 8 permitem investigar quais parâmetros objetivos melhor correspondem a essa diferença perceptível no limiar de discriminação:

- **Loudness** (ΔN). A análise das diferenças de *loudness* revela que as medianas para todos os cinco pares variaram de $-0,33$ a $0,14$ sone. O valor absoluto de todas essas medianas está consideravelmente abaixo da DLO de $0,5$ sone estabelecida por You e Jeon (2008). Isso indica que as diferenças de *loudness* entre os sinais de HATS e microfone convoluído com HRIR em sua forma original, ou seja, sem correção espectral, já era baixa o suficiente para que houvesse diferença na sensação de volume sonoro percebida pelos sujeitos;
- **Sharpness** (ΔS). para a *sharpness*, a mediana das diferenças se manteve baixa na maioria dos casos. Contudo, para o Par 5 (compressor a 4500 RPM), a mediana de ΔS foi de $0,09$ acum, valor próximo à DLO de $0,08$ acum. Isso sugere que a *Sharpness* pode ter contribuído para a percepção da diferença em sinais com alto conteúdo de alta frequência, mas não foi um fator consistente em todas as fontes;

- **Roughness (ΔR).** observa-se maior variabilidade e até mudança de sinal entre pares (por exemplo, mediana de +0,19 asper no Par 1 e −0,29 asper no Par 2). Alguns valores excedem o **DLO** reportado na literatura, mas essa comparação deve ser feita com cautela, pois aqui foi adotado o modelo de Daniel–Weber (1997) a partir da implementação disponível na biblioteca MOSQUITO e os estímulos diferem daqueles usados para a determinação do limiar clássico. Assim, ΔR é tratado como indicador exploratório nesse conjunto;
- **MCD.** Diferentemente dos parâmetros psicoacústicos anteriores, a MCD apresenta comportamento mais consistente com a dissimilaridade entre os sinais. As medianas no limiar situam-se em uma faixa compacta, de 2,04 (Par 2) a 3,33 (Par 1), com distribuições estáveis entre pares. Esses resultados sugerem que a **MCD** é um bom candidato a modelo de dissimilaridade para os sinais analisados.

Baseado nesses dados e na carência de estudos sobre a **DLO** para a **MCD**, a análise estatística dos dados obtidos foi refinada. A Figura 34 mostra os valores- p para o teste de Dunn da **MCD**. Observa-se que o sinal de refrigerador é o único que apresenta diferença em relação aos demais, com $p < 0,05$ para todos os casos. Para os sinais de compressor, não é possível rejeitar a hipótese nula, ou seja, não há evidências de diferença estatística para o parâmetro em relação às amostras.

Figura 34 – Matriz de valores p do teste de Dunn para comparações dos valores de MCD referentes α_{crit} de cada sujeito.

Par 1	1.000	0.000	0.041	0.000	0.000
Par 2	0.000	1.000	0.640	1.000	1.000
Par 3	0.041	0.640	1.000	1.000	0.385
Par 4	0.000	1.000	1.000	1.000	1.000
Par 5	0.000	1.000	0.385	1.000	1.000
	Par 1	Par 2	Par 3	Par 4	Par 5

Fonte: Autor.

Assim, os valores para sinais de compressor foram agrupados, foi rodado um *bootstrap*⁶

⁶ O *bootstrap* é um método estatístico de reamostragem que permite estimar a distribuição de um

para definir o valor do [MCD](#) e respectiva zona de confusão referente ao intervalo de confiança de 95%. Obteve-se a mediana de 2,14 com intervalo inferior de 2,04 e superior de 2,41.

parâmetro (neste caso, a mediana da [MCD](#)) e determinar, por inferência, o respectivo intervalo de confiança (zona de confusão) sem assumir uma distribuição teórica específica ([Montgomery; Runger, 2016](#)).

3.6 TESTE MUSHRA PARA AVALIAÇÃO DA DIFERENÇA PERCEPTIVA ENTRE OS SINAIS

A etapa anterior, baseada no método [PEST](#), permitiu estimar os valores críticos do fator de correção (α_{crit}) correspondentes ao limiar de discriminação. No entanto, esse procedimento não quantifica a magnitude da diferença percebida entre os sinais. Para suprir essa lacuna, foi implementada uma adaptação do método [MUSHRA](#), cujo objetivo foi quantificar a diferença entre o sinal de referência e sinais corrigidos com diferentes fatores de correção espectral.

As amostras utilizadas foram as mesmas empregadas nos testes PEST, porém com variação dos fatores de correção, conforme descrito nas seções subsequentes.

3.6.1 Implementação da interface de teste

A interface desenvolvida para o teste é mostrada na Figura 35.⁷ Antes do início, cada participante preenchia uma tela de cadastro semelhante à do teste anterior (Figura 30a), garantindo a coleta dos dados básicos do voluntário.

Foram avaliados dez estímulos sonoros em cada sessão, incluindo uma referência e uma âncora ocultas, seguindo a recomendação da norma ITU-R BS.1534-3 (2014). Diferentemente do procedimento clássico, não foram adicionadas âncoras intermediárias, pois o objetivo não era comparar a qualidade de sistemas, mas sim avaliar diferenças decorrentes da limitação física do aparato de captura. Assim, não havia degradação artificial do sinal, mas apenas discrepâncias originadas da ausência de caracterização perfeita da HRTF.

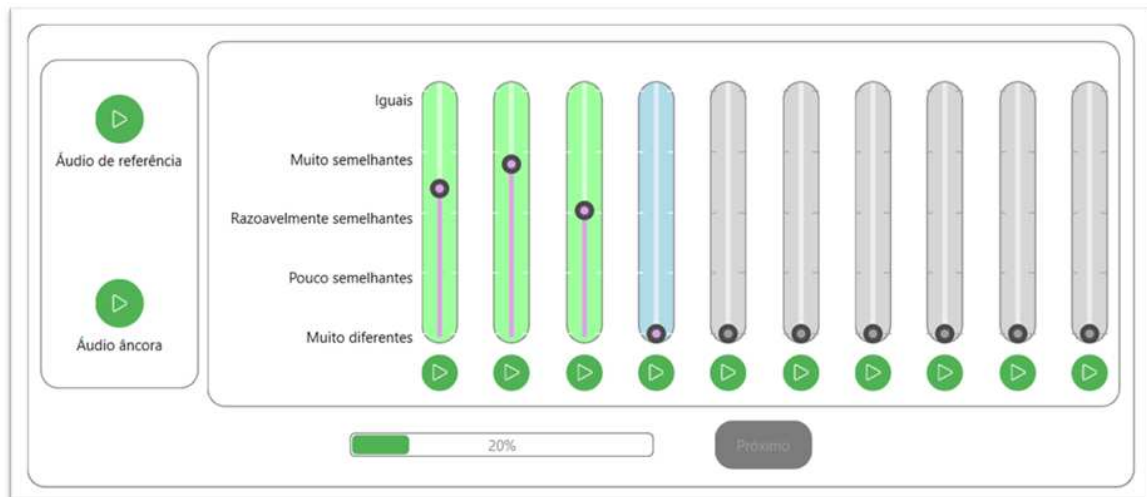
As âncoras verbais foram adaptadas para descritores verbais de cinco níveis: iguais, muito semelhantes, razoavelmente semelhantes, pouco semelhantes e muito diferentes. Antes do teste, os participantes receberam instruções verbais detalhadas, enfatizando que cada estímulo deveria ser comparado diretamente à referência. Explicou-se ainda que a âncora auditiva correspondia à calibração da escala, ou seja, representava a maior diferença possível em relação à referência.

A interface foi configurada de forma a garantir consistência no processo de avaliação: o *slider* correspondente a cada áudio só podia ser movido após sua reprodução, evitando

⁷ Nesse caso, a plataforma foi implementada em português, uma vez que não havia previsão de participação de voluntários estrangeiros.

julgamentos prévios. Os *sliders* já avaliados eram destacados em verde, enquanto o *slider* relativo ao último áudio ouvido permanecia azul, auxiliando o participante na identificação da etapa corrente. Somente após todos os áudios terem sido avaliados era permitido avançar para a próxima etapa, correspondente à análise da fonte sonora seguinte.

Figura 35 – Interface implementada para o teste MUSHRA.



Fonte: Autor.

3.6.2 Amostra sonora para o teste piloto

Para o teste piloto do [MUSHRA](#), os fatores de correção α foram definidos de forma sistemática a partir das diferenças em *loudness*, *sharpness* e *roughness* entre os sinais de HATS e microfone convoluído com HRIR. O algoritmo implementado calculou, para cada valor candidato de α , as discrepâncias em unidades de [DLO](#), de modo a quantificar o afastamento relativo em cada parâmetro psicoacústico.

Com base nesses valores, foram selecionados pontos de α distribuídos ao longo de um contínuo:

1. um $\alpha_{\min}$ no lado negativo, garantindo diferença perceptiva clara em pelo menos um dos parâmetros;
2. um valor de α correspondente ao ponto mais próximo do sinal de referência (mínima diferença);
3. um conjunto de valores intermediários, obtidos por percentis da distância composta em JNDs, assegurando diversidade de condições de teste. Além disso, ao menos um

α positivo foi incluído, representando a aproximação do sinal de teste ao sinal de referência.

No total, nove valores de α foram selecionados, cobrindo desde diferenças amplificadas até aproximações muito próximas à referência. A partir desses fatores, foram gerados os estímulos para o teste piloto, obtidos pela aplicação do filtro de correção espectral ponderada aos pares HATS–microfone convoluído com HRIR. Dessa forma, assegurou-se que os estímulos avaliados pelo júri cobrissem todo o espectro de variação perceptual relevante, desde diferenças claramente audíveis até condições em que os sinais eram muito semelhantes. Na análise dos resultados, os dados de sujeitos que atribuíram escores menores que 90% para a referência oculta foram eliminados. Porém, não foram estabelecidos critérios de exclusão para a âncora oculta, pois os participantes foram instruídos a comparar os sinais sempre com a referência.

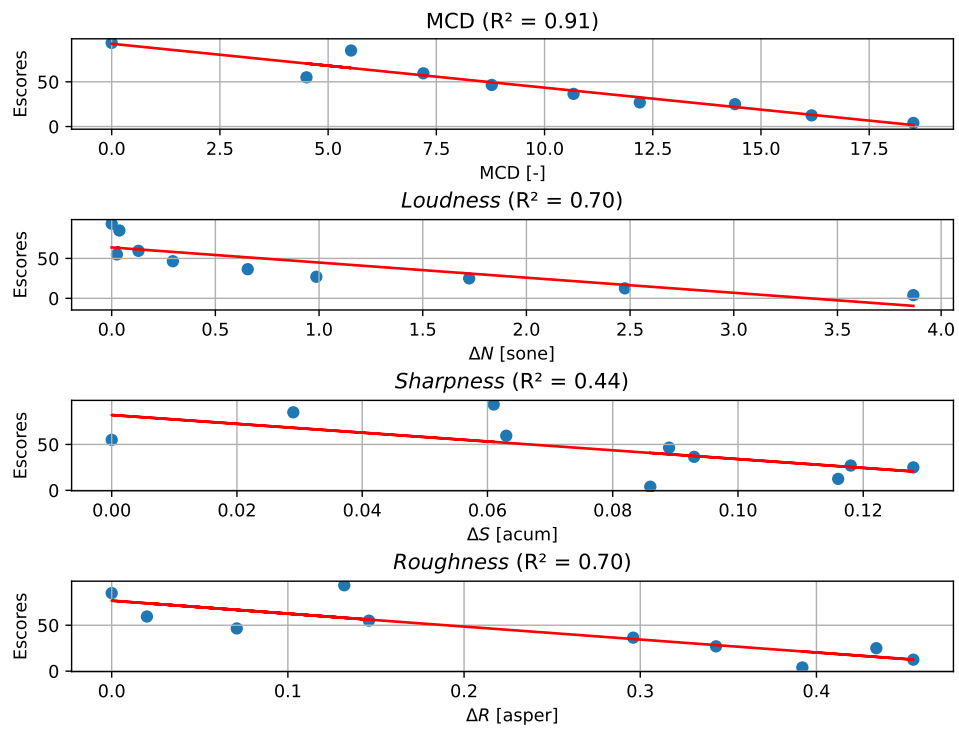
3.6.3 Resultados do teste piloto

As Figuras 36 a 40 apresentam a correlação entre as medianas dos escores atribuídos pelos participantes e os parâmetros psicoacústicos calculados. De forma geral, observou-se que a MCD apresentou a correlação mais consistente e elevada com os escores subjetivos em todas as fontes analisadas, superando os demais parâmetros (*loudness*, *sharpness* e *roughness*). A melhor correlação para os outros parâmetros é de 0,96 para o *sharpness* nos sinais de compressor a 2400 rpm, porém, o mesmo parâmetro apresenta correlação de 0,84 para compressor de velocidade única sendo proporcional ao escore, o que é inconsistente com a diferença perceptiva. Esses resultados indicam que a MCD é o parâmetro mais adequado para estimar a percepção de diferença entre sinais de HATS e microfone convoluído com HRIR com os diferentes níveis de correção espectral utilizados.

Assim, os resultados do teste piloto forneceram a base para a implementação definitiva do método MUSHRA, descrita na subseção anterior. A etapa final foi conduzida com foco na MCD, permitindo uma análise mais robusta e alinhada à percepção dos participantes. Os valores de referência de MCD foram definidos de forma a abranger toda a faixa perceptual relevante, entre 0 (referência oculta) e aproximadamente 20 (âncora, com maior diferença espectral observada). Embora os pontos intermediários tenham sido gerados a partir de uma distribuição quase linear, pequenas variações no espaçamento resultaram da discretização numérica utilizada. A escala horizontal, portanto, representa os níveis efetivos

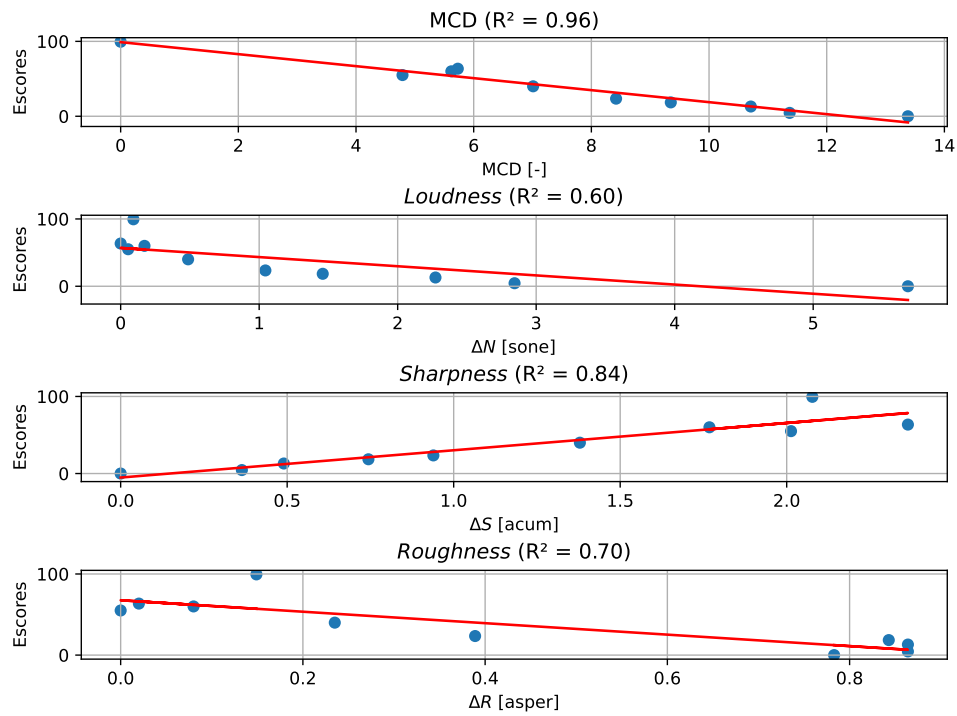
de discrepância entre os sinais de teste e de referência, enquanto os escores MUSHRA, exibidos até 125, incluem a extrapolação gráfica para melhor visualização das distribuições dentro dos violinos.

Figura 36 – Correlação das medianas dos escores MUSHRA com os parâmetros utilizados (sinal de refrigerador).



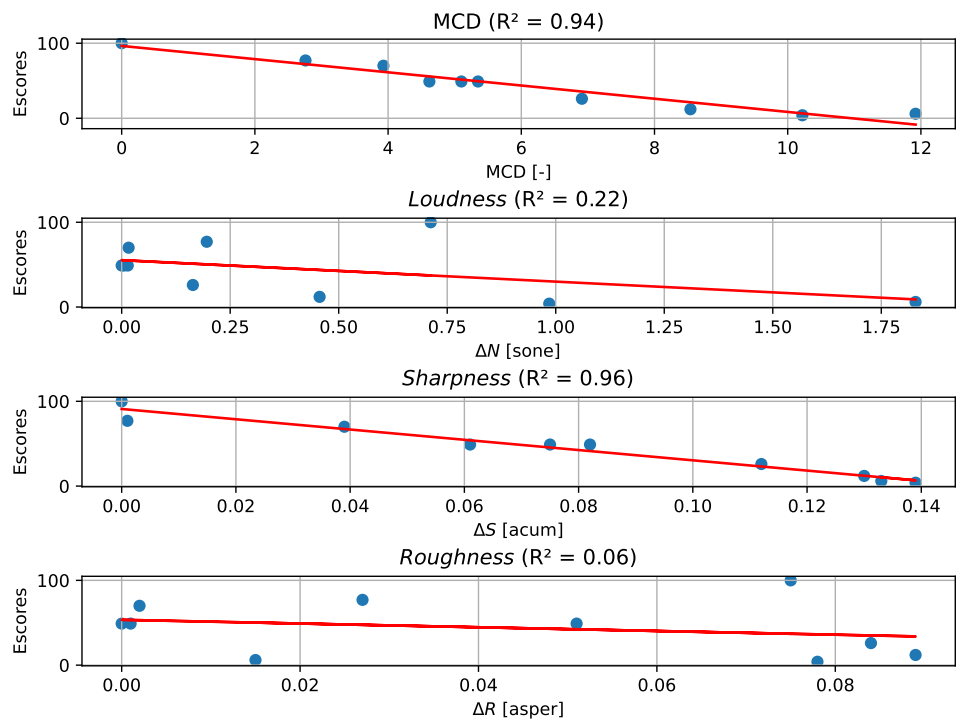
Fonte: Autor.

Figura 37 – Correlação das medianas dos escores MUSHRA com os parâmetros utilizadas (sinal de compressor de velocidade única).



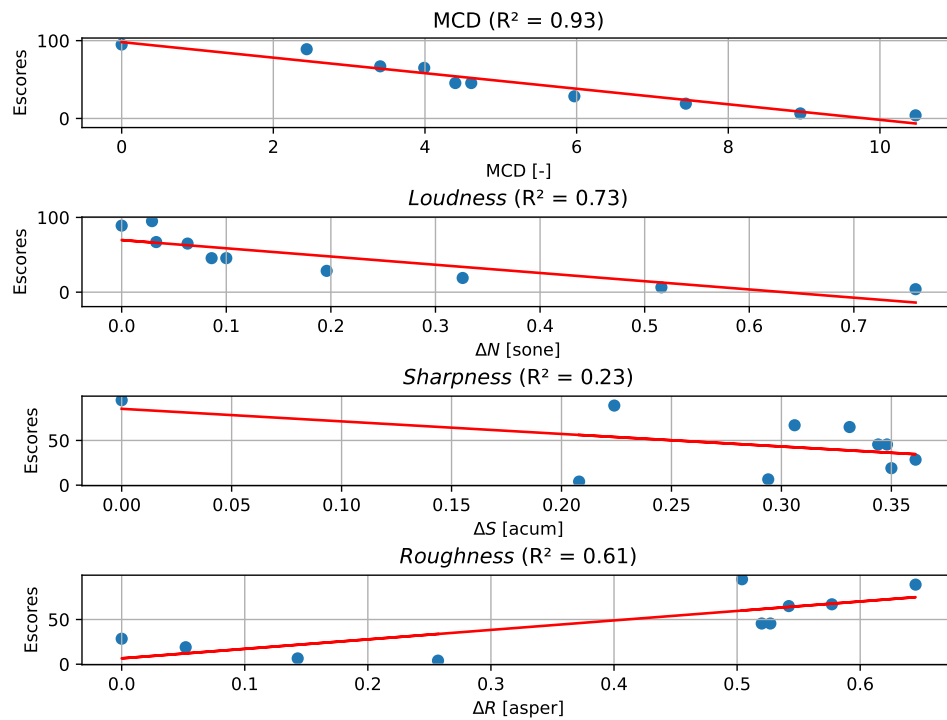
Fonte: Autor.

Figura 38 – Correlação das medianas dos escores MUSHRA com os parâmetros utilizados (sinal de compressor de velocidade a 2400 rpm).



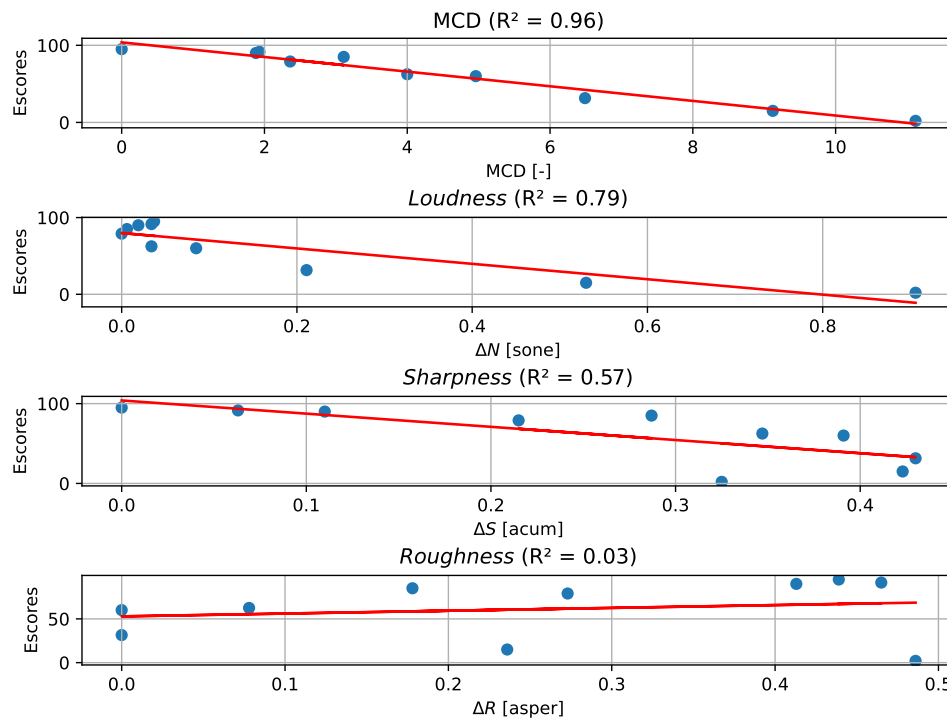
Fonte: Autor.

Figura 39 – Correlação das medianas dos escores MUSHRA com os parâmetros utilizados (sinal de compressor de velocidade a 3000 rpm).



Fonte: Autor.

Figura 40 – Correlação das medianas dos escores MUSHRA com os parâmetros utilizados (sinal de compressor de velocidade a 4500 rpm).



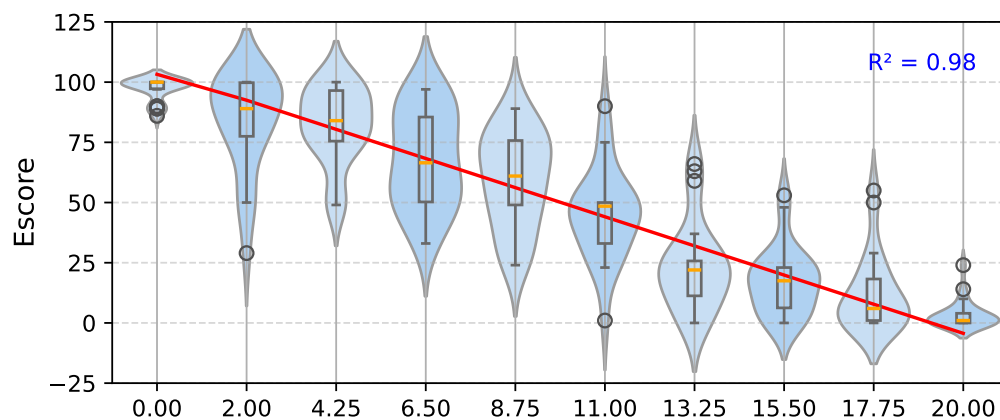
Fonte: Autor.

3.6.4 Resultados do teste final

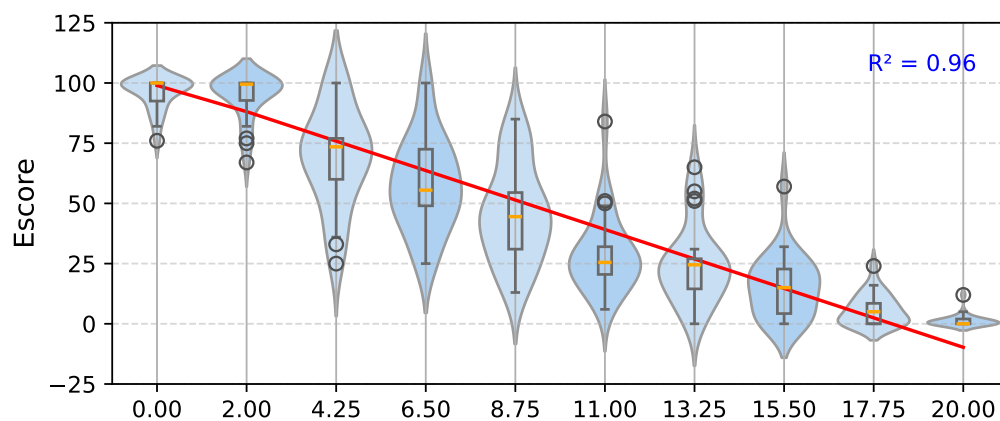
Figuras 41 e 42 mostram a dispersão dos escores MUSHRA com os respectivos valores de R^2 obtidos em relação à MCD para as fontes mecânicas e eletroacústica, respectivamente. Observa-se alta relação linear, sendo o pior caso de $R^2 = 0,96$ para o compressor de velocidade única, evidenciando que o MCD se mostrou como um modelo consistente para representar as diferenças percebidas pelo júri.

Com relação à variabilidade dos escores, para todos os casos foi observada uma maior dispersão nas regiões intermediárias da MCD. Para os sinais de referência e âncora a variabilidade foi baixa em todos os casos, exceto para o compressor a 4500 rpm (Figura 42c). É importante ressaltar que os dados de sujeitos que atribuíram menos de 90% para a âncora oculta foram eliminados das análises estatísticas.

Figura 41 – Distribuição dos escores MUSHRA e correlação com os valores de MCD para as fontes mecânicas.



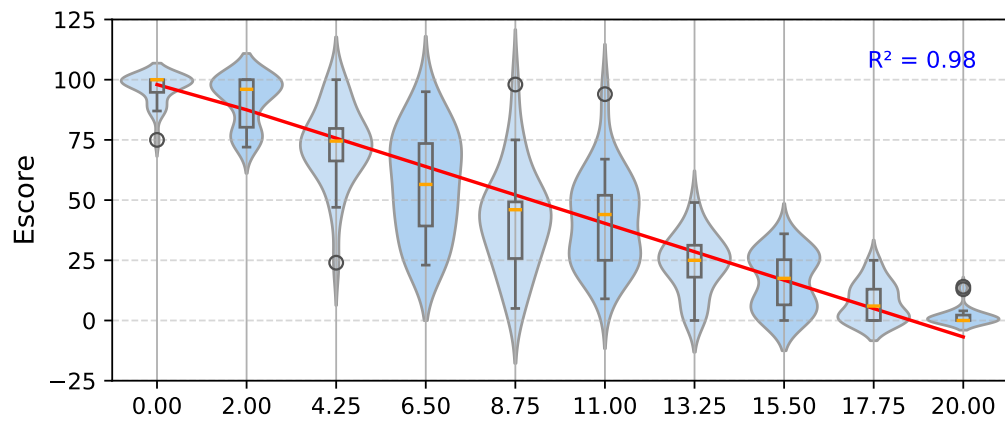
(a) Sinal de refrigerador.



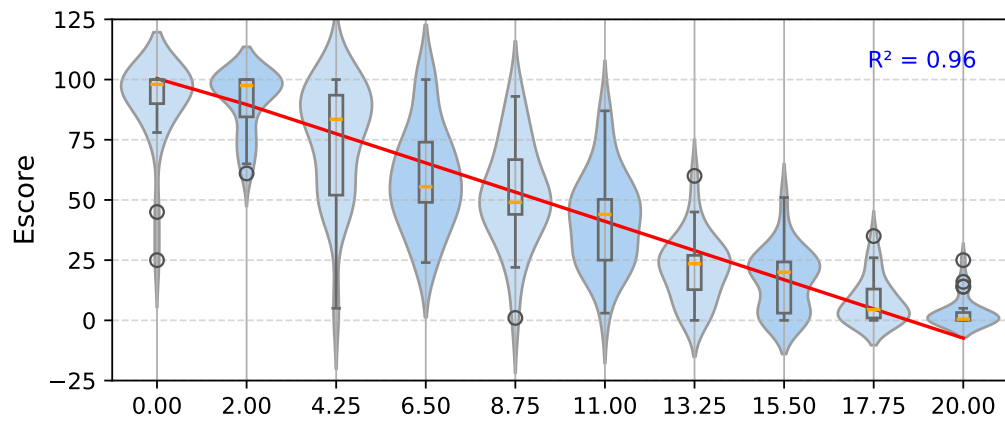
(b) Sinal de compressor velocidade única.

Fonte: Autor.

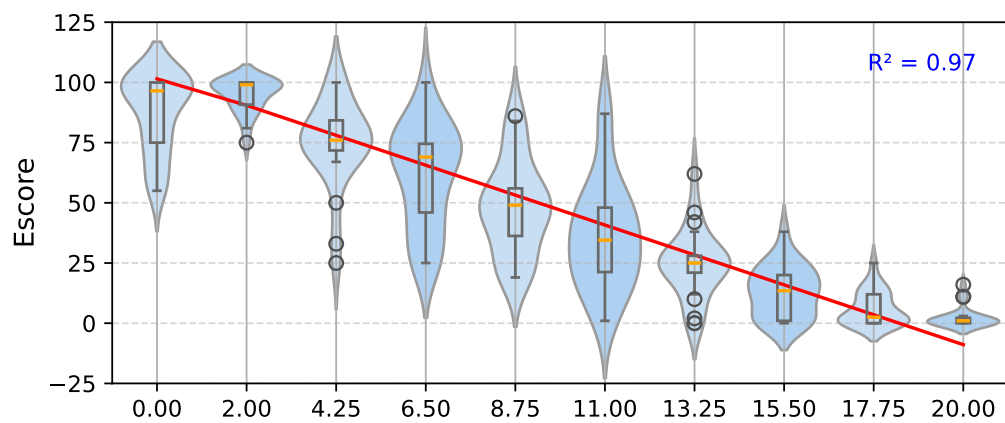
Figura 42 – Distribuição dos escores MUSHRA e correlação com os valores de MCD para os sinais de fonte eletroacústica.



(a) Sinal de compressor a 2400 rpm.



(b) Sinal de compressor a 3000 rpm.



(c) Sinal de compressor a 4500 rpm.

Fonte: Autor.

3.7 REVISÃO E SÍNTESE DO CAPÍTULO

Neste capítulo, foi descrito um estudo comparativo entre sinais obtidos com [HATS](#) e microfone convoluído com HRIR em câmara semi-anecoica. O objetivo foi avaliar em que medida um microfone pode substituir ou complementar o uso do manequim na aquisição de sinais biauriculares para a avaliação da qualidade sonora de sistemas de refrigeração.

Primeiramente, realizou-se a medição da [HRTF](#) para a posterior convolução da respectiva [HRIR](#) aos sinais de microfone, permitindo uma comparação direta com os sinais obtidos com o [HATS](#). O procedimento experimental envolveu três fontes sonoras: um compressor hermético, um refrigerador e um monitor de áudio. Essa abordagem permitiu analisar tanto fontes mecânicas, cuja variabilidade é dominada pela instabilidade operacional, quanto uma fonte eletroacústica, utilizada para isolar a variabilidade ocasionada pelo reposicionamento dos receptores.

Os resultados objetivos da análise espectral mostraram boa concordância entre os sinais de [HATS](#) e microfone em frequências de até 1 kHz. Contudo, discrepâncias significativas de até 10 dB foram observadas acima de 1 kHz, atribuídas a limitações da [HRTF](#), que não captura interações complexas com o piso da câmara e a consequente propagação de ondas não planas. Adicionalmente, o compressor apresentou maior variabilidade espectral, devido à sua menor potência sonora (resultando em menor [SNR](#)) e à instabilidade inerente à montagem experimental.

A análise de variabilidade dos parâmetros psicoacústicos revelou resultados distintos para cada tipo de fonte. Para o monitor de áudio, no qual se avaliou o efeito do reposicionamento, os parâmetros de *loudness* e *roughness* apresentaram diferenças sistematicamente superiores aos limiares de percepção (DLOs). Em contrapartida, as diferenças de *sharpness* ficaram abaixo da [DLO](#) estabelecida na literatura. Para as fontes mecânicas (compressor e refrigerador), embora tenham sido identificadas diferenças estatisticamente significativas, os desvios nas medianas de *loudness*, *sharpness* e *roughness* permaneceram abaixo das respectivas [DLOs](#), indicando que, para essas fontes, os dois métodos de captura forneceram resultados equivalentes. Já para a , a comparação entre sinais de HATS e microfone convoluídos com HRIR sempre extrapolaram a DLO obtida, sugerindo que esse parâmetro é sensível a pequenas diferenças entre os sinais.

A avaliação subjetiva foi estruturada em duas etapas complementares. Primeiramente, o método PEST foi utilizado para estimar o fator de correção espectral crítico α_{crit} no

qual os participantes começavam a discriminar os sinais. Os resultados mostraram que para as fontes mecânicas as diferenças eram mais evidentes, enquanto para o monitor de áudio os sinais eram tão similares que o erro espectral precisou ser extrapolado (valores de α negativos) para se tornar audível. Em seguida, um teste [MUSHRA](#) foi implementado para quantificar a magnitude da diferença percebida.

Os resultados do teste [MUSHRA](#) indicaram que a [MCD](#) apresentou a correlação mais consistente e elevada com os escores subjetivos, superando *loudness*, *sharpness* e *roughness*. No teste final, observou-se uma forte relação linear entre a MCD e os escores do júri para todas as fontes sonoras ($R^2 \geq 0,96$), validando a [MCD](#) como um indicador robusto da diferença percebida entre os sinais.

4 COMPARAÇÃO DE SINAIS OBTIDOS POR DIFERENTES HATS E CARACTERIZAÇÃO DE FONES DE OUVIDO

O capítulo anterior apresentou a comparação entre gravações realizadas com [HATS](#) e com microfone convoluído com [HRIR](#) em ambiente controlado, permitindo avaliar a correspondência espectral entre diferentes técnicas de aquisição. A partir desses resultados, a [MCD](#) foi adotada como parâmetro objetivo para quantificação das diferenças entre sinais.

Este capítulo descreverá um estudo complementar no qual se realizou uma comparação entre três [HATS](#) em câmara semi-anecoica e em uma sala de reuniões do [LVA](#), além de critérios de fones de ouvido e em relação à robustez da equalização. O objetivo principal é avaliar a dispersão espectral de sinais bem como da [MCD](#) entre sinais ao utilizar 3 [HATSs](#) distintos:

- HSU III.2, mesmo do estudo anterior, sem simulador de pina e com canal bloqueado;
- Kemar: com simulador de pina, canal auditivo e impedância de orelha média;
- Cortex MK 2 restaurado: para esse manequim foram confeccionadas novas orelhas a partir de um modelo 3D das originais (simulador de pina), mas foi adaptado para utilizar microfone de baixo custo no canal bloqueado.

4.1 MEDIÇÃO DOS PARÂMETROS DE FONES DE OUVIDO

Com o objetivo de investigar as características de diferentes modelos de fones de ouvido e avaliar seu impacto na reprodução de sinais para testes subjetivos, foi realizada uma caracterização detalhada de quatro equipamentos. Os fones de ouvido avaliados foram:

- Sennheiser HD 600 (aberto, circumauricular);
- Sennheiser HD 58 (aberto, circumauricular);
- Sennheiser HD 206 (fechado, circumauricular);
- AKG K92 (fechado, circumauricular).

As medições foram conduzidas utilizando o [HATS](#) Kemar que foi adaptado com uma orelha prototipada contendo um microfone MEMS na entrada do canal auditivo, conforme ilustrado na [Figura 43](#). Essa montagem permitiu a captura simultânea da resposta em

frequência em três pontos: na entrada do canal auditivo aberto, na entrada do canal auditivo bloqueado e na posição da membrana timpânica.

Figura 43 – Orelha prototipada com o microfone MEMS na entrada do canal auditivo.



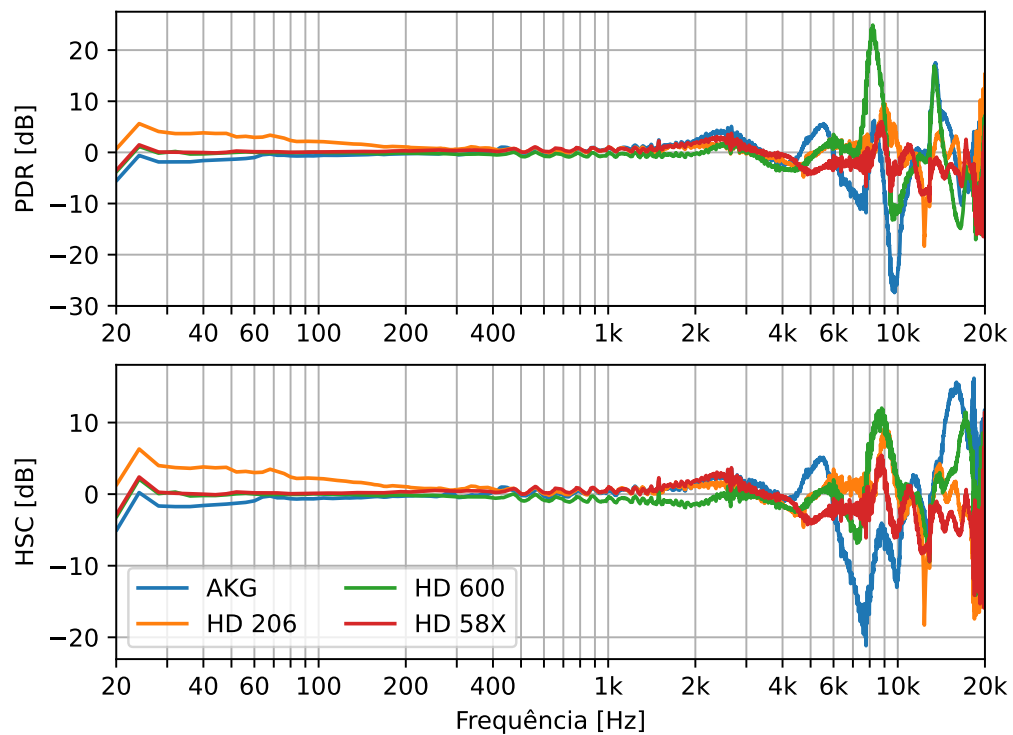
Fonte: Autor.

4.1.1 Critérios de Abertura (PDR e HSC)

Inicialmente, foram avaliados os critérios de abertura dos fones para verificar o quanto cada modelo se aproxima de uma condição de campo livre, ideal para a reprodução biauricular (ver Seção 2.2.2). A Figura 44 apresenta os parâmetros de PDR e HSC dos fones de ouvido.

Os modelos abertos (HD 58X e HD 600) apresentam valores de PDR e HSC próximos de 0 dB entre 200 Hz e 1 kHz, indicando um acoplamento acusticamente transparente e interferência menor na região da primeira ressonância do canal auditivo. Em contraste, os fones fechados (HD 206 e AKG K92) exibem valores elevados em baixas frequências, o que é coerente com uma vedação mais eficaz que aumenta a carga acústica no canal (efeito de oclusão). A proximidade entre PDR e HSC nos fones abertos reforça que a transferência sonora para o tímpano é similar à do campo livre, validando-os como mais adequados para a reprodução fidedigna de sinais biauriculares.

Figura 44 – Critérios de abertura dos fones de ouvido: PDR na figura superior e HSC na figura inferior.



Fonte: Autor.

4.1.1.1 FRF dos fones de ouvido

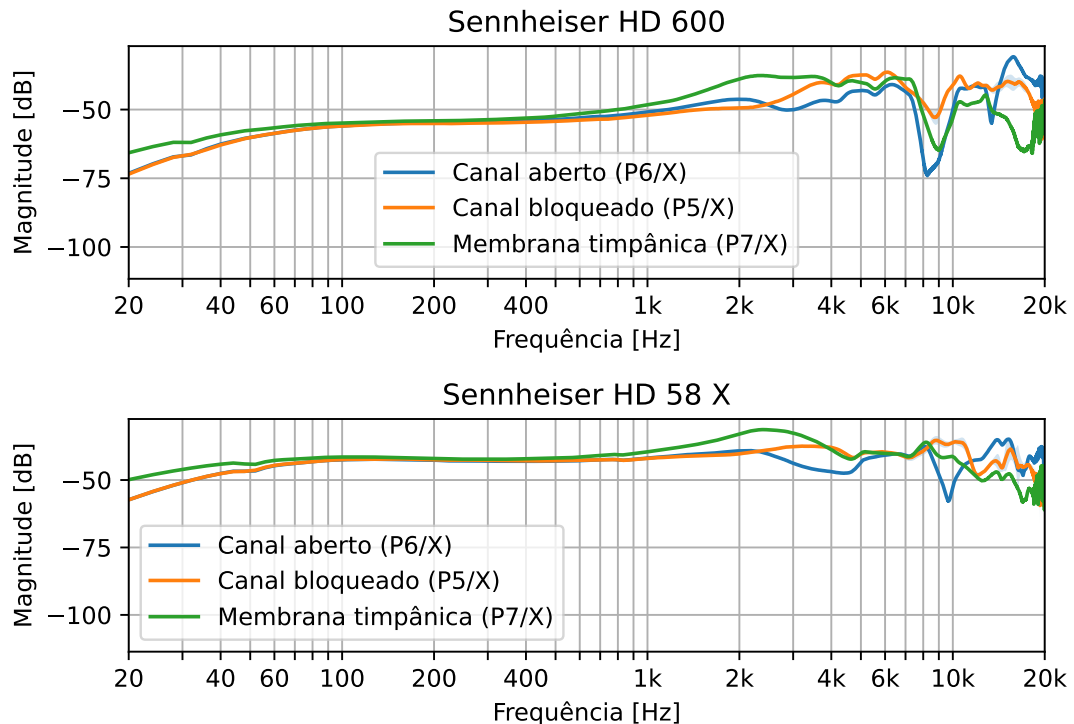
A análise da magnitude da resposta em frequência foi agrupada por tipo de fone para facilitar a comparação. As Figuras 45 e 46 apresentam as FRFs médias para as três posições de medição, com a variabilidade (desvio padrão) obtida a partir de 10 reposicionamentos.

Como mostra a Figura 45, os fones abertos (HD 600 e HD 58 X) exibem um comportamento consistente e uma resposta relativamente plana, especialmente na região de médias frequências. A Figura 46, por outro lado, revela que os fones fechados (AKG K92 e HD 206) possuem FRFs com picos e vales mais pronunciados, característicos de sistemas com maior ressonância interna devido à vedação.

Adicionalmente, uma análise das curvas revela comportamentos comuns a todos os modelos. A FRF obtida na membrana timpânica (P7/X - curvas verdes) exibe um ganho proeminente na faixa entre 2 e 5 kHz, o que é consistente com a ressonância primária do canal auditivo. Também se observa um vale acentuado (anti-ressonância) entre 8 e 10 kHz, que ocorre exclusivamente nas medições com o canal auditivo aberto (P6/X - curvas azuis). Esse fenômeno é característico de fones circumauriculares e resulta de um cancelamento

acústico na cavidade formada entre o fone e a orelha do manequim.

Figura 45 – Espectros de magnitude médios da FRF dos fones de ouvido abertos, considerando 10 reposicionamentos.



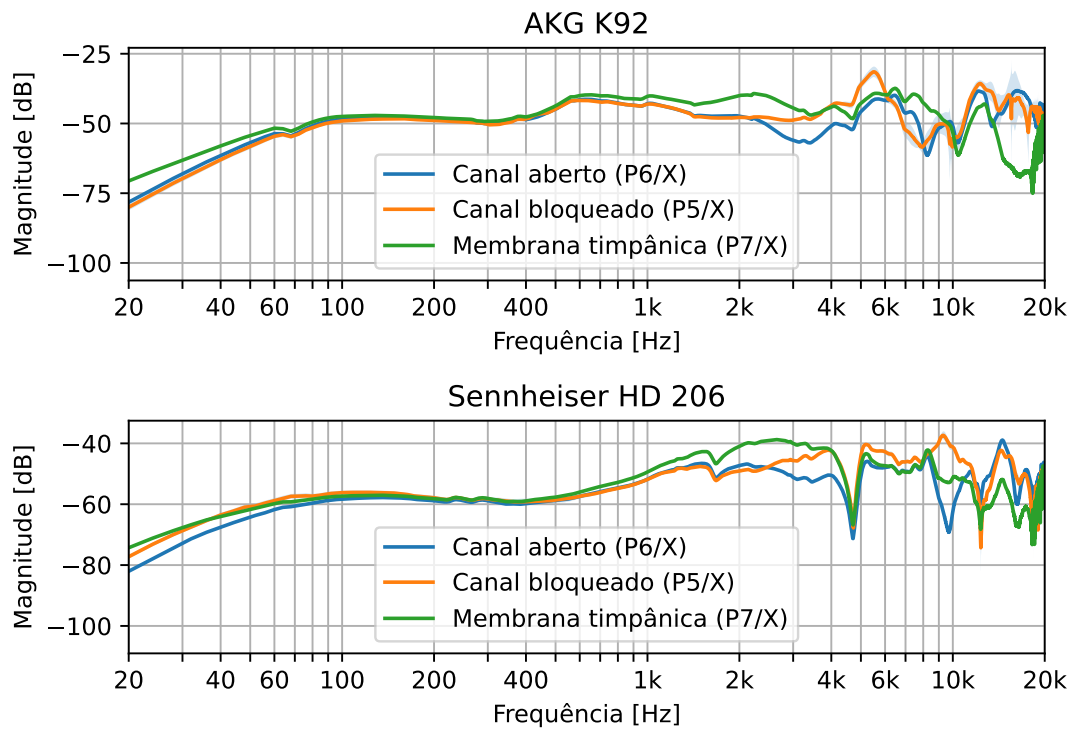
Fonte: Autor.

4.1.2 Espectros de atraso de grupo

Para avaliar a resposta de fase dos fones, o atraso de grupo foi analisado. A Figura 47 compara o comportamento de um fone aberto (HD 58 X) e um fechado (HD 206).

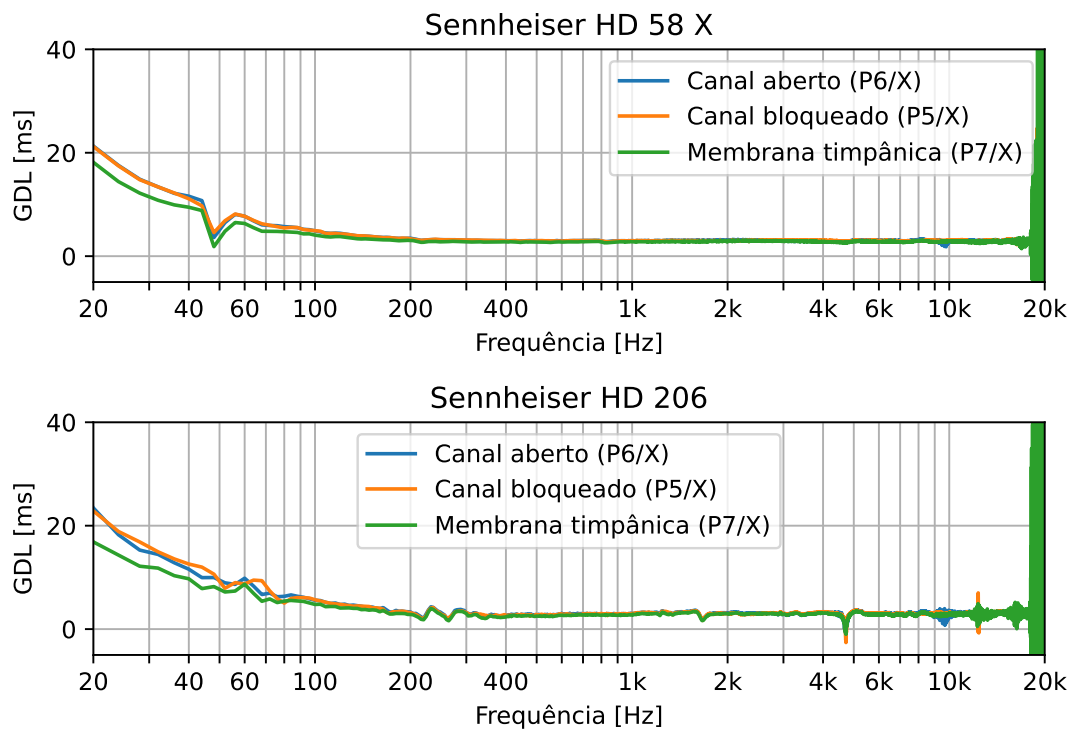
A análise da média revela que as três condições de medição (canal aberto, bloqueado e tímpano) apresentam resultados quase idênticos acima de 100 Hz para ambos os fones. O HD 58 X exibe um [GDL](#) praticamente constante em toda a faixa de médias e altas frequências, indicando uma excelente resposta a transientes. O HD 206, por sua vez, apresenta maior variação, especialmente em baixas frequências. Diante desse comportamento observado na resposta de fase do fone aberto, optou-se por equalizar apenas o espectro de magnitude da FRF dos fones.

Figura 46 – Espectros de magnitude médios da FRF dos fones de ouvido fechados, considerando 10 reposicionamentos.



Fonte: Autor.

Figura 47 – Atraso de Grupo médio para um fone aberto (HD 58 X) e um fechado (HD 206).

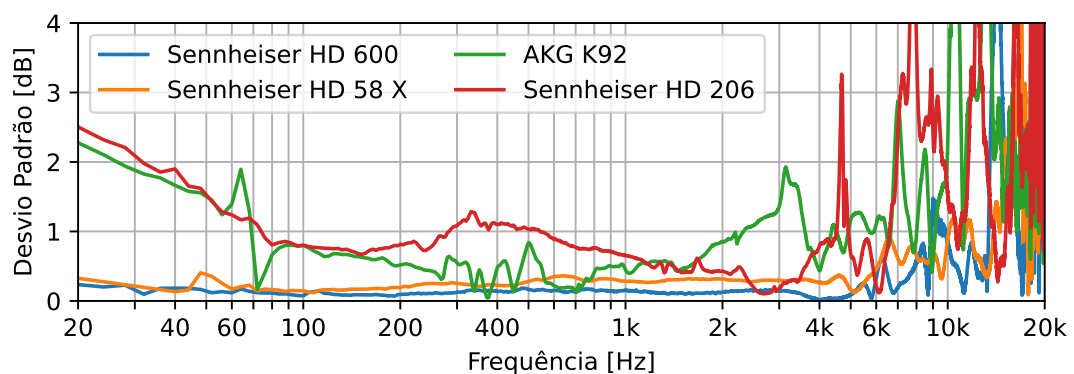


Fonte: Autor.

4.1.3 Variabilidade da Magnitude da FRF

A variabilidade das medições decorrente do reposicionamento dos fones de ouvido foi avaliada a partir do desvio padrão dos espectros de magnitude da FRF. A Figura 48 apresenta um comparativo entre os quatro modelos de fone, considerando os dados obtidos na posição da membrana timpânica (P7), em que os efeitos de acoplamento são mais evidentes.

Figura 48 – Desvio padrão dos espectros de magnitude da FRF dos 4 fones para os 10 reposicionamentos e gravação na membrana timpânica.



Fonte: Autor.

Observa-se um comportamento distinto entre os fones de tipo aberto e fechado. Os modelos abertos, Sennheiser HD 600 e HD 58 X, apresentam uma variabilidade notavelmente baixa e similar entre si, com o desvio padrão não excedendo 0,5 dB até aproximadamente 6 kHz. Esse resultado indica um acoplamento acústico robusto e com boa repetibilidade.

Em contraste, os fones fechados exibem maior variabilidade. O Sennheiser HD 206 apresenta a maior desvio-padrão em baixas frequências, o que pode ser atribuído à sua arquitetura, que dificulta uma vedação perfeita e consistente, permitindo que pequenas aberturas entre a almofada e a cabeça alterem significativamente a resposta de graves. O AKG K92 também demonstra uma variabilidade consideravelmente maior que a dos modelos abertos. De forma geral, para altas frequências, os picos agudos curvas de desvio padrão coincidem com as regiões de anti-ressonância (vales profundos) nos espectros de magnitude das FRFs (ver Figuras 45 e 46), em que pequenas alterações no percurso acústico resultam em grandes variações de nível. Por fim, nota-se que o desvio padrão aumenta consideravelmente para todos os fones em altas frequências, especialmente após

a primeira ressonância do canal auditivo (acima de ≈ 6 kHz), em que os comprimentos de onda curtos são mais sensíveis a pequenas alterações de posicionamento.

4.1.4 Síntese e Justificativa da Escolha do Fone

Com base nas análises objetivas, o fone de ouvido Sennheiser HD 58 X foi selecionado para a realização dos testes subsequentes. A escolha se justifica por seu desempenho superior e mais adequado à reprodução binauricular de alta fidelidade, conforme os seguintes pontos:

1. Sua arquitetura aberta otimiza os critérios de PDR e HSC, garantindo baixa interferência na impedância de entrada do canal auditivo;
2. Sua resposta em frequência é consistente e apresenta baixa variabilidade entre múltiplos reposicionamentos, garantindo repetibilidade;
3. Seu espectro de fase é linear para médias e altas frequências, o que é evidenciado pelo atraso de grupo constante nessa faixa, o que assegura uma reprodução livre de distorções de fase significativas;
4. Apresenta um desempenho eletroacústico muito similar ao do modelo de referência Sennheiser HD 600, porém com um custo significativamente inferior — aproximadamente R\$ 1100,00, menos da metade do valor do HD 600 ($\approx$ R\$ 2600,00) na data da aquisição — representando uma excelente relação custo-benefício.

Dessa forma, a combinação de características técnicas robustas e um custo-benefício favorável consolidam o HD 58 X como a escolha ideal para os procedimentos experimentais deste trabalho.

4.1.5 Comparativo da FRF do fone utilizando os diferentes HATSs

Para investigar a influência do manequim na medição, a caracterização do fone Sennheiser HD 58 X foi estendida aos três HATS utilizados nos ensaios. A Figura 49 apresenta um comparativo da FRF média (gráfico superior) e do respectivo desvio-padrão (gráfico inferior) para as medições realizadas na membrana timpânica.

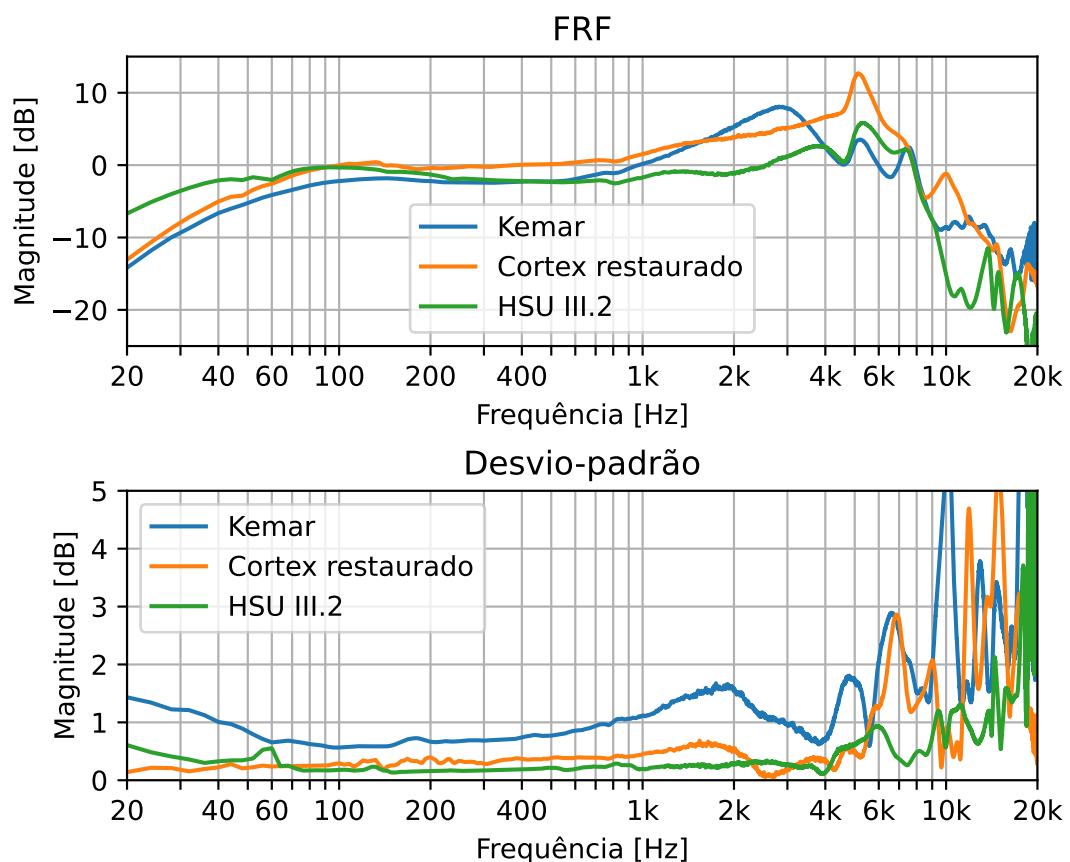
A análise das FRFs médias revela que, embora a tendência geral das curvas seja similar, existem diferenças notáveis na magnitude, especialmente na região de ressonância

do canal auditivo (entre 2 e 6 kHz). Isso evidencia que a geometria específica de cada manequim impacta diretamente a resposta em frequência medida.

A análise da variabilidade, apresentada no gráfico de desvio-padrão, mostra comportamentos distintos para cada HATS. O manequim Kemar exibe a maior variabilidade na faixa de 1 a 4 kHz, o que pode ser atribuído à complexidade de seu simulador de canal auditivo, que forma uma cavidade maior e, conseqüentemente, mais sensível a pequenas variações no acoplamento acústico com o fone. Em contrapartida, o HSU III.2 demonstra a menor variabilidade na maior parte da faixa de frequência, o que pode ser relacionado à sua geometria mais simplificada.

Um comportamento particular é observado no Cortex restaurado, que apresenta a maior variabilidade em altas frequências. Conforme a Figura 49, um pico de variabilidade proeminente em torno de 5 kHz na curva de desvio-padrão corresponde diretamente a uma ressonância com alto fator de qualidade na FRF média deste manequim.

Figura 49 – Espectro de magnitude da FRF e desvio-padrão do fone HD 58 X com os 3 HATS utilizados nos ensaios.



Fonte: Autor.

Regiões com ressonâncias de alto fator de qualidade são intrinsecamente mais sensíveis a pequenas alterações no acoplamento, o que explica por que a magnitude medida flutua significativamente a cada reposicionamento do fone. A origem dessa ressonância acentuada pode ser uma consequência da geometria específica da pina restaurada ou, mais provavelmente, da própria resposta e montagem do microfone de baixo custo utilizado neste manequim, que pode ter introduzido uma característica acústica não presente nos outros modelos. Além disso, o modelo de pina empregado no manequim Cortex difere daquele utilizado no Kemar, o que também pode contribuir para as diferenças observadas nas curvas de FRF obtidas.

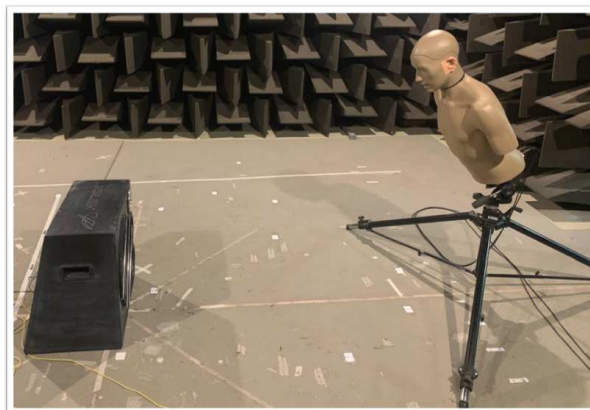
Portanto, os resultados apresentados nesta seção demonstram que a repetibilidade das medições não depende apenas do fone, mas também da complexa interação deste com a morfologia de cada [HATS](#). Tal constatação reforça a importância de utilizar um conjunto casado de [HATS](#) e fone de ouvido.

4.2 MEDIÇÕES EM CÂMARA SEMI-ANECOICA COM OS 3 HATS

Para expandir a análise comparativa entre microfone e [HATS](#) apresentada no Capítulo 3, esta seção apresenta os resultados obtidos em condições similares utilizando 3 [HATS](#).

O procedimento experimental consistiu em posicionar uma fonte sonora no centro da câmara e realizar medições sequenciais com cada um dos três manequins, mantendo sempre a mesma posição em relação à fonte. A Figura 50 ilustra a montagem experimental, mostrando o [HATS](#) Kemar posicionado em frente à caixa acústica utilizada nos ensaios.

Figura 50 – HATS Kemar posicionado no local de medição.



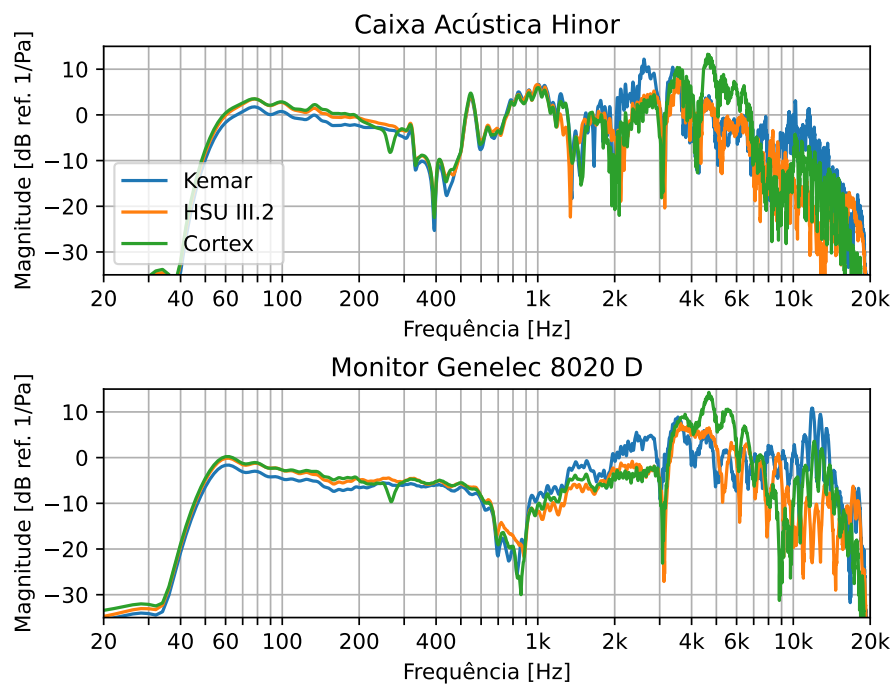
Fonte: Autor.

4.2.1 FRFs entre fontes sonoras e HATS obtidas no experimento

As Figuras 51 e 52 mostram as FRFs entre a fonte sonora (sinal de excitação gerado no computador) e cada HATS (pressão sonora no microfone), sem e com a equalização do fone de ouvido HD 58 X, respectivamente.

Analisando as curvas sem equalização (Figura 51), observa-se uma boa concordância entre os três manequins em baixas frequências, até aproximadamente 1 kHz, para ambas as fontes sonoras. Acima dessa região, surgem divergências significativas que refletem as características morfológicas de cada HATS.

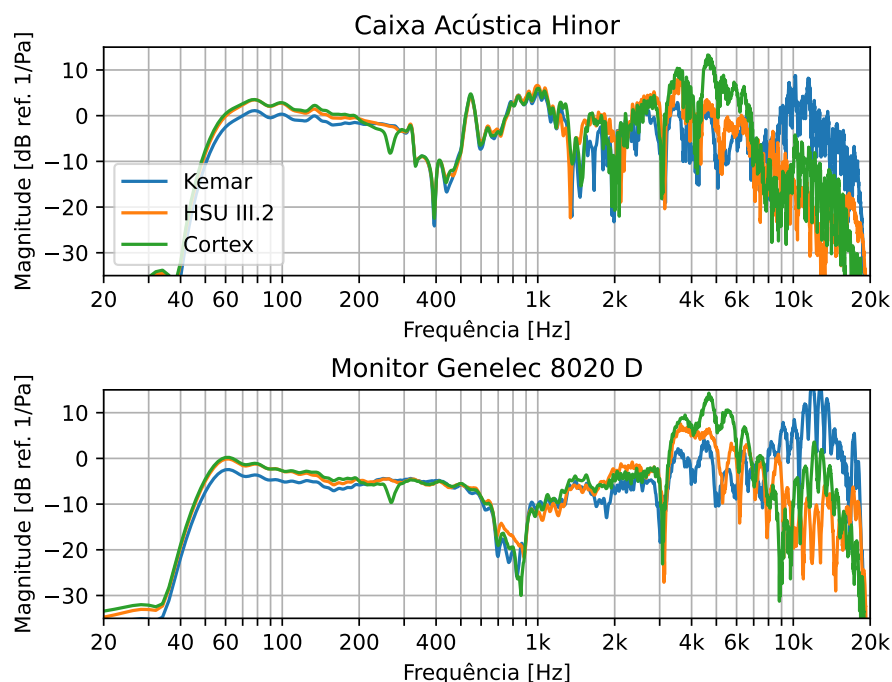
Figura 51 – Espectro de magnitude das FRFs entre fonte e receptor utilizando os 3 HATS e as duas fontes sonoras.



Fonte: Autor.

O HATS Kemar apresenta um ganho de magnitude proeminente em torno de 3 kHz para as duas fontes sonoras, comportamento esperado e consistente com a ressonância principal de seu simulador de canal auditivo. O HATS Cortex, por sua vez, exibe a mesma ressonância aguda em aproximadamente 5 kHz que já havia sido identificada na análise de variabilidade dos fones (Seção 4.1), reforçando que esta é uma característica intrínseca do manequim.

Figura 52 – Espectro de magnitude das FRFs entre fonte e receptor utilizando os 3 HATS e as duas fontes sonoras com a correção do fone de ouvido.



Fonte: Autor.

Já a Figura 52 apresenta as mesmas FRFs após a aplicação do filtro de equalização do fone de ouvido HD 58 X, específica para cada manequim. A expectativa inicial era de que este processo fizesse as três curvas convergirem, uma vez que o filtro atua sobre o sinal de excitação, modificando a resposta do sistema composto fone + HATS. Como a correção visa compensar a resposta eletroacústica do fone e sua interação com o acoplamento do manequim, esperava-se que a influência do sistema de reprodução fosse reduzida e que as FRFs apresentassem maior similaridade entre os diferentes manequins.

No entanto, os resultados demonstram que algumas discrepâncias persistem. Uma possível explicação para este fenômeno reside na carga eletroacústica distinta que cada HATS impõe ao transdutor do fone. Como o filtro de equalização é gerado para compensar a interação única de cada par fone–HATS, a correção não anula as diferenças intrínsecas entre as FRFs dos próprios manequins. Isso leva a uma conclusão metodológica crucial: a simples utilização de um fone FEC não é suficiente para garantir a intercambialidade de medições entre diferentes HATSs.

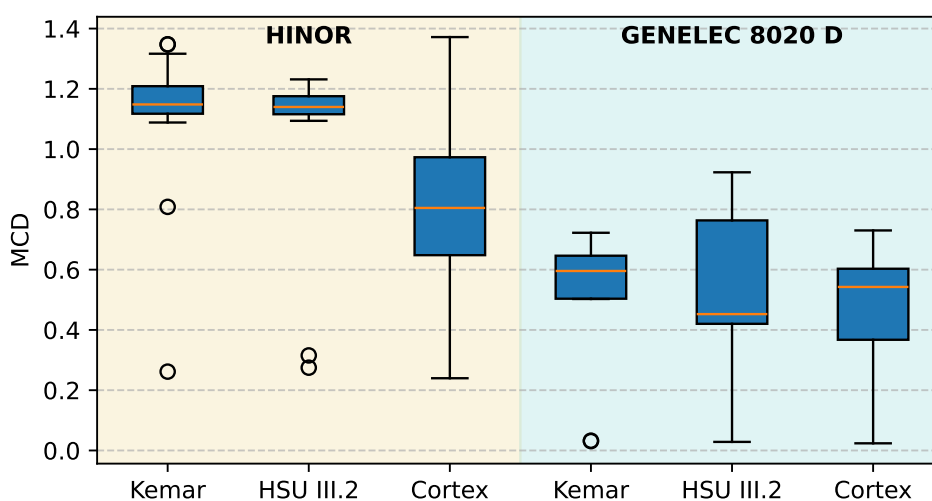
4.2.2 Comparação de Sinais Convoluídos com as Respostas Impulsivas

Para analisar a variabilidade decorrente do reposicionamento dos receptores, foi realizada uma simulação similar à descrita na Seção 3.4. O sinal do compressor a 4500 rpm foi convoluído com o conjunto das 15 respostas impulsivas biauriculares obtidas para cada um dos 3 HATSs e para cada uma das 2 fontes sonoras (Caixa Acústica Hinor e Monitor Genelec 8020 D).

A partir dos sinais convoluídos foi quantificada a diferença entre os sinais por meio da MCD. Para cada conjunto de 15 medições (uma para cada HATS e fonte), o sinal correspondente ao primeiro posicionamento foi definido como referência. Em seguida, a MCD foi calculada entre esta referência e os sinais dos 14 reposicionamentos subsequentes.

A Figura 53 apresenta os boxplots com a distribuição dos valores de MCD obtidos para variabilidade intra-HATS. A análise da figura revela duas tendências principais. Primeiramente, ao comparar as duas fontes, o Monitor Genelec resulta em valores de MCD consistentemente mais baixos e com menor dispersão para todos os HATSs. Isso indica uma maior repetibilidade, que pode ser atribuída às dimensões reduzidas da fonte Genelec, cujo comportamento se aproxima mais de uma fonte pontual, minimizando interações complexas com a geometria dos manequins.

Figura 53 – Variabilidade de reposicionamento do MCD para os 3 HATS e 2 fontes utilizando o sinal do compressor a 4500 rpm.



Fonte: Autor.

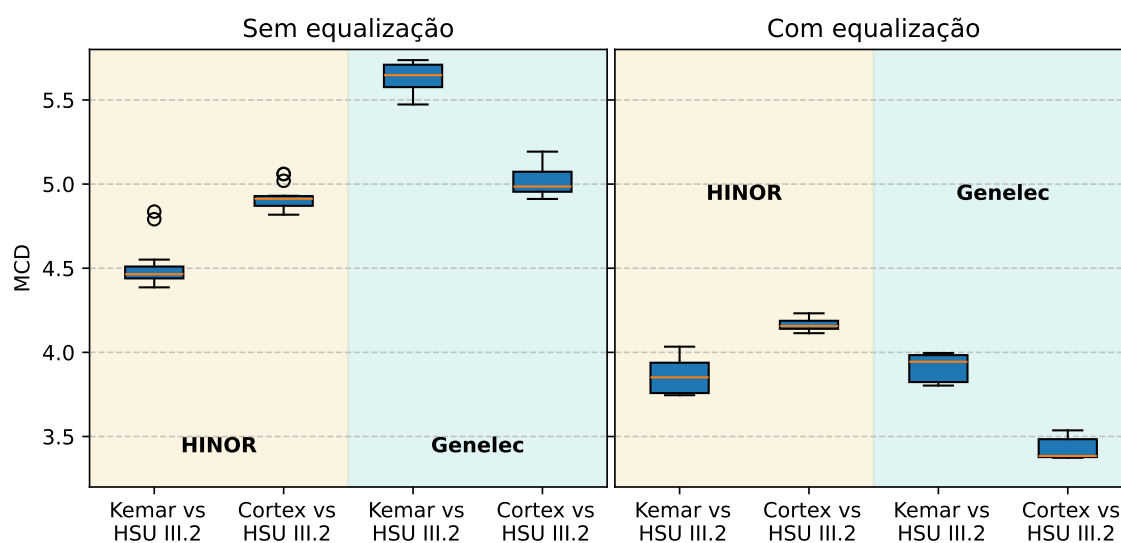
Em segundo lugar, para a Caixa Acústica Hinor, as diferenças entre os manequins são mais pronunciadas. Enquanto Kemar e HSU III.2 apresentam uma variabilidade

similar e mais contida, o **HATS** Cortex restaurado exibe tanto a maior mediana quanto a maior dispersão dos valores. Esse resultado corrobora as análises da Seção 4.1, que já apontavam para uma maior instabilidade nas medições com o manequim Cortex, reforçando a hipótese de que suas características construtivas o tornam os resultados mais sensíveis ao reposicionamento.

Apesar das diferenças observadas entre as condições, é importante notar que os valores absolutos de **MCD** se mantiveram relativamente baixos, tendo valor máximo de 1,4. Isso sugere que, embora existam variações estatisticamente mensuráveis, a variabilidade espectral decorrente reposicionamento dos manequins tende a ser imperceptível.

Já a Figura 54 apresenta os boxplots com a distribuição dos valores de **MCD** entre os **HATSs**, utilizando o HSU III.2 como referência. A análise compara os resultados para as duas fontes sonoras, nas condições sem e com a aplicação da equalização do fone de ouvido.

Figura 54 – Comparação da MCD entre os HATS, utilizando o HSU III.2 como referência, para as condições sem e com equalização do fone de ouvido.



Fonte: Autor.

No painel esquerdo (Sem equalização), observa-se que as maiores diferenças (valores de **MCD** mais elevados) ocorrem com a fonte Genelec. Isso pode ser atribuído à resposta de frequência mais plana deste monitor, que tende a excitar e evidenciar mais as particularidades espectrais de cada manequim.

No painel direito (com equalização), nota-se que a aplicação da correção do fone de ouvido reduz consistentemente a **MCD** em todos os casos. A equalização, ao compensar

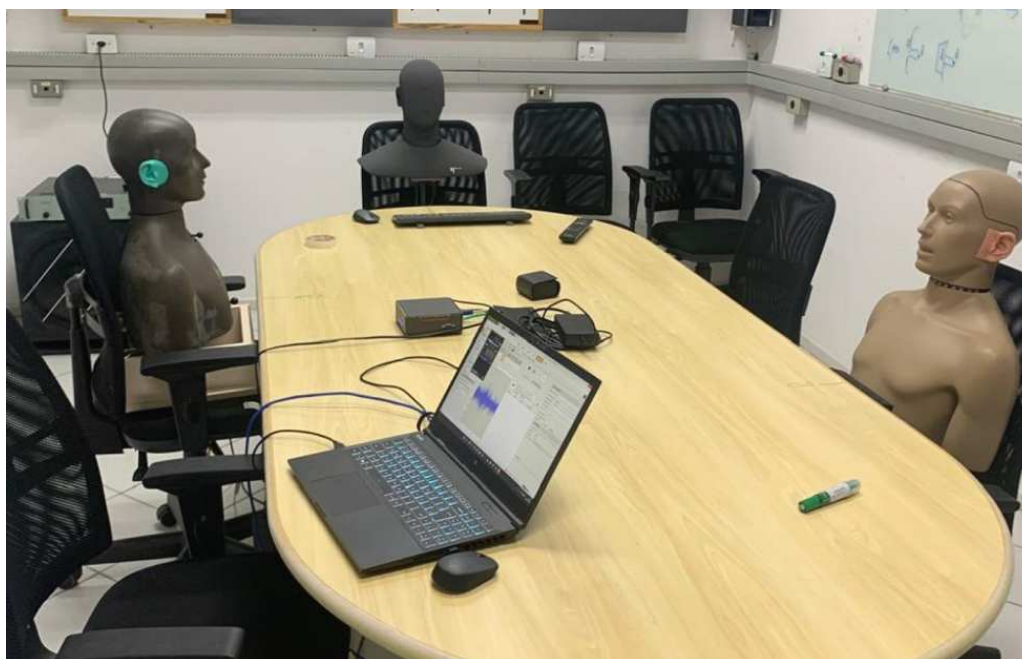
a resposta do sistema fone-HATS, remove uma parte significativa das discrepâncias da cadeia de reprodução, aproximando as curvas. No entanto, a correção não elimina completamente as diferenças, principalmente para a fonte Hinor. Para a MCD a equalização é mais eficiente possivelmente devido à composição espectral do sinal do compressor, que possui componentes predominantes em altas frequências, que fica fora da região de maior discrepância mostrada na Figura 52.

4.3 GRAVAÇÕES NA SALA DE REUNIÕES

A etapa final da avaliação experimental teve como objetivo comparar o desempenho dos três HATSs em um ambiente não anecoico. Para tal, as medições foram realizadas em uma sala de reuniões do LVA, buscando avaliar as diferenças nas BRIRs capturadas sob diferentes ângulos de incidência sonora.

A metodologia consistiu em definir três posições fixas para as fontes sonoras e três posições para os receptores, conforme detalhado no mapa de medições da Figura 57. Em cada uma das nove combinações de fonte-receptor, os HATSs foram intercambiados sequencialmente. Este procedimento garantiu que cada manequim ocupasse as mesmas localizações, permitindo uma comparação direta dos resultados e resultando em um total de 27 BRIRs distintas ($3 \text{ fontes} \times 3 \text{ posições} \times 3 \text{ HATSs}$). A Figura 55 ilustra a montagem com os três manequins posicionados nos pontos de medição definidos na sala.

Figura 55 – HATS posicionados nos pontos de medição.



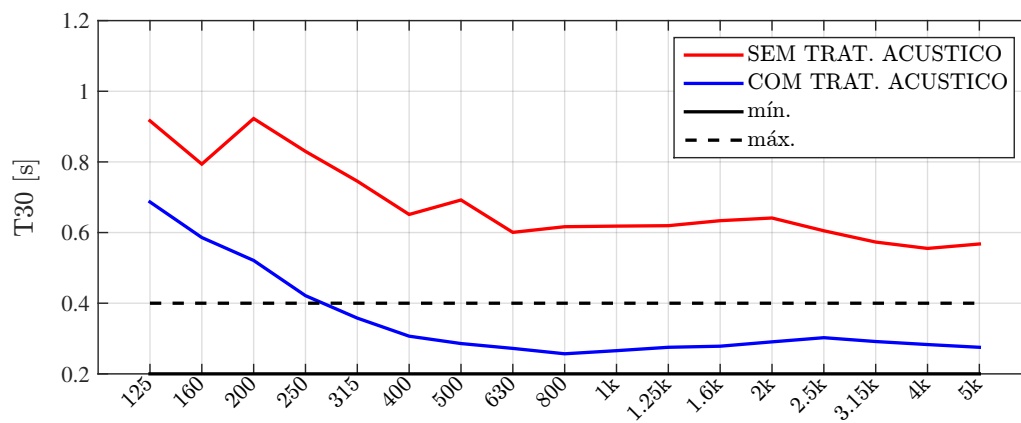
Fonte: Autor.

4.3.1 Características do ambiente

A sala de reuniões na qual as medições foram realizadas possui formato retangular, com dimensões de 5,12 m de comprimento por 3,25 m de largura e 3,36 m de pé-direito, totalizando um volume de aproximadamente 56 m³. O ambiente conta com tratamento acústico, e o tempo de reverberação resultante (T30), apresentado em bandas de terço de

oitava na Figura 56, situa-se em torno de 0,3 s nas frequências médias. A partir destas características, a Frequência de Schroeder para a sala foi estimada em aproximadamente 193 Hz.

Figura 56 – Tempo de reverberação da sala em bandas de terço de oitava.



Fonte: Cortesia Leandro Barbosa (Wave Consultoria).

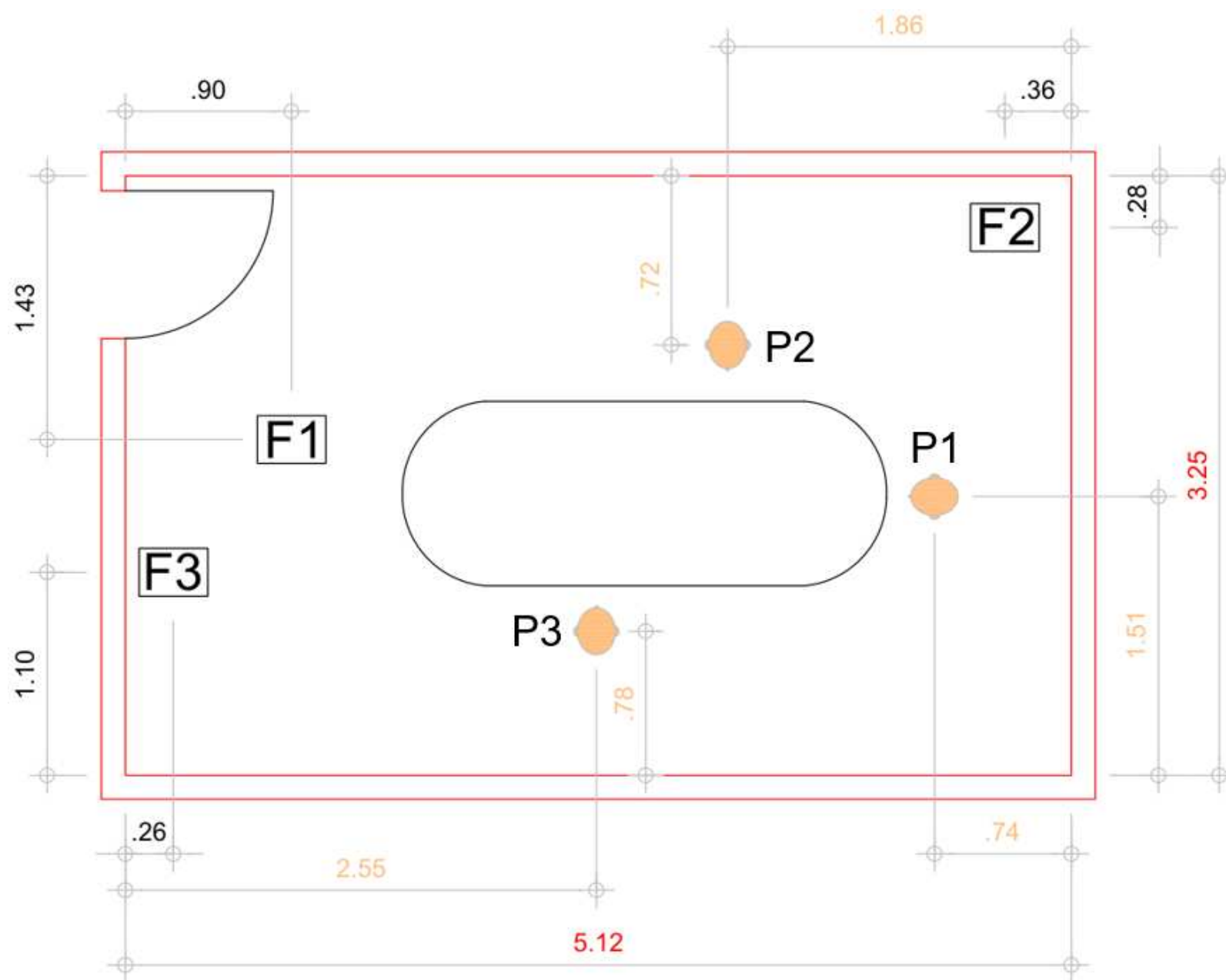


Figura 57 – Mapa de medição na sala de reuniões com as posições de fonte (F1, F2, F3) e receptor (P1, P2, P3).

O ambiente, com acomodação para até oito pessoas, é equipado com um sistema de áudio do tipo *Home Theater* (Sony DAV-DZ77T), ilustrado na Figura 58. Composto por uma unidade central (*receiver*), quatro caixas satélite e dois *subwoofers* em um arranjo 4.2, o sistema foi utilizado como uma das fontes sonoras para as medições de BRIR. Para tal, o sinal de excitação foi enviado a partir da interface de áudio Behringer UCA 222 diretamente para a entrada ótica do *receiver*.

Figura 58 – Sistema de áudio instalado na sala.



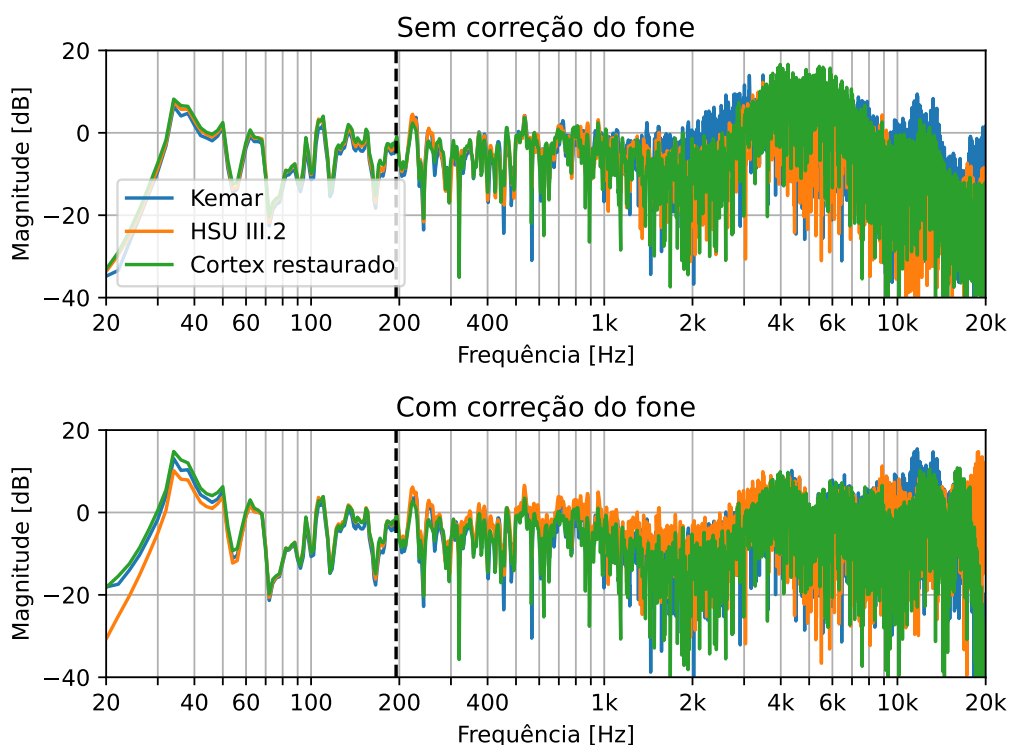
Fonte: Autor

4.3.2 FRFs entre fontes sonoras e HATS obtidas no experimento

As Figuras 59 e 60 apresentam as respostas em frequência para os três HATSs na Posição 1, com excitação pelo monitor de áudio Genelec, representadas em banda estreita e em bandas de 1/6 de oitava, respectivamente, sendo exibidos apenas os dados do canal esquerdo.

A representação em banda estreita (Figura 59) permite observar as características modais da sala, com ressonâncias bem definidas em baixas frequências (baixa densidade modal) e um comportamento mais complexo em altas frequências, acima da Frequência de Schroeder indicada.

Figura 59 – Resposta em frequência em banda estreita (lado esquerdo, $\Delta f = 2$ Hz) para os três HATS na posição 1 utilizando o monitor de áudio Genelec para excitar o ambiente. Linha tracejada preta mostra a Frequência de Schroeder da sala.



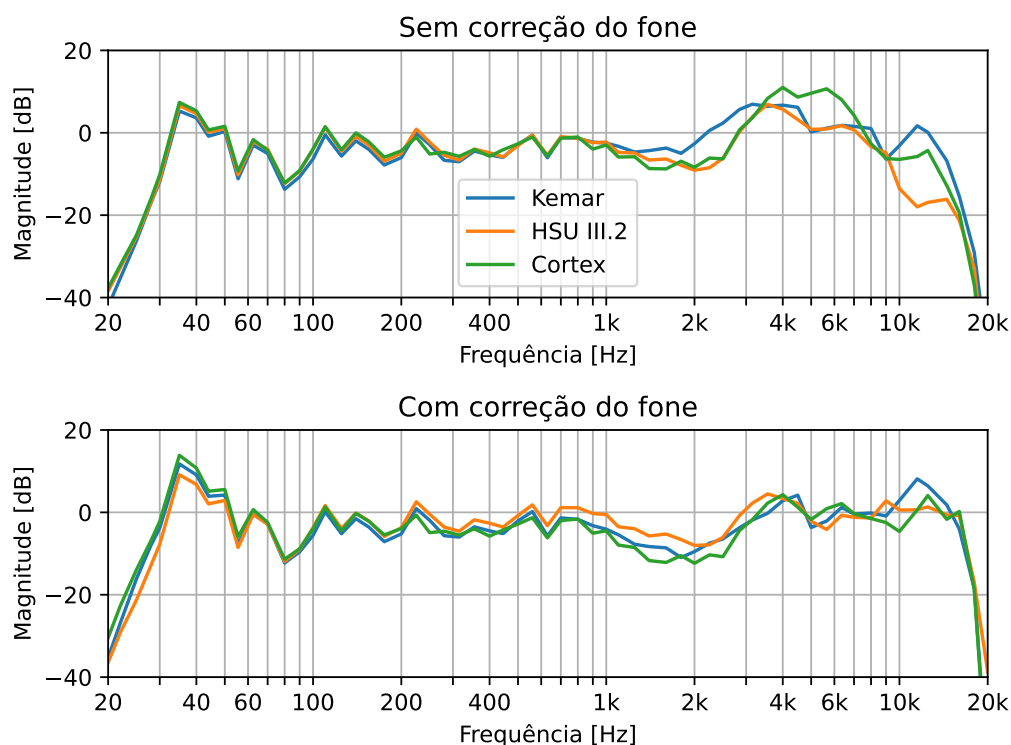
Fonte: Autor.

Ao comparar a diferença entre os receptores, os resultados confirmam a tendência observada na câmara semi-anecoica: após a aplicação da correção do fone de ouvido, a ressonância do canal auditivo do Kemar é atenuada, reduzindo a magnitude da [FRF](#) nesta região. Esse efeito pode ser melhor visualizado na [Figura 60](#), havendo maior concordância entre as curvas, principalmente na região acima de 1 kHz.

A [Figura 60](#)¹, por sua vez, ao suavizar as flutuações modais, facilita a comparação da tendência espectral entre os manequins. Os resultados confirmam o comportamento observado na câmara semi-anecoica: no painel superior (sem correção), as curvas divergem, cada uma com sua assinatura espectral. No painel inferior (com correção), o efeito da equalização é notável. Além de atenuar a ressonância do canal auditivo do Kemar, a correção promove uma maior concordância entre as três curvas, especialmente na região acima de 1 kHz.

¹ Todas as FRFs são mostradas no Apêndice [B](#).

Figura 60 – Resposta em frequência em bandas de 1/6 de oitava (lado esquerdo) para os três HATS na posição 1 utilizando o monitor de áudio Genelec para excitar o ambiente.

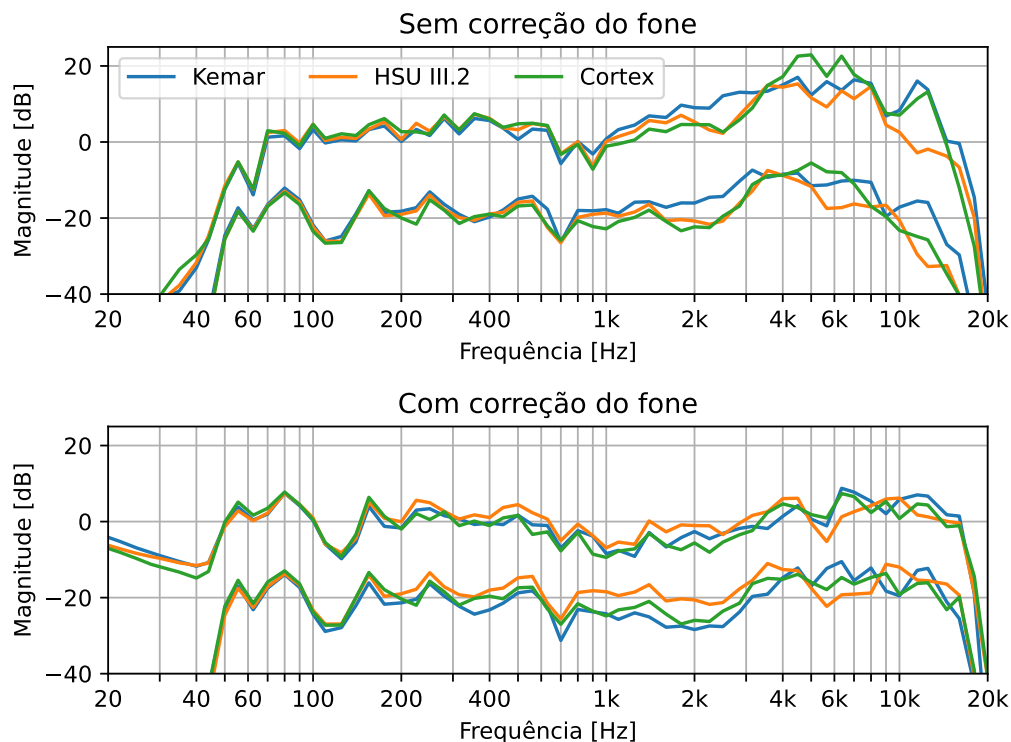


Fonte: Autor.

Para avaliar a robustez do método de calibração dos fones em um cenário mais complexo, foi analisada uma condição de escuta assimétrica, com os HATSs posicionados no Ponto 2, onde a orelha direita estava mais diretamente exposta à fonte sonora (sistema de áudio Sony). A Figura 61 apresenta as FRFs biauriculares para essa configuração, antes e depois da correção do fone.

No painel superior (sem correção), a assimetria é evidente: o canal direito de todos os HATSs (curvas deslocadas em -20 dB) apresenta maior energia em altas frequências devido à menor obstrução pela cabeça (sombra acústica). O resultado mais revelador, contudo, está no painel inferior (com correção). Ao contrário do observado no caso simétrico (Figura 60), a aplicação da equalização nesta condição não resultou em uma convergência clara das curvas. As discrepâncias entre os HATSs permanecem proeminentes e, em certas regiões, são até mesmo realçadas. Nota-se, por exemplo, que na faixa entre 1 e 4 kHz, a resposta do canal direito do Kemar é visivelmente atenuada em comparação com a do HSU III.2, que se torna a de maior magnitude.

Figura 61 – Respostas em frequência (FRFs) biauriculares para os três HATS na Posição 2, utilizando o sistema de áudio da sala como fonte sonora. Para facilitar a visualização, a curva do canal direito foi deslocada artificialmente em -20 dB em relação à do canal esquerdo.



Fonte: Autor.

Esse resultado sugere que a eficácia da equalização de fones é limitada em cenários acusticamente complexos e direcionais. Embora a correção possa equalizar a resposta do fone em si, os resultados sugerem que ela não é capaz de compensar as diferenças fundamentais na captação direcional (inerentes à [HRTF](#) de cada manequim) em um campo sonoro assimétrico. Isso reforça a conclusão de que, mesmo com um fone FEC devidamente equalizado, as características morfológicas e eletroacústicas do [HATS](#) continuam sendo um fator dominante na caracterização de condições com fortes pistas espaciais.

4.3.3 Análise da MCD entre áudios convoluídos

Para realizar uma análise quantitativa abrangente, os três sinais anecoicos de compressor (2400, 3000 e 4500 rpm) foram convoluídos com todas as [BRIRs](#) medidas na sala de reuniões. Subsequentemente, a [MCD](#) foi calculada para cada sinal resultante, comparando sempre os resultados dos [HATSs](#) Kemar e Cortex com os do HSU III.2, que foi utilizado como referência.

Este procedimento gerou um extenso conjunto de 216 valores de **MCD**, abrangendo todas as combinações de fontes, posições, sinais e estados de equalização. A tabela completa com todos estes dados é apresentada no Apêndice B, Tabela 11. As análises a seguir se concentrarão em extrair as tendências mais relevantes deste conjunto de dados.

4.3.3.1 Análise da Média Interauricular na Comparação entre HATS através da **MCD**

Para realizar uma análise global das diferenças, a Tabela 9 sintetiza os dados apresentando a média interauricular da **MCD** para todas as condições de medição. A tendência mais clara e consistente é o efeito da equalização: em praticamente todos os cenários, a aplicação da correção do fone de ouvido reduz os valores de **MCD**, o que era esperado e valida a eficácia do processo em mitigar as disparidades entre as cadeias de medição.

Tabela 9 – Médias interaurais da MCD entre os HATS de teste (Kemar, Cortex) e o HATS de referência (HSU III.2) para todas as condições de medição.

			Posição 1			Posição 2			Posição 3		
			Genelec	Hinor	Sony	Genelec	Hinor	Sony	Genelec	Hinor	Sony
Sem correção	2400 rpm	Cortex	4.82	5.82	5.02	4.87	6.52	4.61	4.73	6.20	6.85
		Kemar	5.63	6.79	6.14	5.62	7.46	5.89	6.26	6.94	6.56
	3000 rpm	Cortex	4.67	5.57	4.95	4.78	6.10	4.73	4.58	6.30	6.91
		Kemar	5.74	6.80	6.29	5.45	7.24	5.73	5.87	6.75	6.77
	4500 rpm	Cortex	5.01	5.73	5.39	4.86	6.16	4.69	4.70	6.40	6.70
		Kemar	5.76	6.82	6.73	5.26	7.37	6.04	5.94	6.98	6.58
Com correção	2400 rpm	Cortex	4.54	5.49	5.04	4.62	5.49	4.55	4.69	5.80	5.89
		Kemar	5.13	5.42	5.78	4.81	4.94	4.56	4.69	4.96	5.12
	3000 rpm	Cortex	4.74	5.36	5.12	4.53	5.48	4.81	4.43	5.87	6.06
		Kemar	5.42	5.76	6.01	4.63	4.57	4.71	4.83	5.01	5.43
	4500 rpm	Cortex	4.98	5.52	5.33	4.80	5.57	4.58	4.66	6.32	5.96
		Kemar	5.26	5.63	5.86	5.04	5.05	4.96	4.72	5.17	5.33

Fonte: Autor.

Como um exemplo notável, a maior diferença observada para o **HATS** Kemar na condição sem correção apresenta um **MCD** de 7,46, na Posição 2 com a fonte Hinor e o sinal de 2400 rpm que foi reduzida para 4,57 após a equalização. Apesar dessa melhora geral, a tabela também evidencia que as diferenças relativas entre as fontes e posições se mantêm, indicando que a assinatura acústica de cada condição é atenuada, mas não completamente eliminada pelo processo e calibração da cadeia de gravação e reprodução dos sinais binaurais.

4.3.3.2 Análise da Assimetria Interauricular na Comparação entre HATS através da MCD

Para finalizar a análise, foi investigado se a diferença espectral entre o **HATS** de referência (HSU III.2) e os outros manequins (Kemar e Cortex) era simétrica entre os canais esquerdo e direito. Para isso, calculou-se a diferença interauricular da **MCD**, definida como a diferença entre os valores obtidos para os canais direito e esquerdo. Valores próximos de zero indicam simetria entre os canais (erro interauricular reduzido), enquanto valores positivos ou negativos elevados indicam assimetrias que podem impactar o equilíbrio da imagem sonora.

Matematicamente, a diferença interauricular da MCD foi obtida por:

$$\Delta_{\text{MCD}}^{(\text{HATS})} = \text{MCD}_R(\text{HSU}, \text{HATS}) - \text{MCD}_L(\text{HSU}, \text{HATS}), \quad (4.1)$$

em que MCD_R e MCD_L representam, respectivamente, os valores da MCD entre o **HATS** de referência e o manequim de teste para os canais direito e esquerdo. A Tabela ?? apresenta os resultados para todas as condições de medição e respectivas convoluções com os sinais de compressores nas diferentes posições de microfone.

A análise dos resultados sem correção revela que a combinação do **HATS** Kemar com a fonte sonora Sony consistentemente gera as maiores assimetrias interauriculares, com valores que chegam a 4.81. Este efeito é particularmente pronunciado na Posição 3, sugerindo uma forte interação entre a geometria do Kemar, sua assimetria em relação à fonte sonora, já que sua orelha esquerda está voltada para os alto-falantes.

O resultado mais significativo, no entanto, é o efeito da equalização do fone de ouvido (painel inferior). A aplicação da correção reduz de forma significativa a magnitude das diferenças interauriculares em quase todas as condições. Isso indica que o processo de equalização individual, ao corrigir a resposta em frequência de cada canal individualmente, compensa as assimetrias presentes na interação fone-**HATS**. Embora a diferença espectral entre os manequins não seja totalmente eliminada, o erro residual após a correção torna-se significativamente mais equilibrado entre os canais esquerdo e direito, sendo um fator fundamental para garantir a estabilidade e o correto posicionamento da imagem sonora na reprodução binauricular.

Tabela 10 – Diferenças interauriculares da MCD (Direito - Esquerdo) entre os HATS de teste (Kemar, Cortex) e o HATS de referência (HSU III.2) para todas as condições de medição.

			Posição 1			Posição 2			Posição 3		
			Genelec	Hinor	Sony	Genelec	Hinor	Sony	Genelec	Hinor	Sony
Sem correção	2400 rpm	Cortex	-0.61	-1.41	-0.39	-0.36	-1.25	-0.16	-0.13	-0.49	1.15
		Kemar	1.92	1.66	0.79	1.31	2.24	2.38	4.09	2.15	4.43
	3000 rpm	Cortex	-0.16	-1.40	-0.30	-0.39	-0.63	0.13	-0.47	-0.36	1.33
		Kemar	1.32	1.89	1.30	1.35	2.46	1.66	3.34	1.71	4.81
	4500 rpm	Cortex	-0.66	-1.47	0.02	-0.20	-1.41	-0.30	-0.46	0.09	0.53
		Kemar	1.05	0.86	0.70	0.56	1.61	1.30	3.15	1.74	4.28
Com correção	2400 rpm	Cortex	-0.36	-0.86	0.39	-0.10	-0.64	-0.46	-0.36	-0.73	0.24
		Kemar	-0.01	-0.37	0.38	0.73	0.35	-0.67	0.71	0.16	1.81
	3000 rpm	Cortex	-0.24	-0.73	0.27	-0.37	-0.77	0.10	-0.62	-0.58	0.36
		Kemar	-0.08	-0.52	0.94	0.66	0.42	-1.01	1.20	0.88	1.74
	4500 rpm	Cortex	-0.45	-0.87	0.34	-0.62	-1.55	-0.08	-0.41	-0.20	0.13
		Kemar	0.19	-0.19	0.68	0.08	0.25	-0.48	0.05	-0.33	1.67

Fonte: Autor.

5 CONSIDERAÇÕES FINAIS

Este trabalho investigou a influência de diferentes receptores ([HATSs](#) e microfone) na gravação de sinais biauriculares de sistemas de refrigeração de uso doméstico. Foram obtidas os parâmetros utilizados para avaliar qualidade sonora desses produtos bem como implementados testes subjetivos para avaliar a diferença percebida entre estímulos sonoros de mesmas fontes sonoras obtidas pelos diferentes receptores.

No primeiro estudo realizado, foram comparados sinais obtidos com microfone e [HATS](#) em câmara semi-anecoica. A configuração experimental permitiu avaliar a estabilidade operacional das fontes mecânicas (compressor hermético e refrigerador) e a variabilidade ocasionada pela reposição do receptor ao utilizar fonte eletroacústica para excitar o campo acústico.

Os resultados dos espectros obtidos evidenciaram que as diferenças em termos espectrais abaixo de 1 kHz podem ser atribuídas à variabilidade operacional das fontes mecânicas. Acima dessa frequência, a [HRTF](#) referente a uma posição fonte-receptor não contempla os efeitos da reflexão do piso para todas as fontes sonoras utilizadas. Em relação aos parâmetros psicoacústicos, foram observadas diferenças sistemáticas entre [HATS](#) e microfone a partir da análise estatística dos resultados obtidos. Algumas diferenças excedem os valores de [DLO](#) apresentados na literatura. Porém, para o *roughness*, esses valores não são representativos em função da diferença dos modelos utilizados.

Para avaliar se essas diferenças são perceptíveis por sujeitos, foi implementada uma ferramenta para avaliar o limiar de discriminação entre sinais sonoros, integrando o método [PEST](#) a uma técnica de correção ponderada baseada em filtragem espectral. O método desenvolvido teve como principal foco reduzir o número de iterações em testes subjetivos ao concentrar a atenção do ouvinte na zona de maior incerteza perceptiva, sendo mais eficiente que métodos convencionais como 3AFC ou ABX.

A partir dos fatores críticos de correção espectral (α_{crit}) foi possível obter as respectivas diferenças de *loudness* (ΔN), *sharpness* (ΔS), *roughness* (ΔR) e [MCD](#). Dentre os modelos avaliados, a [MCD](#) apresentou a melhor correspondência com os valores críticos, revelando baixa variabilidade entre quatro dos cinco pares de sinais analisados, sendo o par referente ao refrigerador classificado como *outlier*. A partir desses quatro pares, foi possível estimar uma [DLO](#) e a respectiva zona de confusão para o parâmetro.

A fim de quantificar as diferenças perceptivas entre os sinais, foi elaborada uma

adaptação do teste **MUSHRA** utilizando valores de correção espectral predefinidos. No teste piloto, esses valores foram escolhidos com base em um modelo de diferença perceptiva definido *a priori*, o qual revelou que a **MCD** foi o parâmetro dominante nas análises de correlação preliminares. Assim, os valores de correção espectral foram ajustados para o teste final de modo a garantir equidistância em relação à **MCD**, com passos próximos à **DLO** estimada no teste anterior. Os resultados obtidos corroboram que esse parâmetro é um excelente preditor de diferença perceptiva, apresentando correlação mínima de $R^2 = 0,96$, permitindo concluir que a **MCD** se destaca como o modelo mais sensível e consistente para representar a percepção de diferença entre os sinais analisados.

A partir desses resultados, foi elaborado um novo estudo com objetivo de investigar a diferença de sinais gravados em campo livre e em ambiente típico a partir de diferentes **HATSs**. Além disso, foram avaliados efeitos da equalização de um fone de ouvido com os diferentes **HATSs** utilizados. Os critérios de abertura revelaram o fone Sennheiser HD 58 X como o modelo mais adequado dentre os caracterizados para a reprodução binaural dentre os medidos.

Em ambiente controlado, as medições em câmara semi-anecoica permitiram avaliar isoladamente os efeitos da resposta dos diferentes **HATSs** e a influência da equalização dos fones de ouvido. Na ausência de equalização, verificou-se que fontes com resposta espectral mais ampla, como a Genelec, tendem a evidenciar com maior clareza as diferenças entre os manequins, resultando em valores mais elevados de **MCD**. A aplicação da equalização reduziu essas discrepâncias de forma consistente, compensando as variações na resposta conjunta do sistema fone-HATS e aproximando as curvas entre os dispositivos. Ainda assim, pequenas variações persistiram, especialmente para a fonte Hinor, indicando que, embora eficaz, a correção espectral não elimina completamente as particularidades acústicas de cada manequim.

Em ambiente não anecoico, as gravações realizadas na sala de reuniões revelaram que a assimetria das fontes sonoras em relação ao **HATS** exerce influência significativa sobre as respostas obtidas. Quando a fonte não se encontrava alinhada com o eixo de simetria do manequim, observou-se maior discordância entre os canais esquerdo e direito, refletindo o impacto da incidência angular e da interação com as reflexões do ambiente. A equalização dos fones, embora útil para reduzir as diferenças espectrais médias, mostrou-se menos efetiva sob essas condições, sugerindo que, em contextos reais, a variabilidade espacial e as características do ambiente passam a dominar o comportamento das medidas.

O conjunto de experimentos realizados nesta pesquisa permitiu compreender de forma integrada as relações entre medições objetivas e diferenças perceptivas em sistemas de reprodução binauricular. A implementação do método de correção espectral ponderada e a definição de um fator crítico de ajuste (α_{crit}) mostraram-se eficazes para quantificar diferenças entre sinais gravados por diferentes **HATS** e microfones, possibilitando a obtenção de parâmetros consistentes com as respostas auditivas humanas. Dentre os modelos avaliados, o **MCD** destacou-se como um achado relevante por se mostrar a mais sensível e estável na representação das variações perceptivas entre os sinais. Os parâmetros psicoacústicos de *loudness*, *sharpness* e *roughness*, consolidadas na literatura sobre qualidade sonora de sistemas de refrigeração, foram utilizadas como referência complementar e reforçaram a coerência dos resultados sob diferentes condições experimentais.

Do ponto de vista prático, os resultados contribuem para o avanço das padronizações de medições acústicas voltadas à qualidade sonora, ao demonstrar a importância de uma calibração espectral criteriosa e da escolha adequada do **HATS** e do fone de referência. A abordagem proposta fornece uma base quantitativa e perceptualmente validada que pode subsidiar procedimentos de comparação entre laboratórios e apoiar o desenvolvimento de normas e protocolos para avaliação de fidelidade sonora e auralização. Em síntese, o trabalho consolida um elo entre medições objetivas e percepção auditiva, contribuindo para a evolução das metodologias experimentais em acústica aplicada à percepção e à qualidade sonora.

5.1 SUGESTÕES PARA TRABALHOS FUTUROS

Com base nas evidências e contribuições apresentadas ao longo deste trabalho, identificam-se diversas possibilidades de continuidade que podem aprofundar a compreensão dos mecanismos envolvidos na percepção de diferenças sonoras e aprimorar os procedimentos experimentais adotados. Nesse sentido, destacam-se algumas direções de pesquisa que podem orientar investigações futuras:

- Realização de testes subjetivos adicionais voltados à avaliação da variabilidade intra e inter-sujeito no estudo já implementado, buscando maior robustez estatística na estimativa dos limiares perceptivos;
- Aplicação da metodologia proposta à comparação entre respostas impulsivas obtidas

em salas reais e simuladas, visando investigar a validade perceptiva de modelos de aurilização em diferentes condições acústicas;

- Estudo comparativo entre filtros FIR e IIR quanto à eficiência e estabilidade na equalização biauricular para aplicações em tempo real.
- Validação experimental do [DLO](#) utilizando o método PEST combinado com abordagens de escolha forçada, como ABX ou 3AFC para obter uma função psicométrica mais precisa;
- Desenvolvimento de testes subjetivos de incômodo aplicados a versões distintas de uma mesma fonte sonora, de modo a investigar se o modo de gravação e reprodução influencia a percepção de incômodo sonoro;
- Exploração da manipulação espectral controlada a partir da [MCD](#) para avaliar versões de um mesmo sinal excitando diferentes [SLITs](#);
 - Implementação de manipulações espectrais semelhantes às deste trabalho para sinais estacionários;
 - Extensão da abordagem para sinais não estacionários, empregando processamento em quadros espectrais e reconstrução por *overlap-and-add*;
- Investigação de critérios de oclusão de fones de ouvido por meio de medições com acopladores do tipo *occluded ear simulator* (OI) e tubos de impedância;
- Comparação entre modelos de qualidade sonora aplicados a sistemas de refrigeração e modelos baseados em similaridade perceptiva, de modo a avaliar a relevância das diferenças entre ambas as abordagens.

BIBLIOGRAFIA

ALGAZI, V. R.; AVENDANO, C.; THOMPSON, D. Dependence of subject and measurement position in binaural signal acquisition. **Journal of the Audio Engineering Society**, Audio Engineering Society, v. 47, n. 11, p. 937–947, 1999.

ALGAZI, V. R.; DUDA, R. O.; THOMPSON, D. M.; AVENDANO, C. The cipic hrtf database. *In*: IEEE. PROCEEDINGS of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No. 01TH8575). 2001. p. 99–102.

ALLEN, J. B.; BERKLEY, D. A. Image method for efficiently simulating small-room acoustics. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 65, n. 4, p. 943–950, 1979.

AMERICAN NATIONAL STANDARDS INSTITUTE / ACOUSTICAL SOCIETY OF AMERICA. **ANSI/ASA S3.36-2012: Specification for a Manikin for Simulated in-situ Airborne Acoustic Measurements**. New York, NY, USA, 2012.

ANTONIOU, A. **Digital Signal Processing**. New York, NY: McGraw-Hill Professional, out. 2005.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR IEC 61260-1:2023 – Eletroacústica — Filtros de banda de oitava e fração de oitava — Parte 1: Especificações**. set 2023. <https://www.abntcatalogo.com.br/>. Primeira edição, publicada em 19 de setembro de 2023.

BARNWELL III, T. P.; CLEMENTS, M.; QUACKENBUSH, S. **Objective measures for speech quality testing**. Prentice Hall, Englewood Cliffs, New Jersey, 1988.

BENDAT, J. S.; PIERSON, A. G. **Random data: analysis and measurement procedures**. John Wiley & Sons, 2011.

BERANEK, L. L.; MELLOW, T. **Acoustics: sound fields and transducers**. Academic Press, 2012.

BERZBORN, M.; BOMHARDT, R.; KLEIN, J.; RICHTER, J.-G.; VORLÄNDER, M. The ITA-Toolbox: An open source MATLAB toolbox for acoustic measurements and signal processing. *In*: PROCEEDINGS of the 43th Annual German Congress on Acoustics, Kiel, Germany. 2017. v. 6.

BLAUERT, J. **Spatial hearing**. 2. ed. London, England: MIT Press, nov. 1996.

BLAUERT, J.; JEKOSCH, U. Sound-quality evaluation—a multi-layered problem. **Acta acustica united with acustica**, S. Hirzel Verlag, v. 83, n. 5, p. 747–753, 1997.

BOLEY, J.; LESTER, M. Statistical analysis of ABX results using signal detection theory. *In*: 10. AUDIO engineering society convention. 2009. v. 127.

BRANDÃO, E. **Acústica de Salas: projeto e modelagem**. [S.l.]: Edgard Blucher, 2016.

BRINKMANN, F.; ASPÖCK, L.; ACKERMANN, D.; LEPA, S.; VORLÄNDER, M.; WEINZIERL, S. A round robin on room acoustical simulation and auralization. **The Journal of the Acoustical Society of America**, AIP Publishing, v. 145, n. 4, p. 2746–2760, 2019.

BRINKMANN, F.; KREUZER, W.; THOMSEN, J.; DOMBROVSKIS, S.; POLLACK, K.; WEINZIERL, S.; MAJDAK, P. Recent advances in an open software for numerical HRTF calculation. **Journal of the Audio Engineering Society**, Audio Engineering Society, n. 7/8, p. 502–514, 2023.

BRUM, R. **Estudo da influência de diferentes técnicas de auralização na avaliação da qualidade sonora de produtos**. 2020. f. 102. Dissertação de Mestrado – Universidade Federal de Santa Catarina, Florianópolis.

BUCKLEIN, R. The audibility of frequency response irregularities. **Journal of the Audio Engineering Society**, v. 29, n. 3, p. 126–131, 1981.

BURKHARD, M.; SACHS, R. Anthropometric manikin for acoustic research. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 58, n. 1, p. 214–222, 1975.

CHOUEIRI, E. Y. Optimal crosstalk cancellation for binaural audio with two loudspeakers. **Princeton University**, v. 28, 2008.

CLARK, D. High-resolution subjective testing using a double-blind comparator. **Journal of the Audio Engineering Society**, Audio Engineering Society, v. 30, n. 5, p. 330–338, 1982.

COX, T.; D'ANTONIO, P. **Acoustic absorbers and diffusers**. 3. ed. London, England: CRC Press, set. 2020.

DAL MORO, P. M. M. **Estudo da influência do ambiente de instalação e da condição de montagem na percepção sonora de refrigerador doméstico**. 2019. Dissertação (Mestrado) – Universidade Federal de Santa Catarina, Florianópolis, SC.

DANIEL, P.; WEBER, R. Psychoacoustical roughness: Implementation of an optimized model. **Acta Acustica**, v. 83, p. 113–123, 1997.

DEMOLINER, G. **Avaliação da qualidade sonora de um ciclo típico de operação de um refrigerador**. 2017. Dissertação (Mestrado) – Universidade Federal de Santa Catarina, Florianópolis, SC.

DEUTSCHES INSTITUT FÜR NORMUNG. **DIN 45692:2009-08 - Measurement technique for the simulation of the auditory sensation of sharpness**. Published on 1.8.2009. Berlin, Germany, ago. 2009.

DI, G. An improved psychoacoustic annoyance model based on tonal noises. *In*: INSTITUTE OF NOISE CONTROL ENGINEERING, 8. INTER-NOISE and NOISE-CON Congress and Conference Proceedings. 2016. v. 253, p. 540–549.

EARGLE, J. **Loudspeaker Handbook**. London, England: Chapman e Hall, jan. 1997.

ENQVIST, P.; KARLSSON, J. Minimal Itakura-Saito distance and covariance interpolation. *In*: IEEE. 2008 47th IEEE Conference on Decision and Control. 2008. p. 137–142.

FARINA, A. Advancements in impulse response measurements by sine sweeps. *In*: AUDIO ENGINEERING SOCIETY. AUDIO engineering society convention 122. 2007.

FARINA, A. Silence Sweep: a novel method for measuring electro-acoustical devices. *In*: PRESENTED at the 126th Convention of the Audio Engineering Society. Munich, Germany, maio 2009.

FARINA, A. *et al.* Simultaneous measurement of impulse response and distortion with a swept-sine technique. **Preprints-Audio Engineering Society**, Audio Engineering Society; 1999, 2000.

FASTL, H.; ZWICKER, E. **Psychoacoustics: facts and models**. Springer Science & Business Media, 2007. v. 22.

FRIEDLANDER, B.; PORAT, B. The modified Yule-Walker method of ARMA spectral estimation. **IEEE Transactions on Aerospace and Electronic Systems**, IEEE, n. 2, p. 158–173, 2007.

FRITZ, C.; CROSS, I.; MOORE, B. C.; WOODHOUSE, J. Perceptual thresholds for detecting modifications applied to the acoustical properties of a violin. **The Journal of the Acoustical Society of America**, AIP Publishing, v. 122, n. 6, p. 3640–3650, 2007.

GELFAND, S. A. **Hearing: An introduction to psychological and physiological acoustics**. CRC Press, 2017.

GRAY, A.; MARKEL, J. Distance measures for speech processing. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, IEEE, v. 24, n. 5, p. 380–391, 2003.

HAMMERSHØI, D.; MØLLER, H. Methods for binaural recording and reproduction. **Acta Acustica united with Acustica**, S. Hirzel Verlag, v. 88, n. 3, p. 303–311, 2002.

HAMMERSHØI, D.; MØLLER, H. Sound transmission to and within the human ear canal. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 100, n. 1, p. 408–427, 1996.

HARTMANN, W. M. **Signals, sound, and sensation**. Springer Science & Business Media, 2004.

HAYES, M. H. **Statistical digital signal processing and modeling**. John Wiley & Sons, 1996.

HE, Z. **Digital Signal Processing: Theory and Application**. People's Posts & Telecommunications Press, Beijing, 1983.

INGLE, V.; KOGON, S.; MANOLAKIS, D. **Statistical and adaptive signal processing**. Artech, 2005.

INTERNATIONAL ELECTROTECHNICAL COMMISSION. **IEC 60318-1:2009 Electroacoustics – Simulators of human head and ear – Part 1: Ear simulator for the measurement of supra-aural and circumaural earphones**. Geneva, Switzerland, 2009.

INTERNATIONAL ELECTROTECHNICAL COMMISSION. **IEC 60318-4:2010 Electroacoustics – Simulators of human head and ear – Part 4: Occluded-ear simulator for the measurement of earphones coupled to the ear by means of ear inserts**. Geneva, Switzerland, 2010.

INTERNATIONAL ELECTROTECHNICAL COMMISSION. **IEC 60318-7:2022 Electroacoustics – Simulators of human head and ear – Part 7: Head and torso simulator for the measurement of sound sources close to the ear**. Geneva, Switzerland, 2022.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **Acoustics – Methods for calculating loudness – Part 1: Zwicker method**. 2017. Last reviewed and confirmed in 2022. This version remains current.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **Acoustics – Methods for calculating loudness – Part 2: Moore-Glasberg method**. 1. ed. 2017. ISO Standard. Last reviewed and confirmed in 2022.

ISO 3741:2010. **Acoustics — Determination of sound power levels and sound energy levels of noise sources using sound pressure — Precision methods for reverberation test rooms**. International Organization for Standardization Geneva, Switzerland, 2010.

ISO 3745:2012. **Acoustics — Determination of sound power levels and sound energy levels of noise sources using sound pressure — Precision methods for anechoic rooms and hemi-anechoic rooms**. International Organization for Standardization Geneva, Switzerland, 2012.

ISO 8253-2. **Acoustics – Audiometric test methods – Part 2: Sound field audiometry with pure-tone and narrow-band test signals**. Geneva, Switzerland, 2009.

ISO, BSEN; STANDARD, BRITISH. **ISO 3382-2. Acoustics—Measurements of Room Acoustics Parameters - Part 2: Reverberation time in ordinary rooms**. 2009.

ITAKURA, F.; SAITO, S. Analysis synthesis telephony based on the maximum likelihood method. *In: PROCEEDINGS of the 6th International Congress on Acoustics*. Los Alamitos, CA: IEEE, 1968. p. c-17–20.

ITU R BS1116-3. **Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems**. Fev. 2015. Series BS – Broadcasting service (sound).

ITU-R BS.1534-3. **Recommendation ITU-R BS.1534-3: Method for the subjective assessment of intermediate quality level of audio systems**. 2014.

JEON, J. Y.; YOU, J.; CHANG, H. Y. Sound radiation and sound quality characteristics of refrigerator noise in real living environments. **Applied acoustics**, Elsevier, v. 68, n. 10, p. 1118–1134, 2007.

JIANG, X.; NING, L.; GEORGIU, T. T. Distances and Riemannian metrics for multivariate spectral densities. **IEEE Transactions on Automatic Control**, IEEE, v. 57, n. 7, p. 1723–1735, 2012.

KAERNBACH, C. *et al.* Simple adaptive testing with the weighted up-down method. **Perception & psychophysics**, v. 49, n. 3, p. 227–229, 1991.

KANETADA, N.; YAMAMOTO, R.; MIZUMACHI, M. Evaluation of sound quality of high resolution audio. **Proc. IEEE/IIAE International Conference on Intelligent Systems and Image Processing**, 2013.

KARJALAINEN, M.; PIIRILÄ, E.; JÄRVINEN, A.; HOUPANIEMI, J. Comparison of loudspeaker equalization methods based on DSP techniques. *In: AUDIO ENGINEERING SOCIETY. AUDIO Engineering Society Convention 102.* 1997.

KOMINEK, J.; BLACK, A. W. The CMU Arctic speech databases. *In: SSW.* 2004. p. 223–224.

KUBICHEK, R. Mel-cepstral distance measure for objective speech quality assessment. *In: IEEE. PROCEEDINGS of IEEE pacific rim conference on communications computers and signal processing.* 1993. v. 1, p. 125–128.

KUTTRUFF, H.; VORLÄNDER, M. **Room acoustics.** Crc Press, 2024.

LATHI, B. P.; DING, Z. **Modern Digital and Analog Communication.** 5. ed. Cary, NC: Oxford University Press, fev. 2018. (Oxford Series in Electrical and Computer Engineering).

LEE, C.; HORNER, A.; WU, B. The effect of mp3 compression on the timbre space of sustained musical instrument tones. **Journal of the Audio Engineering Society**, Audio Engineering Society, v. 61, n. 11, p. 840–849, 2013.

LEONARD, B. **Listening Through the Room: The influence of acoustic environment on task-oriented critical listeners.** McGill University (Canada), 2015.

LI, G.; ZHENG, C.; LI, X.; YU, T.; BLEECK, S.; SANG, J. Evaluation of headphone phase equalization on sound reproduction. **Applied Acoustics**, Elsevier, v. 156, p. 208–216, 2019.

LOIZOU, P. C. **Speech enhancement: theory and practice.** CRC press, 2007.

LONG, M. **Architectural acoustics.** Elsevier, 2005.

LYONS, R. G. **Understanding digital signal processing.** 3. ed. Philadelphia, PA: Prentice Hall, nov. 2010.

MALLAT, S. **A wavelet tour of signal processing.** Elsevier, 1999.

MARTIN, D.; KING, R.; WOSZCZYK, W.; MASSENBURG, G.; FRANCISCO, M. de. Can We Hear The Difference? Testing the Audibility of Artifacts in High Bit Rate MP3 Audio. *In: AUDIO ENGINEERING SOCIETY. AUDIO Engineering Society Convention 141.* 2016.

MARTIN, D.; MASSENBURG, G. Advanced technical ear training: Development of an innovative set of exercises for audio engineers. *In: AUDIO ENGINEERING SOCIETY. AUDIO Engineering Society Convention 139.* 2015.

MASIERO, B.; FELS, J. Perceptually robust headphone equalization for binaural reproduction. *In: AUDIO ENGINEERING SOCIETY NEW YORK. AUDIO engineering society convention.* 2011. v. 130, p. 1–7.

MASIERO, B. S. **Individualized binaural technology: measurement, equalization and perceptual evaluation.** Logos Verlag Berlin GmbH, 2012. v. 13.

MIDDLEBROOKS, J. C.; MAKOUS, J. C.; GREEN, D. M. Directional sensitivity of sound-pressure levels in the human ear canal. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 86, n. 1, p. 89–108, 1989.

MØLLER, H. Fundamentals of binaural technology. **Applied acoustics**, Elsevier, v. 36, n. 3-4, p. 171–218, 1992.

MØLLER, H.; HAMMERSHØI, D.; JENSEN, C. B.; SØRENSEN, M. F. Evaluation of artificial heads in listening tests. **Journal of the Audio Engineering Society**, Audio Engineering Society, v. 47, n. 3, p. 83–100, 1999.

MØLLER, H.; JENSEN, C. B.; HAMMERSHØI, D.; SØRENSEN, M. F. Binaural synthesis: obtaining the correct sound at the eardrums. *In: BINAURAL synthesis.* 1995.

MØLLER, H.; SØRENSEN, M. F.; JENSEN, C. B.; HAMMERSHØI, D. Binaural technique: Do we need individual recordings? **Journal of the Audio Engineering Society**, Audio Engineering Society, v. 44, n. 6, p. 451–469, 1996.

MONTGOMERY, D. C.; RUNGER, G. C. **Estatística Aplicada e Probabilidade para Engenheiros.** 6. ed.: LTC, 2016.

MOORE, B. C.; GLASBERG, B. R. A revision of Zwicker's loudness model. **Acta Acustica united with Acustica**, European Acoustics Association, v. 82, n. 2, p. 335–345, 1996.

MORISE, M.; YOKOMORI, F.; OZAWA, K. World: a vocoder-based high-quality speech synthesis system for real-time applications. **IEICE TRANSACTIONS on**

Information and Systems, The Institute of Electronics, Information e Communication Engineers, v. 99, n. 7, p. 1877–1884, 2016.

MÜLLER, S.; MASSARANI, P. Transfer-function measurement with sweeps. **Journal of the Audio Engineering Society**, Audio Engineering Society, v. 49, n. 6, p. 443–471, 2001.

MÜLLER, S.; SOARES, Z. M. D. Caracterização de microfones de medição. *In*: SOCIEDADE DE ENGENHARIA DE ÁUDIO (AES BRASIL). ANAIS do 11º Congresso de Engenharia de Áudio – 17ª Convenção Nacional da AES Brasil. São Paulo, SP, maio 2013. p. 1–6. Apresentado no 11º Congresso de Engenharia de Áudio, 17ª Convenção Nacional da AES Brasil, 7 a 9 de maio de 2013.

MUNSON, W.; GARDNER, M. B. Standardizing auditory tests. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 22, 5_Supplement, p. 675–675, 1950.

MURRAY, L. **Sound design theory and practice: Working with sound**. Routledge, 2019.

OPPENHEIM, A. V. **Sinais e Sistemas**. 2. ed. [S.l.]: Pearson Prentice Hall, 2010.

OPPENHEIM, A. V.; SCHAFER, R. W. **Discrete-Time Signal Processing (3rd Edition) (Prentice-Hall Signal Processing Series)**. 3. ed. [S.l.]: Prentice Hall, ago 2009.

PAUL, S. Binaural recording technology: A historical review and possible future developments. **Acta acustica united with Acustica**, European Acoustics Association, v. 95, n. 5, p. 767–788, 2009.

PLINGE, A.; JACOB, F.; HAEB-UMBACH, R.; FINK, G. A. Acoustic microphone geometry calibration: An overview and experimental evaluation of state-of-the-art algorithms. **IEEE Signal Processing Magazine**, IEEE, v. 33, n. 4, p. 14–29, 2016.

PRAS, A.; ZIMMERMAN, R.; LEVITIN, D.; GUASTAVINO, C. Subjective evaluation of mp3 compression for different musical genres. *In*: AUDIO ENGINEERING SOCIETY. AUDIO Engineering Society Convention 127. 2009.

PRAWDA, K.; SCHLECHT, S. J.; VÄLIMÄKI, V. Robust selection of clean swept-sine measurements in non-stationary noise. **The Journal of the Acoustical Society of America**, AIP Publishing, v. 151, n. 3, p. 2117–2126, 2022.

PROAKIS, J.; SALEHI, M. **Digital Communications**. 5. ed. New York, NY: McGraw-Hill Professional, nov. 2007.

RAFAELY, B.; TOURBABIN, V.; HABETS, E.; BEN-HUR, Z.; LEE, H.; GAMPER, H.; ARBEL, L.; BIRNIE, L.; ABHAYAPALA, T.; SAMARASINGHE, P. Spatial audio signal processing for binaural reproduction of recorded acoustic scenes—review and challenges. **Acta Acustica**, EDP Sciences, v. 6, p. 47, 2022.

RAMOS, G.; LOPEZ, J. J. Filter design method for loudspeaker equalization based on IIR parametric filters. **Journal of the Audio Engineering Society**, Audio Engineering Society, v. 54, n. 12, p. 1162–1178, 2006.

RIFE, D. D.; VANDERKOOY, J. Transfer-function measurement with maximum-length sequences. **Journal of the Audio Engineering Society**, Audio Engineering Society, v. 37, n. 6, p. 419–444, 1989.

ROGOWSKA, A. Audibility of lossy compressed musical instrument tones. *In*: AUDIO ENGINEERING SOCIETY. AUDIO Engineering Society Convention 138. 2015.

RUMSEY, F. **Spatial audio**. Routledge, 2012.

SAN MILLÁN-CASTILLO, R.; LATORRE-IGLESIAS, E.; GLESSER, M.; WANTY, S.; JIMÉNEZ-CAMINERO, D.; ÁLVAREZ-JIMENO, J. M. MOSQUITO: an open-source and free toolbox for sound quality metrics in the industry and education. *In*: INSTITUTE OF NOISE CONTROL ENGINEERING, 5. INTER-NOISE and NOISE-CON Congress and Conference Proceedings. 2021. v. 263, p. 1164–1175.

SCHLIEPER, R.; LI, S.; PREIHS, S.; PEISSIG, J. Estimation of the Headphone "Openness" Based on Measurements of Pressure Division Ratio, Headphone Selection Criterion, and Acoustic Impedance. *In*: AUDIO ENGINEERING SOCIETY. AUDIO Engineering Society Convention 145. 2018.

SCHLIEPER, R.; LI, S.; PREIHS, S.; PEISSIG, J. The relationship between the acoustic impedance of headphones and the occlusion effect. *In*: AUDIO ENGINEERING SOCIETY. AUDIO Engineering Society Conference: 2019 AES International Conference on Headphone Technology. 2019.

SHIN, K.; HAMMOND, J. (ed.). **Fundamentals of Signal Processing for Sound and Vibration Engineers**. [S.l.]: John Wiley & Sons, Ltd, 2008.

SILVEIRA, H. A.; VERGARA, E. F. M.; LENZI, A. Modelos psicofísicos aplicados na detecção de sinais impulsivos em refrigeradores domésticos. *In*: XII Congresso/Congresso Iberoamericano de Acústica, XXIX Encontro da Sociedade Brasileira de Acústica. Florianópolis, SC, Brasil, 2022.

SMITH, S. W. **The scientist and engineer's guide to digital signal processing**. California Technical Pub, dez. 1997.

SOULODRE, G. A.; LAVOIE, M. Subjective evaluation of MPEG layer II with spectral band replication. *In: AES 117th Convention*. 2004.

TAYLOR, M.; FORBES, S.; CREELMAN, C. Sequential dependence in signal detection, and parameter estimation by sequential tracking (PEST). **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 42, 5_Supplement, p. 1195–1195, 1967.

THADEN, R. **Auralisation in building acoustics**. 2005. Tese (Doutorado) – Bibliothek der RWTH Aachen.

VÖLK, F. Headphone selection for binaural synthesis with blocked auditory canal recording. *In: 132ND AES Convention*. 2012.

VORLÄNDER, M. **Auralization**. Springer, 2020.

WALKER, G. T. On periodicity in series of related terms. **Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character**, The Royal Society London, v. 131, n. 818, p. 518–532, 1931.

WATKINSON, J. **Art of digital audio**. Routledge, 2013.

WECKESSER, W. Signal processing with SciPy: Linear filters. **Signal processing with SciPy: Linear filters**, 2016.

WELCH, P. D. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. **IEEE Transactions on audio and electroacoustics**, v. 15, n. 2, p. 70–73, 1967.

WETHERILL, G.; LEVITT, H. Sequential estimation of points on a psychometric function. **British Journal of Mathematical and Statistical Psychology**, Wiley Online Library, v. 18, n. 1, p. 1–10, 1965.

WIGHTMAN, F. L.; KISTLER, D. J. Headphone simulation of free-field listening. II: Psychophysical validation. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 85, n. 2, p. 868–878, 1989.

WIKIPEDIA. **Minimum phase** — Wikipedia, The Free Encyclopedia. 2025. <http://en.wikipedia.org/w/index.php?title=Minimum%20phase&oldid=1261612951>. [Online; accessed 21-April-2025].

WOODHOUSE, J.; MANUEL, E.; SMITH, L.; WHEBLE, A.; FRITZ, C. Perceptual thresholds for acoustical guitar models. **Acta Acustica united with Acustica**, European Acoustics Association, v. 98, n. 3, p. 475–486, 2012.

YOU, J.; JEON, J. Y. Just noticeable differences in sound quality metrics for refrigerator noise. **Noise Control Engineering Journal**, Institute of Noise Control Engineering, v. 56, n. 6, p. 414–424, 2008.

ZÖLZER, U. **Dafx**. Edição: Udo Zolzer. 2. ed. Hoboken, NJ: Wiley-Blackwell, mar. 2011.

ZWICKER, E.; FASTL, H. **Psychoacoustics: Facts and models**. Springer Science & Business Media, 2013. v. 22.

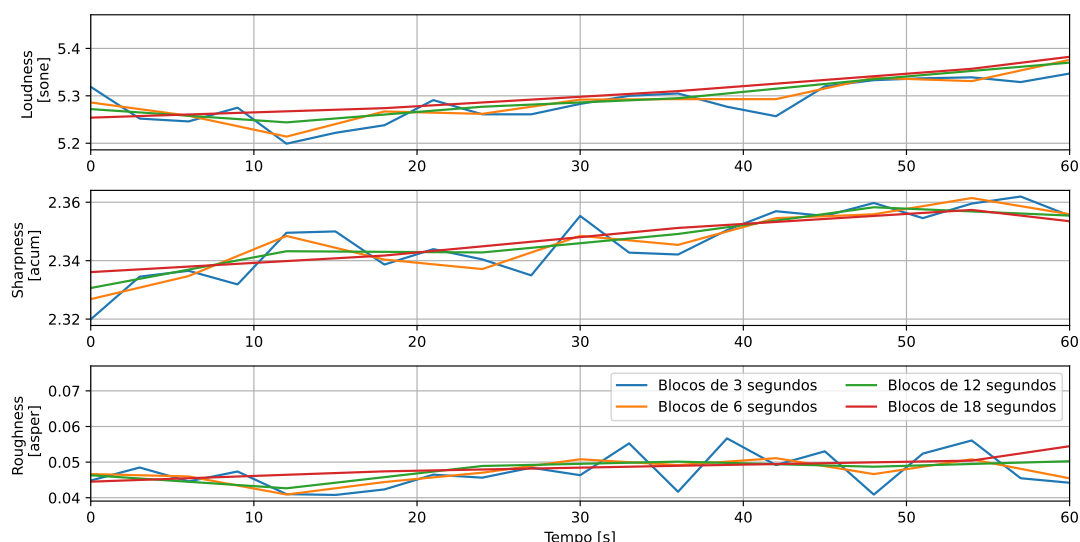
ZWICKER, E.; SCHARF, B. A model of loudness summation. **Psychological review**, American Psychological Association, v. 72, n. 1, p. 3, 1965.

APÊNDICES

APÊNDICE A – ANÁLISE DE CONVERGÊNCIA DAS MÉTRICAS EM FUNÇÃO DO TAMANHO DO BLOCO

Com o objetivo de avaliar a influência do tamanho do bloco na estabilidade das métricas psicoacústicas, foram comparadas janelas de 3, 6, 12 e 18 s aplicadas a um mesmo sinal de referência. A Figura 62 apresenta a evolução temporal das métricas de *Loudness*, *Sharpness* e *Roughness*. Observa-se que, embora os blocos mais curtos apresentem maior variabilidade local, a média e a tendência global convergem rapidamente, indicando que blocos de 3 s são suficientes para representar adequadamente a evolução temporal dessas métricas.

Figura 62 – Análise de convergência das métricas psicoacústicas em função do tamanho do bloco para o sinal de compressor de velocidade única.



APÊNDICE B – DADOS OBTIDOS NA SALA DE REUNIÕES

Tabela 11 – MCD para os sinais de compressor convoluídos com as BRIRs obtidas nas diferentes condições de medição na sala de reuniões.

				Posição 1			Posição 2			Posição 3		
				Genelec	Hinor	Sony	Genelec	Hinor	Sony	Genelec	Hinor	Sony
Sem correção	2400 rpm	Cortex	Direito	4.52	5.11	4.83	4.69	5.90	4.53	4.67	5.95	7.42
			Esquero	5.12	6.52	5.22	5.05	7.15	4.69	4.80	6.44	6.27
		Kemar	Direito	6.59	7.62	6.53	6.27	8.58	7.08	8.30	8.02	8.78
			Esquero	4.67	5.96	5.74	4.97	6.34	4.71	4.21	5.87	4.35
	3000 rpm	Cortex	Direito	4.59	4.87	4.80	4.58	5.79	4.79	4.35	6.12	7.58
			Esquero	4.75	6.27	5.10	4.97	6.42	4.66	4.82	6.47	6.24
		Kemar	Direito	6.41	7.75	6.95	6.13	8.47	6.56	7.55	7.60	9.18
			Esquero	5.08	5.85	5.64	4.78	6.02	4.90	4.20	5.89	4.37
	4500 rpm	Cortex	Direito	4.68	4.99	5.40	4.76	5.46	4.54	4.47	6.45	6.97
			Esquero	5.33	6.46	5.38	4.96	6.87	4.84	4.93	6.36	6.44
		Kemar	Direito	6.28	7.25	7.08	5.54	8.18	6.69	7.52	7.85	8.72
			Esquero	5.24	6.39	6.38	4.98	6.57	5.39	4.37	6.11	4.44
Com correção	2400 rpm	Cortex	Direito	4.36	5.06	5.23	4.57	5.17	4.32	4.51	5.43	6.01
			Esquero	4.72	5.92	4.85	4.67	5.81	4.78	4.88	6.16	5.77
		Kemar	Direito	5.12	5.23	5.97	5.17	5.11	4.23	5.04	5.04	6.02
			Esquero	5.14	5.60	5.59	4.44	4.76	4.90	4.33	4.88	4.21
	3000 rpm	Cortex	Direito	4.62	5.00	5.26	4.35	5.09	4.86	4.12	5.58	6.24
			Esquero	4.86	5.73	4.99	4.72	5.86	4.76	4.74	6.16	5.87
		Kemar	Direito	5.38	5.50	6.48	4.96	4.78	4.21	5.43	5.45	6.30
			Esquero	5.46	6.02	5.54	4.30	4.36	5.22	4.23	4.56	4.55
	4500 rpm	Cortex	Direito	4.75	5.09	5.50	4.49	4.80	4.54	4.46	6.22	6.02
			Esquero	5.20	5.96	5.16	5.11	6.35	4.62	4.87	6.42	5.90
		Kemar	Direito	5.35	5.54	6.21	5.08	5.17	4.72	4.75	5.01	6.16
			Esquero	5.16	5.73	5.52	5.00	4.92	5.20	4.70	5.33	4.50

Figura 63 – FRFs em bandas de 1/6 de oitava para as 3 posições de medição (linhas) e os 3 HATS (colunas), utilizando o monitor de áudio Genelec como fonte sonora. Sinais de referência do HSU III.2.

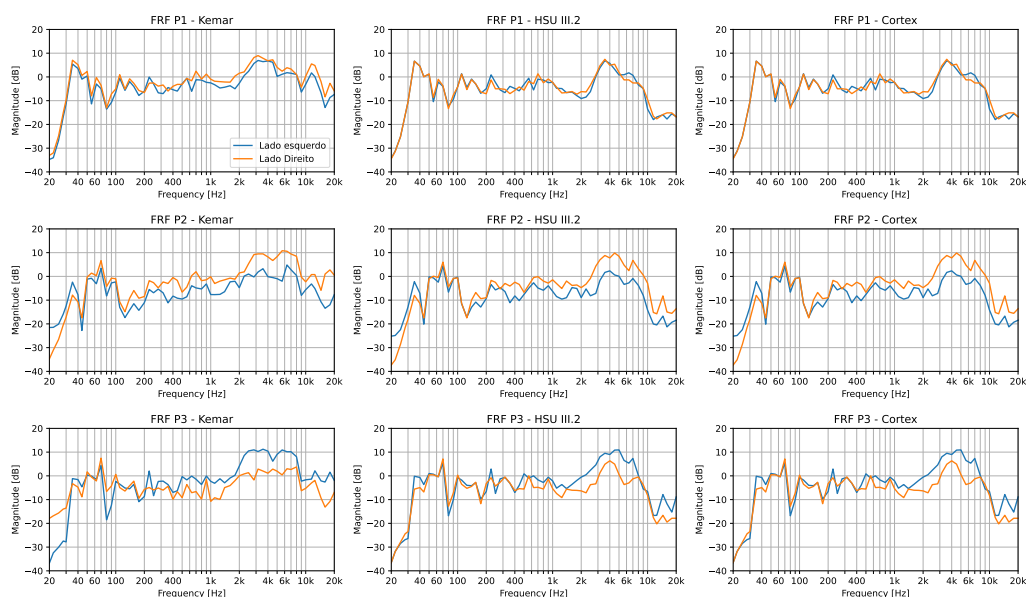


Figura 64 – FRFs em bandas de 1/6 de oitava para as 3 posições de medição (linhas) e os 3 HATS (colunas), utilizando a caixa acústica Hinor como fonte sonora.

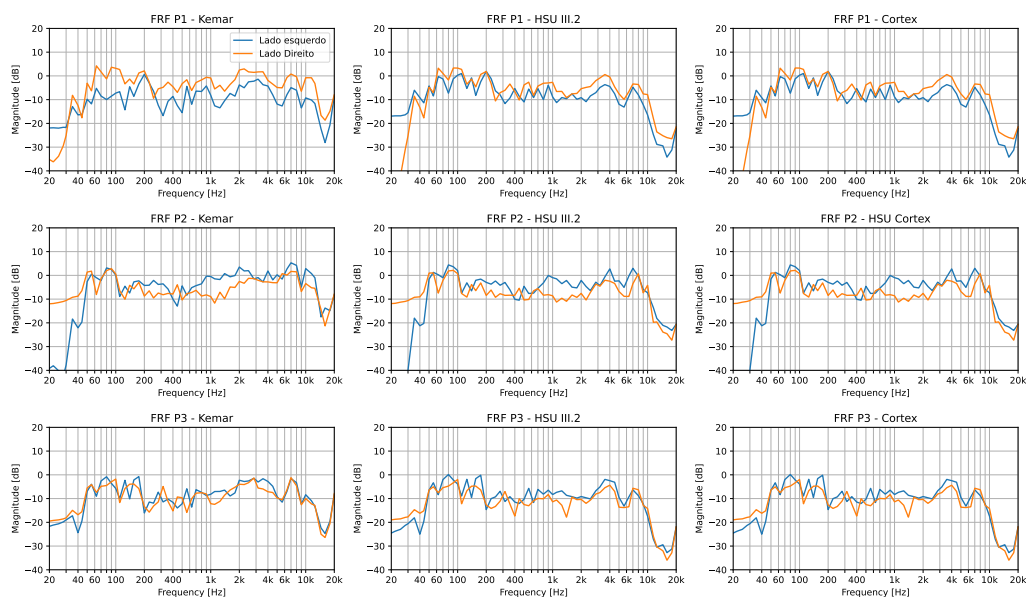
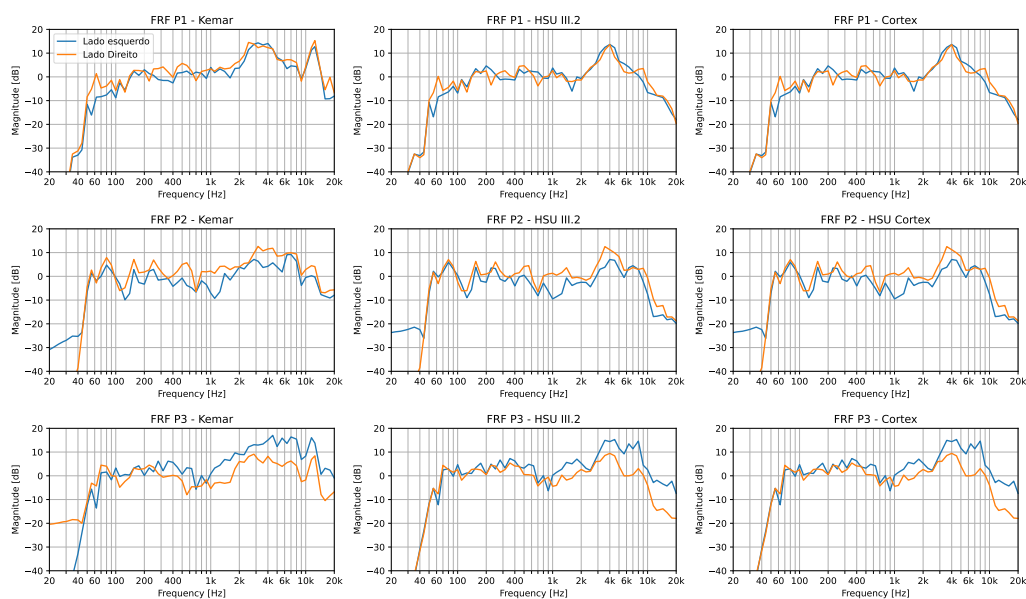
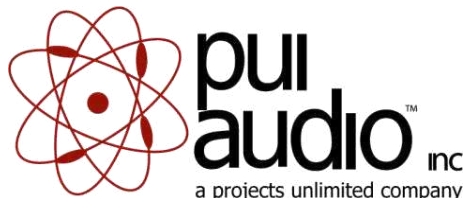


Figura 65 – FRFs em bandas de 1/6 de oitava para as 3 posições de medição (linhas) e os 3 HATS (colunas), utilizando o sistema multimídia Sony como fonte sonora.



ANEXOS

**ANEXO A – FICHA TÉCNICA DO MICROFONE UTILIZADO NA
RESTAURAÇÃO DO HATS CORTEX**



Data Sheet

AOM-5024L-HD-R

PUI Audio's all-new **HD Series** microphones use premium-grade FETs and diaphragms for high sensitivity and superior signal-to-noise ratio. Each microphone features GSM buzz-blocking capacitors. Upgrade the ECM microphone that you use today with a PUI Audio **HD Series** microphone.

The 9.7mm diameter **AOM-5024L-HD-R** is designed for extreme fidelity in even the quietest settings from 20 Hz to 20 kHz.

Features:

- 9.7mm diameter
- 5mm height
- -24 dB sensitivity
- 80 dB signal-to-noise ratio
- True 20 Hz to 20 kHz performance

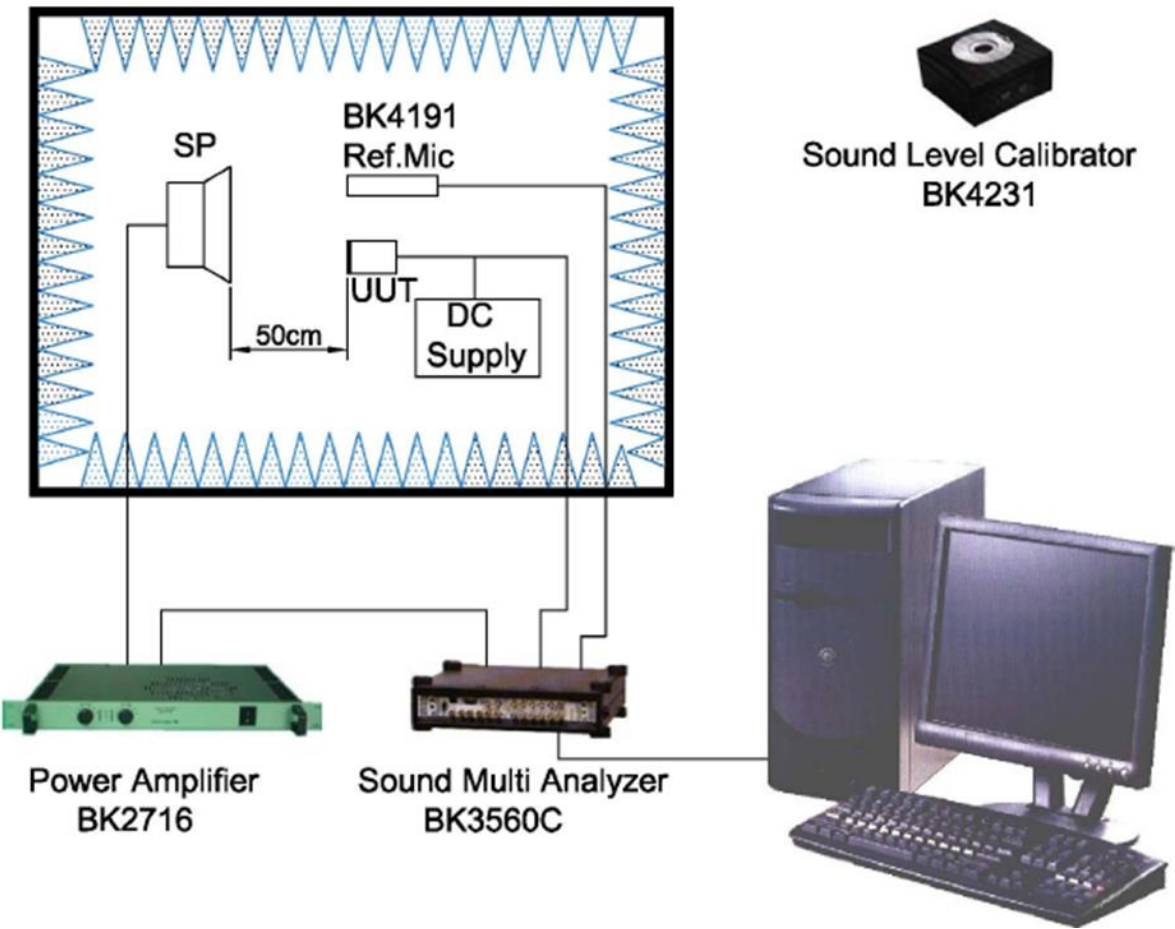
Specifications

Parameters	Values	Units
Sensitivity (1 kHz @ 50cm) 0 dB=1V/Pa	-24 ±3	dB
Rated Voltage	3	VDC
Output Impedance (@ 1 kHz)	2.2	kΩ
Current consumption (3VS with 2.2 kΩ RL)	500	μA
Signal-to-Noise Ratio (1kHz, 94 dB input, A-weighted)	80	dB
Decreasing Voltage (3VS to 2VS)	-3	dB
Frequency Range	20 ~ 20,000	Hz
Operating Voltage Range	1 ~ 10	VDC
Maximum SPL Input (THD<3%)	110	dB
Directivity	Omni-directional	-

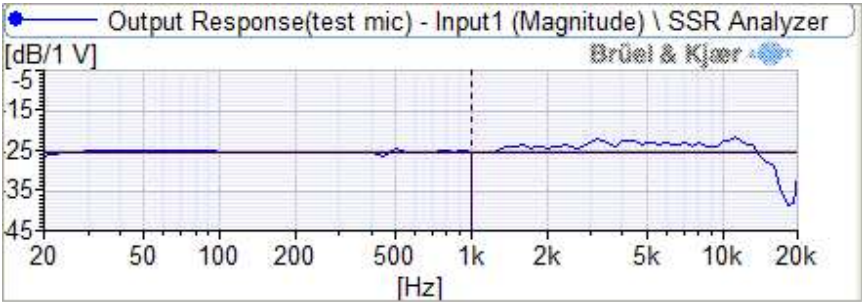
Specifications (continued)

Operating Temperature	-30 ~ +70	°C
Storage Temperature	-40 ~ +85	°C
Weight	<0.3	Grams

Measurement Method (in Anechoic Chamber)



Typical Frequency Response (measured at 50cm with 3V input and 94 dB source)

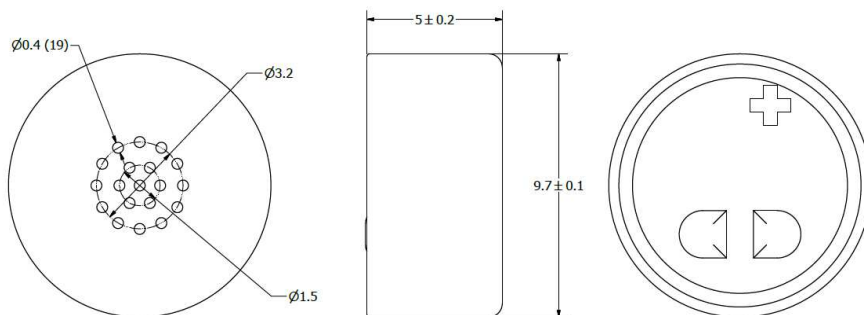


Reliability Testing

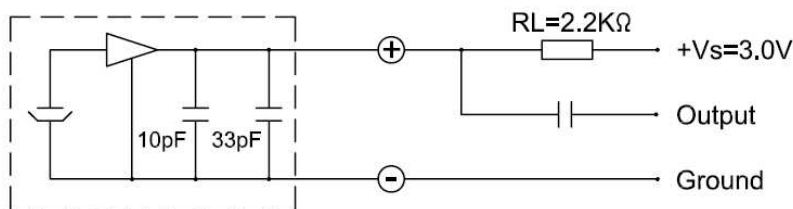
Type of Test	Test Specifications
High Temperature Test	200 hours at $+70^{\circ}\text{C} \pm 3^{\circ}\text{C}$ followed by two hours in normal room temperature
Low Temperature Test	200 hours at $-25^{\circ}\text{C} \pm 3^{\circ}\text{C}$ followed by two hours in normal room temperature
Humidity Test	200 hours at $+40^{\circ}\text{C} \pm 3^{\circ}\text{C}$ with relative humidity at 90% to 95% followed by 2 hours in normal room temperature
Temperature Cycle Testing	30 minutes at -25°C , 10 minutes at 20°C , 30 minutes at $+70^{\circ}\text{C}$, 10 minutes at 20°C for five cycles, followed by 2 hours in normal room temperature
Vibration Test	10 to 55 Hz for 1 minute with 1.52mm distance, followed by a two hour 3 axis test in packaging
Drop Test	Drop microphones in packaging onto concrete floor from 1 meter height in each of 3 axis
ESD Test (according to IEC 6100)	1. Contact discharge - Discharge 6000 VDC from capacitor into microphone output through 330Ω resistor ten times. 2. Air discharge - Discharge 8000 VDC into sound hole of the microphone ten times.

After each test, the speaker's SPL shall be ± 3 dB of the original SPL

Dimensions



Recommended Drive Circuit



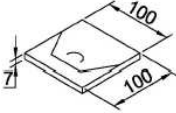
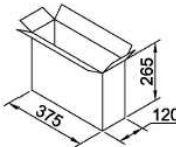
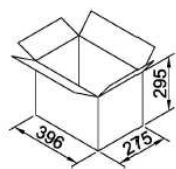
Microphone Handling Precautions

High temperature and/or static electricity may damage microphones. To ensure careful handling, we suggest following these precautions:

- Ensure the power rating of the soldering iron is below 90 watts
- The temperature of the soldering iron must be limited to $360^{\circ}\text{C} \pm 10^{\circ}\text{C}$ ($680^{\circ}\text{F} \pm 50^{\circ}\text{F}$)
- Soldering duration for each terminal shall be at or under 2 seconds
- If practical, use a metal fixture to hold the microphone in-place and to act as a heatsink. A fixture should have appropriate diameter holes drilled through the entire fixture to prevent pressure from being placed on the diaphragm (as below)



Packaging

	Drawing	Qty (pcs.)	Size(mm) L×W×H	Material
Packing		100	100×100×6.5	Paper
Middle Package		10000 (100×100)	375×120×265	Paper
Outer Package		20000 (2×10000)	396×275×295	Paper

This document contains data proprietary to PUI Audio Inc. Any use or reproduction, in any form, without prior written permission of PUI Audio Inc. is prohibited.

©2017, PUI Audio Inc.

Specifications Revisions

Revision	Description	Date
-	Released from Engineering	6/6/2017

Note:

1. Unless otherwise specified:
 - A. All dimensions are in millimeters.
 - B. Default tolerances are $\pm 0.5\text{mm}$ and angles are $\pm 3^\circ$.
2. Specifications subject to change or withdrawal without notice.
3. This part is RoHS 2011/65/EU Compliant.

PUI Audio, Inc. A Projects Unlimited Company, 3541 Stop Eight Road, Dayton, OH 45414
Tel: (937) 415-5901 Fax: (937) 415-5925