

TESE DE DOUTORADO Nº 299

**PREDIÇÃO DE INTENSIDADE SONORA PERCEBIDA (LOUDNESS) PARA
ÁUDIO ESPACIAL**

Leandro da Silva Pires

DATA DA DEFESA: 27/06/2019

Predição de Intensidade Sonora Percebida (*Loudness*)
para Áudio Espacial

Leandro da Silva Pires

Belo Horizonte
2019

SERVIÇO DE PÓS-GRADUAÇÃO DA EE-UFMG

Data de Depósito:

Assinatura: _____

Predição de Intensidade Sonora Percebida (*Loudness*) para Áudio Espacial

Leandro da Silva Pires

***Orientador:* Prof. Dr. Maurílio Nunes Vieira**

***Coorientador:* Prof. Dr. Hani Camille Yehia**

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia da Universidade Federal de Minas Gerais, como requisito para obtenção do título de Doutor em Engenharia Elétrica – Sistemas de Computação e Telecomunicações.

UFMG – Belo Horizonte

Julho de 2019

P667p

Pires, Leandro da Silva.

Predição de intensidade sonora percebida (*loudness*) para áudio espacial [recurso eletrônico] / Leandro da Silva Pires. - 2019.

1 recurso online (342 f. : il., color.) : pdf.

Orientador: Maurílio Nunes Vieira.

Coorientador: Hani Camille Yehia.

Tese (doutorado) - Universidade Federal de Minas Gerais, Escola de Engenharia.

Anexos e apêndices: f. 338-342.

Bibliografia: f. 308-333.

Exigências do sistema: Adobe Acrobat Reader.

1. Engenharia elétrica - Teses. 2. Processamento de sinais - Teses. 3. Radiodifusão - Teses. 4. Telecomunicações - Teses. I. Vieira, Maurílio Nunes. II. Yehia, Hani Camille. III. Universidade Federal de Minas Gerais. Escola de Engenharia. IV. Título.

CDU: 621.3(043)

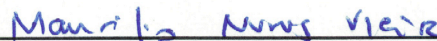
**"Predição de Intensidade Sonora Percebida (loudness) Para
Áudio Espacial"**

Leandro da Silva Pires

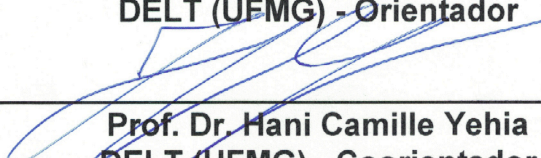
Tese de Doutorado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia da Universidade Federal de Minas Gerais, como requisito para obtenção do grau de Doutor em Engenharia Elétrica.

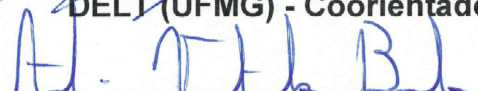
Aprovada em 27 de junho de 2019.

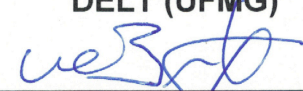
Por:

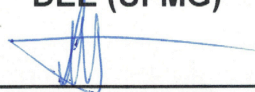


Prof. Dr. Maurílio Nunes Vieira
DELT (UFMG) - Orientador

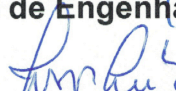

Prof. Dr. Hani Camille Yehia
DELT (UFMG) - Coorientador


Prof. Dr. Adriano Vilela Barbosa
DELT (UFMG)


Prof. Dr. Wallace do Couto Boaventura
DEE (UFMG)


Prof. Dr. Luiz Wagner Pereira Biscainho
COPPE (UFRJ)


Prof. Dr. Alexander Mattioli Pasqual
Divisão de Engenharia Mecânica (ITA)


Prof. Dra. Flávia Magalhães Freitas Ferreira
Programa de Pós-graduação em Engenharia Elétrica (PUC Minas)

AGRADECIMENTOS

À Agência Nacional de Telecomunicações (Anatel), pela oportunidade de trabalho no tema e pelo suporte financeiro ao longo da pesquisa.

A todos os participantes dos testes de escuta pelo suporte dado ao meu trabalho e pela ajuda na obtenção de resultados de melhor qualidade. A observação de efeitos relevantes não seria possível sem suas participações diligentes e seus *feedbacks* das sessões experimentais.

Aos meus orientadores na Universidade Federal de Minas Gerais, Dr. Maurílio Vieira e Dr. Hani Yehia pelo apoio continuado, paciência e incentivo ao longo desta jornada.

Aos membros das bancas examinadoras de qualificação, pelos comentários perspicazes, e de defesa, por aceitarem o convite à apreciação do meu trabalho.

Aos meus supervisores na Universidade de Surrey, Dr. Tim Brookes e Dr. Russell Mason, pela oportunidade de estágio doutoral e pelo acesso ao laboratório, equipamento e instalações de pesquisa. Sem o precioso suporte deles, a condução desta pesquisa não seria possível.

Aos colegas do Centro de Estudos da Fala, Acústica, Linguagem e música (CEFALA) e do *Institute of Sound Recording* (IoSR) pelas trocas, pelas discussões estimulantes e pela diversão.

Aos meus pais e irmão por serem minhas referências de vida.

À minha esposa Luciana pelo apoio incondicional e por não aceitar menos que a excelência.

À Bast pelo companheirismo noturno, cujas patas por sobre o teclado sempre me avisaram da hora de parar.

“A pretensa evidência do sentir não está fundada em um testemunho da consciência, mas no pré-juízo do mundo.”
(Merleau-Ponty (1999, p. 25))

RESUMO

PIRES, L. **Predição de Intensidade Sonora Percebida (*Loudness*) para Áudio Espacial**. 2019. 366 f. Tese (Doutorado em Engenharia Elétrica – Sistemas de Computação e Telecomunicações) – Escola de Engenharia (EE/UFMG), Belo Horizonte – MG.

O controle da intensidade percebida de áudio (*loudness*) na radiodifusão é prática comum e legalmente exigida desde a publicação da Recomendação ITU-R BS.1770, da União Internacional de Telecomunicações (ITU), para medição objetiva de *loudness* em áudio multicanal. Recomendações e regulamentos regionais foram publicados com base no algoritmo ITU-R, inclusive no Brasil. Isto posto, há oportunidades tanto de melhoria da regulamentação nacional à luz das contribuições mais recentes na área, quanto de aprimoramento do modelo ITU-R para medidas em sistemas avançados de áudio espacial. Este trabalho persegue estes dois objetivos ao testar os parâmetros da norma nacional com um controlador de intensidade percebida em tempo real usando descritores de *loudness* voltados para conteúdo de curta duração, além de procurar contribuir com as discussões sobre o tema no âmbito do ITU-R com o desenvolvimento de um modelo de medição objetiva adaptado aos novos formatos de áudio espacial. Este teve um desempenho satisfatório em comparação com outros modelos, embora fosse puramente uma solução de processamento de sinais e suas leituras não se assemelhassem tanto aos resultados subjetivos. Buscando benefícios potenciais de um modelo mais orientado à percepção, realizou-se testes de escuta para avaliação dos parâmetros posicionais de distância, azimute e elevação, cujos resultados serviram de base para a obtenção de curvas de correção de ganho e nova ponderação direcional para o modelo ITU-R. Os resultados gerais apontam para avanços tanto na frente regulatória quanto na de padronização, seja pela elaboração de uma estratégia de melhorias propostas para a norma brasileira de intensidade percebida, seja pela comparação deste novo algoritmo de predição com a fortuna crítica de modelos de *loudness* por meio de medições realizadas em conteúdo para sistemas de reprodução de áudio espacial multicanal. O modelo desenvolvido obteve a melhor relação de compromisso entre erros de predição (RMSE*), correlação das estimações com os resultados dos testes subjetivos, e tempo médio de execução.

Palavras-chave: Loudness, Radiodifusão, Auralização, Áudio Espacial, Testes Subjetivos, Processamento de Sinais.

ABSTRACT

PIRES, L. **Predição de Intensidade Sonora Percebida (*Loudness*) para Áudio Espacial**. 2019. 366 f. Tese (Doutorado em Engenharia Elétrica – Sistemas de Computação e Telecomunicações) – Escola de Engenharia (EE/UFMG), Belo Horizonte – MG.

Loudness control for broadcasting is a common and legally required practice since the International Telecommunication Union (ITU) Recommendation ITU-R BS.1770 for objective measurements in multichannel audio. Recommendations and regulations based on the ITU-R algorithm have been published worldwide, including Brazil. There is scope for improving national regulations in light of recent contributions to the field, and also for adapting the ITU-R model to measurements in advanced audio systems. This work pursues these two goals by testing the parameters of the Brazilian standard with a real-time loudness controller using short-form descriptors and by developing a new objective measurement model adapted to the new spatial audio formats. The proposed method performed well compared to other loudness models, although it was purely signal processing based and its readings were not very close to subject responses. The potential benefits of a more perceptually motivated model led to a PhD placement in the Institute of Sound Recording at the University of Surrey (UK), where listening tests were conducted to assess positional parameters of distance, azimuth and elevation, whose results served as a basis for deriving gain correction curves and a new directional weighting for the ITU-R model. General results point to advancements in the regulatory and standardization fronts, either by the elaboration of a strategy to improve the Brazilian standard of loudness, or by comparing this new prediction method with the critical fortune of loudness models through measurements on audio content for multichannel reproduction systems. The developed model resulted in the best trade-off between prediction errors (RMSE*), correlation between predictions and subject responses, and mean run time.

Key-words: Loudness, Broadcasting, Auralization, Spatial Audio, Listening Tests, Signal Processing.

LISTA DE ILUSTRAÇÕES

Figura 2.1 – Estrutura do Ouvido Humano (externo, médio e interno) . . .	42
Figura 2.2 – Posições de máxima ressonância da membrana basilar para tons puros.	44
Figura 2.3 – Visão esquemática do processamento de sinais no sistema auditivo.	45
Figura 2.4 – Contornos de mesmo <i>loudness</i> para sons biauriculares de direção frontal.	51
Figura 2.5 – As primeiras curvas de ponderação em frequência utilizadas em modelos de <i>loudness</i> de faixa única foram denominadas A, B e C.	52
Figura 2.6 – Equivalência das escalas sone e phon para um tom senoidal de 1 kHz.	53
Figura 2.7 – Conforme dois tons se aproximam em frequência, cresce a sobreposição de suas áreas de vibração na membrana basilar.	55
Figura 2.8 – Distribuição do nível de pressão sonora pelas bandas críticas na escala Bark para um ruído rosa de 40 dB/Hz @ 1 kHz.	57
Figura 2.9 – Determinação do formato de um filtro auditivo pela abordagem do ruído dentado em Patterson (1976).	58
Figura 2.10–Comparação entre as bandas críticas de Zwicker (1961) e as ERBs de Glasberg e Moore (1990).	58
Figura 2.11–Loudness de um tom persistente de 2 kHz com 57 dBSPL como uma função da duração.	62
Figura 2.12–Mascaramento temporal de <i>loudness</i> de um tom de 2 kHz, 60 dBSPL e 5 ms, ocorrido antes de um ruído de excitação uniforme (UEN) com uma diferença de tempo Δt	63
Figura 2.13–Funções teóricas de <i>loudness</i> mono e biauricular em somatórios biauriculares para potências $n = 0,6$ e $n = 0,35$	66

Figura 2.14–Diagrama polar de captação de uma resposta ao impulso de uma sala de estar.	69
Figura 2.15–(a) Atenuação a_D , necessária para produção das curvas de mesmo <i>loudness</i> de um tom puro em campo livre para o mesmo tom em campo difuso, em função da frequência do tom. (b) Contornos de mesmo <i>loudness</i> de 30, 60 e 90 phon em campo livre e em campo difuso.	70
Figura 2.16–(a) Respostas em frequência de filtros de sombreamento de cabeça . (b) Retardos de grupo associados às diferenças de tempo interaural.	73
Figura 2.17–(a) Respostas ao impulso relativas à cabeça (HRIRs) nos ouvidos direito (em vermelho) e esquerdo (em azul) para uma fonte sonora situada 30 graus à direita do ouvinte. (b) Funções de transferência relativas à cabeça (HRTFs) no ouvido esquerdo para 5 fontes posicionadas como um sistema <i>surround</i>	75
Figura 3.1 – Diagrama em blocos do modelo de detecção de energia	79
Figura 3.2 – Curva ROC para detecção da energia de um tom de 1 kHz com duração de 100 ms e com uma razão sinal-ruído de –10 dB.	82
Figura 3.3 – Contornos de mesmo índice de <i>loudness</i> do modelo multi-faixa de Stevens (1961).	86
Figura 3.4 – Atenuação correspondente ao fator de transmissão necessário em condição de campo livre (a_0) ou para campo difuso (a_{0D}).	90
Figura 3.5 – (a) Função de transferência do nível sonoro de campo livre para o nível sonoro timpânico. (b) Curva de atenuação do ouvido médio.	91
Figura 3.6 – Padrão de excitação de um tom de 1 kHz calculado pelas saídas dos filtros auditivos em função de suas frequências centrais.	92
Figura 3.7 – (a) Circuito RC simulando a duração do decaimento do <i>loudness</i> específico (b) Processamento de <i>loudness</i> de rajadas de 5 kHz com durações de 10 e 100 ms.	98

Figura 3.8 – (a) Saída do modelo de Glasberg e Moore (2002) em resposta a um tom de 4 kHz com duração de 200 ms, ilustrando as curvas de <i>loudness</i> instantâneo, de curta e longa duração. (b) Nível de <i>loudness</i> de curta duração em função da duração de tons de 1 e 4 kHz.	100
Figura 3.9 – Curvas A, B, C e D de ponderação em frequência constantes da norma IEC:60651.	104
Figura 3.10–Curva ITU-R BS.468 de ponderação para medição de intensidade de ruído de equipamentos eletrônicos.	105
Figura 3.11–Curva B revisada em baixa frequência (<i>RLB</i>) de ponderação para conteúdo de radiodifusão proposta por Soulodre (2004) .	106
Figura 3.12–(a) Resposta em frequência do filtro de compensação dos efeitos acústicos da cabeça como uma média das respostas ao longo dos ângulos de incidência mais comuns num ambiente de escuta multicanal (b) Curva de ponderação K utilizada no modelo de <i>loudness</i> ITU-R BS.1770, resultado da combinação das respostas em frequência do pré-filtro e do filtro <i>RLB</i>	108
Figura 3.13–Diagrama em blocos do algoritmo de <i>loudness</i> multicanal ITU-R BS.1770	109
Figura 3.14–Diagrama em blocos do medidor de pico verdadeiro conforme especificações no Anexo 2 da Rec. ITU-R BS.1770-3	112
Figura 3.15–Distribuição de <i>Loudness</i> com limiares de fechamento e de Faixa de <i>Loudness</i> para o filme “Matrix” em masterização para DVD	117
Figura 3.16–Medidas consolidadas de <i>loudness</i> conforme Rec. ITU-R BS.1770-4 para um segmento de teledramaturgia com faixa dinâmica larga.	118
Figura 3.17–Diagrama em blocos do medidor de pico verdadeiro conforme especificações no Anexo 2 da Rec. ITU-R BS.1770-3	119
Figura 4.1 – Modelos de definição de áudio imersivo: a) Áudio baseado em canais, b) Áudio baseado em objetos e c) Áudio baseado em cenas	131
Figura 4.2 – Diagrama em blocos do detetor de conteúdo de formato curto	140

Figura 4.3 – Diagrama em blocos do controlador de <i>loudness</i>	147
Figura 4.4 – Treinamento e sintonia do detetor de conteúdo de formato curto.	151
Figura 4.5 – Integrações de <i>loudness</i> ao longo de um trecho de um programa de televisão que intercala passagens de diálogo com números musicais altamente comprimidos em dinâmica	152
Figura 4.6 – Captura de tela do controlador automático de <i>loudness</i> atuando no mesmo conteúdo de áudio da Figura 4.5 com duração de 160 segundos.	153
Figura 4.7 – (a) Construção de uma fonte imagem. (b) Fontes imagem construídas para uma sala em duas dimensões.	157
Figura 4.8 – Diagrama em blocos do modelo de <i>loudness</i> proposto com o objetivo de se contabilizar os efeitos acústicos da sala por meio de sua resposta ao impulso e preservar o agnosticismo do modelo de áudio baseado em cenas no que se refere ao número e à disposição de alto-falantes no sistema de reprodução do consumidor final.	160
Figura 4.9 – Ajustes de nível em relação ao conteúdo de referência (a) por tipo de conteúdo (b) por participante.	162
Figura 4.10–Resposta em frequência e diagrama de captação da sala virtual nº 1	163
Figura 4.11–Resposta em frequência e diagrama de captação da sala virtual nº 2	164
Figura 4.12–Resposta em frequência e diagrama de captação da sala virtual nº 3	165
Figura 4.13–Diferenças médias entre predições para os demais sistemas de áudio em relação à referência de cinco canais. Os segmentos escuros nos gráficos do tipo pizza representam as predições que caíram dentro dos intervalos de confiança dos testes subjetivos.	167
Figura 4.14–Diferenças médias entre predições em relação à referência de cinco canais discriminadas por sistema de reprodução.	169
Figura 5.1 – Laboratório de áudio no formato de sala de audição conforme com a Rec.ITU-R BS.1116 (ITU-R, 2015a)	171

Figura 5.2 – Configuração de medidas de THD	187
Figura 5.3 – Calibração de um tom gravado a 84 dBSPL para um sinal de entrada de –18 dBFS.	188
Figura 5.4 – Distorção harmônica total de sinais acústicos capturados a dois metros de distância da fonte sonora.	189
Figura 5.5 – Arranjo de alto-falantes na perspectiva do ouvinte: (a) ilustração crua da perspectiva do participante (b) representação 2D por pontos centrados nos eixos acústicos dos alto-falantes. . .	191
Figura 5.6 – Ângulos e posições para uma distância fixa entre eixos acústicos de dois alto-falantes Genelec 8020.	192
Figura 5.7 – Medidas de <i>loudness</i> BS.1770 feitas em sinais de teste (tom senoidal e ruído rosa) em convolução com as Respostas Bi-auriculares da Sala ao Impulso (BRIRs) da sala de escuta crítica.	194
Figura 5.8 – Gráficos ampliados no zero de variação angular para o ruído rosa: medidas horizontais próximas a um ângulo de cabeça de 0°.	196
Figura 5.9 – <i>Setup</i> de medidas de loudness do arranjo de alto-falantes em linha.	197
Figura 5.10–Calibração do arranjo experimental	198
Figura 5.11–Interface gráfica do experimento.	208
Figura 5.12–Médias e intervalos de confiança entre usuários por pares de distâncias de teste/referência: ajustes dos participantes em relação ao nível de referência de 70 dBSPL (0 dB).	210
Figura 5.13–Médias e intervalos de confiança agrupados por distância do alto-falante de referência ao ouvinte, referentes aos ajustes feitos pelos participantes do experimento nos níveis de reprodução dos alto-falantes de teste, de tal forma que a sensação de <i>loudness</i> resultante fosse casada com a sensação de <i>loudness</i> produzida por cada alto-falante numa dada distância de referência.	211
Figura 5.14–Análise exploratória dos ajustes de nível realizados pelos participantes do experimento.	212

Figura 5.15–Diagrama em blocos da modificação proposta no algoritmo BS.1770 (adaptado de ITU-R (ITU-R, 2015b)). Blocos de correção de ganho em função da distância foram inseridos na cadeia de processamento multicanal.	215
Figura 5.16–Médias e intervalos de confiança agrupados por distância do alto-falante de referência ao ouvinte, referentes aos ajustes feitos pelos participantes do experimento nos níveis de reprodução dos alto-falantes de teste, de tal forma que a sensação de <i>loudness</i> resultante fosse casada com a sensação de <i>loudness</i> produzida por cada alto-falante numa dada distância de referência.	216
Figura 5.17–Comparações nível a nível dos fatores “alto-falante de teste” e “sala de reprodução”.	218
Figura 5.18–Respostas ao impulso e tempos de reverberação por oitava na sala de escuta crítica BS.1116.	220
Figura 5.19–Respostas ao impulso e tempos de reverberação por oitava na sala de aula comum.	221
Figura 5.20–Sala A: Planta e localização do HATS.	226
Figura 5.21–Sala B: Planta e localização do HATS.	227
Figura 5.22–Sala C: Planta e localização do HATS.	228
Figura 5.23–Sala D: Planta e localização do HATS.	228
Figura 5.24–Sala de escuta crítica em conformidade com a Recomendação BS.1116 do ITU-R (2015a).	229
Figura 5.25–Calibração de fones de ouvido em estúdio com um Simulador de Cabeça e Torso (HATS).	230
Figura 5.26–Resultados do teste piloto da configuração experimental.	234
Figura 5.27– <i>Patch</i> de MaxMSP [®] para o teste de escuta.	235
Figura 5.28–Médias e intervalos de confiança de 95% dos ajustes de nível executados pelos participantes.	236
Figura 5.29–Médias e intervalos de confiança dos participantes agrupados por sala de reprodução.	237
Figura 5.30–Gráfico em linha das respostas dos participantes por todos os níveis dos fatores experimentais “sala de reprodução” e “azimute”.	238

Figura 5.31–Histogramas e gráficos Q-Q para avaliação de normalidade. Os dados estão agrupados por salas sintetizadas.	239
Figura 5.32–Análise exploratória dos dados obtidos experimentalmente.	240
Figura 5.33–As médias dos participantes como função dos tempos de reverberação e dos azimutes foram ajustadas a uma superfície <i>spline</i> cúbica.	244
Figura 5.34–Diferenças entre as medidas de <i>loudness</i> do modelo ajustado e o algoritmo BS.1770, sobrepostas à curva de correção de ganho e às médias de sensibilidade dos participantes.	245
Figura 5.35–Diferenças entre as medidas de <i>loudness</i> do modelo ajustado e o algoritmo BS.1770, sobrepostas à curva de correção de ga- nho e às médias de sensibilidade dos participantes, agrupadas pelos azimutes testados.	246
Figura 5.36–Planos superior, horizontal e inferior do sistema de reprodu- ção “H” de 22.2 canais, a ser usado neste experimento.	253
Figura 5.37–Fotografia panorâmica da sala de escuta crítica capturando os alto-falantes dos planos superior, horizontal e inferior.	253
Figura 5.38–Fotografia da tela do <i>software</i> de operação do <i>Genelec Louds- peaker Management (GLM)</i>	254
Figura 5.39–Ajustes de nível de áudio digital de fontes reais e fantasmas divididas entre a sala de escuta crítica real e sua versão virtual via auralização.	261
Figura 5.40–Matriz de correlação da variável de resposta com as métri- cas espaciais. Os coeficientes de correlação em vermelho indicam correlações significativamente diferentes de zero.	262
Figura 5.41–Gráficos cartesianos das diferenças de somatório biauricular em relação à incidência frontal com ganhos biauriculares de 3 dB and 6 dB.	266
Figura 5.42–Avaliação do desempenho dos participantes.	269
Figura 5.43–Diferenças nas médias dos participantes entre fontes reais e fantasmas. Anotações com setas indicam o par de azimutes dos alto-falantes que geraram esta ou aquela fonte imagem em particular, de mesma elevação.	270

Figura 5.44–Diagramas de caixa de Sensibilidades Direcionais de <i>Loudness</i> (DLS) por posição da fonte sonora, divididos em grupos de fontes reais e fantasmas (casos (i) e (ii)).	271
Figura 5.45–Histogramas e gráficos Q-Q dos dados dos participantes, divididos em grupos de fontes reais e fantasmas (casos (i) e (ii)).	272
Figura 5.46–Predições vs. resíduos de um modelo linear ao qual os dados foram ajustados.	273
Figura 5.47–Correlações com o somatório biauricular de Robinson com ganhos de 3 dB e 6 dB, com a impressão espacial e com o modelo de inibição biauricular de Moore.	282
Figura 5.48–Gráfico de efeitos da regressão linear.	285
Figura 5.49–Gráfico de resposta do modelo. Cada “degrau” de predição corresponde a um conjunto total de participantes e repetições para uma única direção no conjunto de dados de treinamento.	286
Figura 5.50–Diferenças entre medidas objetivas plotadas contra as respostas dos participantes.	288
Figura 5.51–Arcabouço dos sistemas futuros de radiodifusão	290
Figura 5.52–Diagrama em blocos do modelo de <i>loudness</i> proposto: uma versão modificada do modelo ITU-R BS.1770 com correções de ganho como funções da distância e da reverberação, além de nova ponderação direcional levando em conta o efeito da elevação.	293
Figura 5.53–Diferenças entre as médias dos resultados dos modelos para o sistema de reprodução de referência (5.1) e outros métodos de reprodução.	296
Figura 5.54–Diferenças entre as médias dos resultados dos modelos para o sistema de reprodução de referência (5.1) e outros métodos de reprodução, agrupados por método de reprodução.	299

LISTA DE CÓDIGOS-FONTE

Código-fonte 1 – Código fonte cedido pela TC Electronic para o cálculo da LRA (SKOVENBORG, 2012a)	115
---	-----

LISTA DE TABELAS

Tabela 2.1 – Levantamento de estudos sobre o efeito do nível <i>loudness</i> na Duração Crítica (DC).	63
Tabela 3.1 – Valor do fator <i>F</i> conforme larguras utilizadas no banco de filtros do modelo de Stevens para sons estacionários (ISO 532-A).	85
Tabela 3.2 – Ilustração dos principais estágios dos modelos multifaixa de Zwicker <i>et al.</i> (1991) e Moore, Glasberg e Baer (1997) para sons estacionários.	89
Tabela 4.1 – Pesos direcionais inicialmente propostos em (ITU-R, 2014d)	128
Tabela 4.2 – Pesos direcionais incluídos na quarta versão de (ITU-R, 2015b)	129
Tabela 5.1 – Exigências de <i>loudness</i> para programas de formato longo no Reino Unido	174
Tabela 5.2 – Exigências de <i>loudness</i> para programas de formato curto no Reino Unido	175
Tabela 5.3 – Métricas de comparação entre cálculos de <i>loudness</i> e valores de testes perceptivos considerando inclusão do canal LFE. . .	177
Tabela 5.4 – Diferenças absolutas entre medidas de <i>loudness</i> feitas pela Recomendação BS.1770 do ITU-R (2015b) e suas variantes com filtros passa-baixas na saída da curva <i>K</i>	178
Tabela 5.5 – Níveis de pressão sonora e medidas de distorção harmônica total.	188
Tabela 5.6 – Medidas de nível de pressão sonora do tom de teste de 997 Hz (dBSPL)	197
Tabela 5.7 – Estatísticas de qualidade de ajuste dos dados de Razões entre Energia Direta e Energia Reverberante (DRRs) ao modelo de intensidade proporcional ao inverso do quadrado da distância.	222
Tabela 5.8 – Room acoustical properties	225

Tabela 5.9 – Ganhos de somatório biauricular de <i>loudness</i> calculados em contribuição da NHK para o grupo de trabalho de <i>loudness</i> do ITU-R (2014c) e os pesos direcionais propostos por Komori <i>et al.</i> (2015).	249
Tabela 5.10–Ponderação de canais dependente da posição na versão de 2015 do modelo de <i>loudness</i> ITU-R	249
Tabela 5.11–Azimutes e elevações dos alto-falantes que integram o sistema de reprodução 22.2 da sala de escuta crítica.	252
Tabela 5.12–Fontes físicas (alto-falantes) discriminadas por azimute e elevação (θ, ϕ).	258
Tabela 5.13–Fontes fantasmas (trio de alto-falantes adjacentes) com direções discriminadas por azimute e elevação (θ, ϕ).	258
Tabela 5.14–Ganhos de somatório biauricular de <i>loudness</i> calculados e níveis apresentados na contribuição da NHK para o grupo relator de <i>loudness</i> do ITU-R (2014c).	263
Tabela 5.15–Pares de alto-falantes para os casos de balanceamento VBAP 1 e 2, identificados por suas posições (azimute e elevação em graus).	267
Tabela 5.16–Estatísticas de ajuste do Modelo Linear Generalizado de Efeitos Mistos.	275
Tabela 5.17–Ganhos direcionais estimados por solução de um problema de minimização restrita.	279
Tabela 5.18–Ganhos biauriculares estimados por participante.	279
Tabela 5.19–Desempenho de modelos de regressão treinados. valores de RMSE inferiores a 1,79 estão em negrito.	284
Tabela 5.20–Ganhos direcionais estimados pela solução de um problema de regressão.	287
Tabela 5.21–Sub-elementos posicionais do elemento <i>audioBlockFormat</i> para OBA.	291
Tabela 5.22–Estatísticas dos modelos de <i>loudness</i> ordenadas por RMSE*.	298

LISTA DE ABREVIATURAS E SIGLAS

- 2AFC *Two-Alternative Forced Choice* (Escolha Forçada de Duas Alternativas)
- AASI Aparelhos de Amplificação Sonora Individual
- ABC *Australian Broadcasting Corporation* (Empresa de Radiodifusão da Austrália)
- ADM *Audio Definition Model* (Modelo de Definição de Áudio)
- AIC *Akaike Information Criterion* (Critério de Informações de Akaike)
- AM Modulação em Amplitude
- Anatel Agência Nacional de Telecomunicações
- ANOVA . . *Analysis of Variance* (Análise de Variâncias)
- ANSI *American National Standard Institute* (Instituto Nacional Americano de Padrões)
- BASIC . . . *Beginner's All-purpose Symbolic Instruction Code* (Código de Instruções Simbólicas de Uso Geral para Principiantes)
- BBC *British Broadcasting Corporation* (Empresa Britânica de Radiodifusão)
- BCAP *UK Code of Broadcast Advertising* (Código de Propaganda em Radiodifusão do Reino Unido)
- BIC *Bayesian Information Criterion* (Critério de Informações de Bayesiano)
- BRIR *Binaural Room Impulse Response* (Resposta Biauricular da Sala ao Impulso)
- CBS *Columbia Broadcasting System* (Sistema Columbia de Radiodifusão)
- CRC *Communications Research Centre Canada / Centre de Recherches sur les Communications Canada* (Centro de Pesquisas em Comunicações do Canadá)

dBFS	Decibels referentes ao fundo de escala digital
dBSPL	...	Decibels referentes ao nível de pressão sonora de 20 microPascal (20 μ Pa)
dBTP	Pico verdadeiro em decibels relativos a 100% da escala digital
DIC	<i>Deviance Information Criterion</i> (Critério de Informações de Desvio)
DIN	<i>Deutsches Institut für Normung</i> (Instituto Alemão para Padronização)
DL	<i>Differenz Limen</i> (Limiares de Diferença)
DLS	<i>Directional Loudness Sensitivities</i> (Sensibilidades Direcionais de Loudness)
DPP	<i>Digital Production Partnership</i> (Parceria de Produção Digital)
DRR	<i>Direct-to-Reverberant Energy Ratio</i> (Razão entre Energia Direta e Energia Reverberante)
DSP	<i>Digital Signal Processor</i> (Processador Digital de Sinais)
EBU	<i>European Broadcasting Union</i> (União Europeia de Radiodifusão)
ERB	<i>Equivalent Rectangular Bandwidth</i> (Largura de Faixa Retangular Equivalente)
FM	Frequência Modulada
FN	Falsos Negativos
FP	Falsos Positivos
GLM	<i>Generalized Linear Model</i> (Modelo Linear Generalizado)
GLME	...	<i>Generalized Linear Mixed Effects Model</i> (Modelo Linear Generalizado de Efeitos Mistos)
GPR	<i>Gaussian Process Regression Model</i> (Modelo de Regressão de Processo Gaussiano)
HATS	<i>Head and Torso Simulator</i> (Simulador de Torso e Cabeça)
HOA	<i>Higher-Order Ambisonics</i> (Ambisonics de Alta Ordem)
HRIR	<i>Head-Related Impulse Response</i> (Resposta ao Impulso Relativa à Cabeça)
HRTF	<i>Head-Related Transfer Functions</i> (Funções de Transferência Relativas à Cabeça)

- HSD *Honestly Significant Difference* (Diferença Honestamente Significativa)
- IACC *Inter-aural Cross Correlation Coefficient* (Coeficiente de Correlação Cruzada Interauricular)
- IACF *Inter-aural Cross Correlation Function* (Função de Correlação Cruzada Interauricular)
- IEC *International Electrotechnical Commission* (Comissão Internacional de Eletrotécnica)
- IIR *Infinite Impulse Response* (Resposta ao Impulso Infinita)
- IL *Intensity Level* (Nível de Intensidade)
- ILD_{Norm} *Normalized Inter-aural Level Difference* (Diferença de Intensidade Interauricular Normalizada)
- IoSR *Institute of Sound Recording* (Instituto de Gravação de Som)
- IP *Internet Protocol* (Protocolo Internet)
- ISO *International Organization for Standardization* (Organização Internacional para Padronização)
- ITD *Interaural Time Difference* (Diferença de Tempo Interaural)
- ITDG *Initial Time Delay Gap* (Lacuna Inicial de Retardo Temporal)
- ITU *International Telecommunication Union* (União Internacional de Telecomunicações)
- ITU-R *International Telecommunication Union, Radiocommunication Sector* (Setor de Radiocomunicação da União Internacional de Telecomunicações)
- ITV *Independent Television* (Televisão Independente)
- JND *Just Noticeable Difference* (Diferença no Limite do Observável)
- LFE *Low-Frequency Effects* (Efeitos de Baixa Frequência)
- LKFS *Loudness* ponderado pela curva *K*, referente ao fundo de escala digital de 0 dBFS
- LME *Linear Mixed Effects Model* (Modelo Linear de Efeitos Mistos)
- LRA *Loudness Range* (Faixa de *Loudness*)
- LU *Loudness Unit* (Unidade de *Loudness*)
- MC Ministério das Comunicações

- MCC *Matthews Correlation Coefficient* (Coeficiente de Correlação de Matthews)
- MOS *Mean Opinion Score* (Pontuação Média Opinativa)
- MPEG-H . *Moving Picture Experts Group - High efficiency coding and media delivery in heterogeneous environments* (Grupo de especialistas em cinematografia – Codificação de alta eficiência e entrega de mídia em ambientes heterogêneos)
- NHK 日本放送協会 – *Nippon Hōsō Kyōkai* (Empresa de Radiodifusão do Japão)
- NPR *National Public Radio* (Rádio Pública Nacional)
- PA *Public Address* (Endereçamento ao Público)
- PCA *Principal Component Analysis* (Análise de Componentes Principais)
- PEAQ *Perceptual Evaluation of Audio Quality* (Medida Perceptiva da Qualidade do Áudio)
- PESQ *Perceptual Evaluation of Speech Quality* (Avaliação Perceptiva da Qualidade da Fala)
- PPM *Program Peak Meter* (Medidor de Picos de Programação)
- PSE *Point of Subjective Equality* (Ponto de Igualdade Subjetiva)
- QoE *Quality of Experience* (Qualidade da Experiência)
- RF Radiofrequência
- RLB *Revised Low Frequency B-curve* (Curva B revisada em baixa frequência)
- RMS *Root Mean Square* (Raiz Média Quadrática)
- RMSE ... *Root Mean Square Error* (Raiz Média Quadrática do Erro)
- RMSE* .. *Epsilon-insensitive Root Mean-Square Error* (Raiz Média Quadrática do Erro insensível a épsilon)
- ROC *Receiver Operating Characteristic* (Característica de Operação do Receptor)
- S3A *Future Spatial Audio for an Immersive Listener Experience at Home* (Áudio Espacial Futuro para uma Experiência Imersiva do Ouvinte em Casa)

SeAC	Serviço de Comunicação Audiovisual de Acesso Condicionado, ou TV por Assinatura
SPL	<i>Sound Pressure Level</i> (Nível de Pressão Sonora)
SSE	<i>Sum of Squares of Error</i> (Soma dos Quadrados dos Erros)
SSR	<i>Sum of Squares of Regression</i> (Soma dos Quadrados da Regressão)
SST	<i>Sum of Squares of Totals</i> (Soma dos Quadrados dos Totais)
STE	<i>Short-Term Energy</i> (Energia de Curta Duração)
SVD	<i>Singular-Value Decomposition</i> (Decomposição em Valores Singulares)
SVM	<i>Support Vector Machine</i> (Máquina de Vetor de Suporte)
THD	<i>Total Harmonic Distortion</i> (Distorção Harmônica Total)
THD+N	..	<i>Total Harmonic Distortion plus Noise</i> (Distorção Harmônica Total mais Ruído)
UEN	<i>Uniform Exciting Noise</i> (Ruído de Excitação Uniforme)
UHDTV	..	<i>Ultra High Definition Television</i> (Televisão em Ultra Alta Definição)
VBAP	<i>Vector Base Amplitude Panning</i> (Sistema Vetorial de Panorama por Amplitude)
VN	Verdadeiros Negativos
VP	Verdadeiros Positivos
VU	<i>Volume Unit</i> (Unidade de Volume)
WFS	<i>Wave Field Synthesis</i> (Síntese de Campo de Onda)
ZCR	<i>Zero-Crossing Rate</i> (Taxa de Cruzamento de Zeros)

LISTA DE SÍMBOLOS

L_{eq} — Nível sonoro contínuo equivalente

S — Estímulo físico

ψ — Sensação provocada

S_0 — Limiar de estímulo físico

p_{rms} — Pressão sonora eficaz

ρ_0 — Densidade característica do meio

c — Velocidade do som

p_{ref} — Pressão sonora de referência

I_{ref} — Intensidade sonora de referência

Δf_c — Largura de faixa crítica associada à frequência central f_c

$z(f)$ — Frequência f em barks

$BW_{ERB}(f)$ — Largura de faixa na escala ERB

$ERB_N(f)$ — Frequência f na escala ERB

N — Nível de *loudness*

$E_s(t)$ — Integral de sensação

Δt — Intervalo de tempo

g — Ganho biauricular

\tilde{A}_+ — Amplitude complexa da pressão sonora radiada numa distância unitária ao centro da esfera radiante

k — Número de onda

r — Distância do receptor à esfera radiante

a_D — Curva de atenuação para campo difuso

a — Raio da cabeça

f_s — Frequência de amostragem

θ — Ângulo de azimute

τ_{sh} — Retardo de ombros e torso dependente do azimute e da elevação

τ_{p_n} — Retardo do pavilhão auricular associado à n -ésima reflexão

W — Largura de faixa efetiva

N_0 — Densidade espectral de potência do ruído gaussiano branco

Ω — Estatística de saída do modelo de detecção de energia

T — Duração do sinal

N' — Nível de *loudness* específico

S_t — *Loudness* total em sones

P_t — Nível de *loudness* em phons

a_0 — Fator de transmissão do ouvido externo e médio

$LX(f)_n$ — *Loudness* específico nos métodos ITU de avaliação de qualidade da fala (PESQ) e do áudio (PEAQ)

$W(f)$ — Função de ponderação em frequência

z_n — Nível sonoro contínuo equivalente no n -ésimo canal ponderado pela curva $K(L_{eq}(K))$

G_n — Ganho direcional do n -ésimo canal

ϕ — Ângulo de elevação

L_K — *Loudness* ponderado pela curva K em relação ao fundo de escala digital

L_{KG} — *Loudness* entrecortado

Γ_r — Limiar relativo

Γ_a — Limiar absoluto

$E_{\hat{n}}$ — Energia do sinal num único quadro

$Z_{\hat{n}}$ — Taxa de cruzamento de zeros num único quadro

Σ_x — Matriz de covariância

σ — Desvio padrão do núcleo (*kernel*) gaussiano

C — Termo de regularização

Y — Observação da variável de resposta do experimento

ε — Resíduo do modelo linear

ω — Frequência angular

df — Graus de liberdade

L_{mon} — Pressão sonora equivalente necessária para uma estimulação monótica casada com qualquer combinação biauricular

SUMÁRIO

1	INTRODUÇÃO	30
1.1	Justificativa	31
1.2	Trabalhos Progressos	33
1.3	O Caso Brasileiro	35
1.4	Problemas e Propósito	36
1.5	Contribuições	37
1.6	Organização	37
2	SENSAÇÃO DE LOUDNESS	41
2.1	Anatomia e Fisiologia do Sistema Auditivo	42
2.2	Intensidade Percebida	45
2.2.1	<i>Níveis sonoros de intensidade e pressão</i>	48
2.2.2	<i>Nível de Loudness</i>	50
2.3	Efeitos Espectrais	54
2.3.1	<i>Acumulação espectral de loudness</i>	54
2.4	Efeitos Temporais	59
2.4.1	<i>Integração temporal de loudness</i>	59
2.5	Efeitos Espaciais	64
2.5.1	<i>Somatório biauricular de loudness</i>	65
2.5.2	<i>Loudness em campos sonoros</i>	67
3	MODELOS DE LOUDNESS	77
3.1	O Modelo de Detecção de Energia	78
3.2	Modelos Multifaixa	83
3.2.1	<i>Modelos para sons estacionários</i>	84
3.2.2	<i>Modelos para sons não estacionários</i>	96
3.2.3	<i>Outros modelos</i>	99
3.3	Modelos de Faixa Única	101

3.3.1	<i>Nível sonoro contínuo equivalente (L_{eq})</i>	101
3.3.2	<i>Curvas de ponderação</i>	102
3.3.3	<i>Recomendação ITU-R BS.1770</i>	107
3.3.4	<i>Recomendação EBU R.128</i>	113
4	PROPOSTAS E EXPERIMENTOS PRELIMINARES	121
4.1	<i>Loudness na Radiodifusão Digital</i>	122
4.1.1	<i>Consolidação do padrão ITU-R</i>	122
4.1.2	<i>Desenvolvimentos recentes</i>	126
4.1.3	<i>Linhas de investigação</i>	129
4.1.4	<i>Cursos de ação</i>	133
4.2	Controle Automático de Loudness em Conteúdo de Formato Curto	133
4.2.1	<i>Detetor de conteúdo de formato curto</i>	139
4.2.2	<i>Controlador de loudness</i>	147
4.2.3	<i>Resultados e discussões</i>	150
4.3	Medição de Loudness para Áudio Espacial	155
4.3.1	<i>Métodos</i>	156
4.3.2	<i>Proposta</i>	160
4.3.3	<i>Testes</i>	161
4.3.4	<i>Resultados e discussões</i>	166
5	PROPOSTAS E EXPERIMENTOS PRINCIPAIS	170
5.1	Estágio Doutoral	171
5.2	<i>Loudness na Radiodifusão Digital (revisitado)</i>	172
5.2.1	<i>Sob que aspectos a norma brasileira de loudness pode ser revisada?</i>	175
5.2.2	<i>Como o modelo de loudness do ITU-R pode ser aprimorado para áudio imersivo?</i>	179
5.3	Relação entre Loudness e Distância	183
5.3.1	<i>Verificações preliminares</i>	185
5.3.2	<i>Projeto de experimento</i>	199
5.3.3	<i>Experimento principal</i>	207
5.3.4	<i>Modelo de loudness ITU-R como função da distância</i>	214

5.4	Relação entre <i>Loudness</i> e Reverberação	217
5.4.1	<i>Efeito da reverberação</i>	217
5.4.2	<i>Projeto de experimento</i>	224
5.4.3	<i>Experimento principal</i>	233
5.4.4	<i>Modelo de loudness ITU-R como função da reverberação</i>	243
5.5	Relação entre <i>Loudness</i> e Direção	246
5.5.1	<i>Loudness direcional no modelo ITU-R</i>	247
5.5.2	<i>Projeto de experimento</i>	251
5.5.3	<i>Verificação preliminar</i>	256
5.5.4	<i>Teste piloto</i>	259
5.5.5	<i>Experimento principal</i>	264
5.5.6	<i>Estimação de ganhos: problema de otimização</i>	277
5.5.7	<i>Estimação de ganhos: problema de regressão</i>	280
5.5.8	<i>Modelo de loudness ITU-R como função da direção</i>	285
5.6	Medição de Loudness para <i>Áudio Espacial</i> (revisitada)	287
5.6.1	<i>Áudio baseado em objetos</i>	288
5.6.2	<i>Proposta de modelo para objetos sonoros</i>	291
5.6.3	<i>Avaliação do modelo</i>	292
5.6.4	<i>Resultados e discussões</i>	295
6	CONCLUSÃO	301
6.1	Principais Contribuições	303
6.2	Trabalhos Futuros	306
	REFERÊNCIAS	308
	Glossário	334
	APÊNDICE A PUBLICAÇÕES	338
	ANEXO A LINKS INTERESSANTES	340
	ANEXO B DOCUMENTOS DE APOIO	342

INTRODUÇÃO

A sensação de intensidade do som, ou *loudness*, não depende inteiramente da amplitude da onda acústica, mas também de fatores como frequência, espacialidade e duração. No caso de conteúdo de áudio para consumo, a percepção da intensidade sonora não depende somente do ajuste do volume, mas também da origem e do formato do segmento de áudio. Peças veiculadas em emissoras de radiodifusão, em transmissões sobre Protocolo Internet (IP) ou em distribuições físicas, são processadas espectral e dinamicamente para cumprimento de requisitos comerciais, técnicos e estéticos de masterização. Esse processamento pode alterar a percepção de *loudness* do material processado e, por consequência, fazer com que a audiência experimente “saltos” de intensidade percebida entre fontes (radiodifusão sonora, de sons e imagens e *streaming*), entre canais de uma mesma fonte (a exemplo do Serviço de Comunicação Audiovisual de Acesso Condicionado (SeAC)), e entre segmentos de programação num mesmo canal (a exemplo dos blocos de propaganda na radiodifusão de sons e imagens).

Se o *loudness* pudesse ser medido objetivamente de alguma forma, seria então possível controlá-lo no áudio da programação. Historicamente, a percepção da intensidade sonora é objeto de estudo da Psicoacústica e tem sido alvo de experimentos laboratoriais por décadas com o uso de sinais sintéticos com atributos invariantes no tempo. Procedimentos de medida objetiva também trilharam uma esteira de desenvolvimento em busca de um modelo de *loudness* enquanto medidor de material de áudio não estacionário, tal como segmentos de fala e

música, e comerciais de rádio e televisão.

Em resposta a essa necessidade, o Setor de Radiocomunicação da União Internacional de Telecomunicações (ITU-R) avaliou medidores objetivos de intensidade percebida de áudio para posteriormente elaborar uma recomendação sobre o tema. Na ocasião, o método de nível sonoro contínuo equivalente L_{eq} ponderado pela Curva B revisada em baixa frequência (RLB) teve um melhor desempenho para sinais monofônicos (SOULODRE, 2004). Posteriormente, o método $L_{eq}(RLB)$ foi estendido para medidas de *loudness* em áudio multicanal – estéreo e 5.1 à época – com a introdução de um pré-filtro projetado para contabilizar os efeitos acústicos da cabeça. Este algoritmo culminou na Recomendação “*Algorithms to measure audio programme loudness and true-peak audio level* (Algoritmos para medida de *loudness* em programas de áudio e para medida de nível de pico verdadeiro em áudio)” ITU-R BS.1770 (ITU-R, 2015b), padrão de medida de *loudness* para difusão, distribuição e entrega de conteúdo de áudio. Recomendações e regulamentos regionais para controle de *loudness* foram publicados com base no algoritmo ITU-R BS.1770, inclusive no Brasil (MC, 2012).

1.1 Justificativa

A compressão de faixa dinâmica reduz os picos de uma forma de onda acima de um determinado limiar enquanto os níveis abaixo deste mesmo limiar permanecem inalterados. Surgida originalmente na adaptação de gravações a meios/canais naturalmente restritos em faixa dinâmica, a exemplo de fitas, películas e da radiodifusão analógica (BLESSER, 1969), é um efeito frequentemente usado para se fazer com que a trilha de áudio soe mais intensa sem elevar sua amplitude. Ao se comprimir as partes mais “fortes” do som, é possível aumentar a energia do sinal de áudio sem exceder os limites de faixa dinâmica de um bloco quantizador de um dispositivo de reprodução.

Com a digitalização do áudio, a compressão de faixa dinâmica passou a ser usada com fins artísticos. Realce de efeitos sonoros em trilhas de áudio para cinematografia/teledramaturgia, destaque de solistas em concertos e manutenção do baixo volume de músicas reproduzidas em espaços públicos são alguns

exemplos de aplicação (ZÖLZER, 2011). Graças à sua crescente popularidade, o recurso foi apropriado pelo mercado fonográfico e é usado à exaustão na produção musical voltada à execução radiofônica, marcada pelo uso excessivo de compressão para que os produtos de uma dada gravadora pareçam mais empolgantes que os de suas concorrentes, prática essa conhecida como Guerra de Intensidade de Áudio (*Loudness War*) (VICKERS, 2010).

Não tardou para que as vantagens da compressão dinâmica fossem apreciadas pelo mercado publicitário e incorporadas ao conceito de poder de parada (*stopping power*). Oriundo da balística, o dito poder de parada de uma peça publicitária está associado ao potencial de captura da atenção da audiência (WELLS, 1997). Consequentemente, abusos de compressão dinâmica passaram a ocorrer também na produção de peças publicitárias para televisão. Na radiodifusão aberta, por exemplo, é comum o telespectador assistir a um filme na sua sala de estar e, ao aproveitar os intervalos comerciais para se levantar do sofá, conseguir ouvir as propagandas mesmo estando em outro cômodo da casa.

No que tange à conformidade de conteúdos de formato curto tais como comerciais, anúncios e inserções ao vivo, ao longo da adoção prática dos parâmetros constantes das Recomendações BS.1770 do ITU-R (2015b) e R.128 da União Europeia de Radiodifusão – EBU (2014) pela comunidade de profissionais de áudio, os próprios organismos de padronização reavaliaram os papéis de alguns dos descritores de *loudness* utilizados. Segundo a EBU (2016a), para o controle de peças altamente comprimidas, são recomendados os parâmetros “*Loudness Momentâneo Máximo*”, que emprega uma janela deslizante de 400 milissegundos, e o “*Loudness de Curta Duração*” (SKOVENBORG; NIELSEN, 2007), que emprega uma janela deslizante de 3 segundos (SKOVENBORG; NIELSEN, 2007). Por outro lado, o parâmetro “Faixa de *Loudness*” (SKOVENBORG, 2012a) não seria aplicável para descrever o *loudness* dessas peças por ser baseado numa análise estatística de valores de “*Loudness de Curta Duração*” (3 segundos), uma vez que, para comerciais e vinhetas, resulta num conjunto de valores pequeno demais para se obter um resultado significativo. Já os descritores “*Loudness Médio*” e “Nível Máximo de Pico Verdadeiro” por si só são insuficientes para caracterização de comerciais, vinhetas e inserções. Nesse contexto, cabe revisar os descritores de intensidade percebida utilizados no regulamento

do antigo Ministério das Comunicações (MC, 2012) e seus valores de referência.

Repensar a medida de *loudness*, porém, vai além da revisão das normas nacionais, considerando o advento dos formatos de áudio imersivo com mais de seis canais e a recente finalização dos padrões MPEG-H 3D Audio (ISO, 2015) e Dolby AC-4 (KJÖRLING *et al.*, 2016) de codificação de áudio de nova geração para Televisão em Ultra Alta Definição (UHDTV). Muito embora a revisão mais recente da Rec. ITU-R BS.1770 contemple o suporte a um número arbitrário de canais e alto-falantes (KOMORI *et al.*, 2015), tanto o algoritmo padrão quanto os demais modelos de *loudness* existentes não foram suficientemente testados em tais condições. E ao contrário do áudio tradicional baseado em canais, formatos de nova geração, como o áudio baseado em objetos e em cenas, são agnósticos em relação ao número de alto-falantes. Nestes, o conteúdo de áudio é transmitido na forma de objetos em conjunto com metadados que os definam, ou de cenas caracterizadas por coeficientes Ambisonics de Alta Ordem (HOA) de harmônicos esféricos, que descrevem todos os sons e suas propriedades espaciais não estacionárias durante uma transmissão, independentemente do número de fontes e da disposição dos alto-falantes na reprodução (KUECH *et al.*, 2015). Como consequência, para que algoritmos de controle de faixa dinâmica e *loudness* sejam aplicados aos novos padrões de áudio, o conteúdo codificado deve ser auralizado num arranjo virtual de alto-falantes, ou seja, ter a impressão aural das características acústicas de um espaço de reprodução. Os sinais oriundos dos alto-falantes virtuais são controlados em intensidade e dinâmica, e em seguida são novamente convertidos para a forma de metadados até a etapa de renderização (PETERS *et al.*, 2015). Portanto, a adaptação de um método de medida baseado em canais para operação em áudio imersivo é uma meta a ser perseguida.

1.2 Trabalhos Progressos

Os primeiros estudos sobre intensidade sonora percebida datam dos anos 1930 e tiveram suas descobertas bem documentadas na literatura (FLETCHER; MUNSON, 1933; STEVENS, 1957; MOORE, 2012; ZWICKER; FASTL, 2013). Os modelos de predição de *loudness* que surgiram em seguida podem ser classifi-

cados como modelos de faixa única e modelos multifaixa. Os modelos multifaixa dividem o sinal de entrada em múltiplas faixas de frequência que são subsequentemente combinadas numa estimação de intensidade, enquanto que os modelos de faixa única possuem um único caminho de sinal.

Dentre os primeiros modelos multifaixa, somente três modelos projetados para sons estacionários foram reconhecidos pela comunidade internacional: os modelos de Zwicker (1958), Stevens (1961) e Moore e Glasberg (1996). Em 1975, métodos gráficos baseados nos modelos de Stevens (Método A) e Zwicker (Método B) tornaram-se o padrão da Organização Internacional para Padronização (ISO) 532 (ISO, 1975), cujo método ISO 532B teve uma implementação na linguagem (BASIC) publicada em 1984 (ZWICKER; FASTL; DALLMAYR, 1984) e, mais recentemente, foi incluído no padrão do Instituto Alemão para Padronização (DIN) 45631 (ZWICKER *et al.*, 1991). O modelo de Moore foi revisado em 1997 para incluir o cálculo de *loudness* parcial (MOORE; GLASBERG; BAER, 1997) e também foi objeto do padrão norte-americano do Instituto Nacional Americano de Padrões (ANSI) S3.4-2007 (ANSI, 2012), que substituiu sua versão de 1980 baseada no método ISO 532A. Para o caso de sons não estacionários, isto é, nos quais as características temporais e espectrais mudam ao longo de suas durações, dois modelos foram desenvolvidos: o primeiro por Zwicker e Fastl (1999) e o segundo por Glasberg e Moore (2002), sendo aquele incluído na primeira emenda do padrão DIN 45631.

Os modelos de banda única baseiam-se na dependência em frequência da sensação de *loudness* e no emprego de algum detetor de envoltória (GREEN; SWETS, 1966). Uma ponderação em frequência é baseada na aproximação de uma dos contornos de mesmo *loudness* (FLETCHER; MUNSON, 1933). Embora a acumulação espectral de intensidade não possa ser modelada usando um método de faixa única (SCHLITTENLACHER; ELLERMEIER; HASHIMOTO, 2015), estes modelos são tipicamente voltados para sinais de banda larga com conteúdo espectral semelhante, com uma faixa de níveis razoavelmente estreita, como é o conteúdo de radiodifusão.

A medida de integração de energia de longa duração, denominada nível sonoro contínuo equivalente, ou L_{eq} , é um valor eficaz, ou valor de Raiz Média Quadrática (RMS) no tempo (BRIXEN, 2011). Ponderações padronizadas

pela Comissão Internacional de Eletrotécnica (IEC) com a integração L_{eq} são representadas pelas curvas A, B e C, para intensidades fracas, medianas e fortes, respectivamente (IEC, 1979). Todavia os padrões *de facto* na indústria de conteúdo de áudio foram as integrações: $L_{eq}(A)$, decibelímetro padrão, muito usada para controle de faixas de diálogo; e $L_{eq}(M)$, proposta pelos Laboratórios Dolby (1988) para medida de irritabilidade em filmes e trilhas sonoras. Posteriormente, a curva *RLB* foi proposta por Soulodre (2004) para radiodifusão e o método foi então estendido para uma versão multicanal (SOULODRE; LAVOIE, 2005). Após a introdução de um pré-filtro de sombreamento de cabeça na etapa de ponderação em frequência, o novo modelo $L_{eq}(K)$ culminou no padrão ITU-R (2015b), *de jure* para a produção de conteúdo de áudio digital.

O documento ITU-R (2015b) atualmente se encontra em sua quarta versão. Na segunda edição foi introduzido o conceito de *loudness* entrecortado (*gated loudness*)¹ proposto pela União Europeia de Radiodifusão (EBU). A medida de nível de pico verdadeiro proposta por Dash (2014) foi incluída na terceira revisão, e a versão mais recente acomodou a proposta de Komori *et al.* (2015) para sistemas de áudio com mais de cinco canais no cálculo do algoritmo.

1.3 O Caso Brasileiro

A crescente insatisfação dos usuários com os saltos de intensidade entre programas e comerciais na radiodifusão brasileira culminou na promulgação da Lei nº 10.222, de 9 de maio de 2001 (PR, 2001), que objetivou padronizar o volume de áudio das transmissões de rádio e televisão nos espaços dedicados à propaganda. A carta decretou que os serviços de radiodifusão sonora e de sons e imagens padronizariam seus sinais de áudio, de modo a não haver, no momento da recepção, elevação injustificável de volume nos intervalos comerciais, assim como incumbiu o Poder Executivo de criar, no período de cento e vinte dias a contar da data de publicação, os mecanismos necessários à normalização técnica da matéria, bem como à fiscalização de seu cumprimento. Contudo, como não

¹ O *Loudness* entrecortado consiste no uso de uma função portão na saída da etapa de integração, que segmenta as integrações em intervalos de fechamento de 400 milissegundos com 75% de sobreposição, descontados de dois limiares: os blocos de silêncio, medidos a -70 LKFS, e os blocos considerados muito baixos, medidos 10 LU abaixo do nível total medido após eliminação dos blocos inferiores ao primeiro limiar (EBU, 2014).

havia medidores objetivos de *loudness* para radiodifusão à disposição no período, os cento e vinte dias converteram-se em onze anos para a normatização técnica pelo Ministério das Comunicações (MC, 2012) e subsequente elaboração de um procedimento de fiscalização pela Anatel (2014).

Ambas as portarias são baseadas na segunda versão da Recomendação ITU-R BS.1770 e usam os descritores da segunda versão da Recomendação EBU R 128, ambas de 2011. Estas não possuem suporte a sistemas de reprodução com mais de seis canais, e não fazem uso de descritores curtos de *loudness*. Já a Lei nº 10.222/2001 teve sua redação alterada pela Lei nº 12.810, de 15 de maio de 2013 (PR, 2013) e seu escopo reduzido à radiodifusão com tecnologia digital. Consequentemente, os serviços de radiodifusão analógica, de distribuição de conteúdo sobre IP, e de conteúdo audiovisual de acesso condicionado não estão contemplados pelo arcabouço regulatório nacional.

1.4 Problemas e Propósito

Do ponto de vista do regulador, há duas frentes de trabalho a se considerar: i) examinar a norma vigente e elaborar uma estratégia para melhorar sua efetividade à luz da experiência internacional e, ii) pensar em como aprimorar o algoritmo multicanal BS.1770 para os novos formatos de áudio espacial, tendo por motivação as questões de estudo sobre o tema em andamento no ITU-R.

O primeiro problema passa por uma leitura crítica do arcabouço regulatório para identificação dos pontos de melhoria, seguido de verificação experimental da eficiência de suas implementações, e da definição de valores de referência. Já o segundo consiste em recortar a pesquisa de ponta, enxergar como o estado da arte se traduz nas questões de estudo em andamento no ITU-R, e estabelecer um racional de experimentos rumo a um modelo de *loudness* para objetos sonoros perceptivamente motivado.

O propósito desta pesquisa é elaborar um arcabouço teórico consistente para medição de *loudness* em sistemas avançados de reprodução de áudio. Não há modelos listados em padrões internacionais com esta configuração, os modelos consolidados não foram suficientemente testados em áudio baseado em objetos/cenas e estudos mais recentes de *loudness* são ainda incipientes nesse

contexto. Não cabe aqui elaborar toda uma crítica teórica de intensidade subjetiva de áudio, mas sim elencar teoria e ferramentas necessários para implementar e analisar o modelo em questão. Espera-se que futuros trabalhos sobre intensidade percebida de áudio se beneficiem dos resultados produzidos nesta pesquisa.

1.5 Contribuições

Embora o principal objetivo deste projeto de pesquisa seja implementar um modelo de *loudness* para áudio espacial, o trabalho buscou obter tanto resultados científicos quanto *insights* regulatórios, podendo alimentar proposições de futuras normas nacionais.

Primeiramente, o estudo tentou relacionar medição de *loudness* com auralização e com testes subjetivos tanto em termos de implementação quanto em análise de desempenho. Uma cadeia de sinal que abranja desde a renderização dos sinais até um algoritmo de medida de intensidade percebida no *sweet-spot* não foi coberta por trabalhos anteriores à data de redação. Portanto, as implementações descritas neste texto caminham nesta direção².

Por fim, espera-se que o levantamento de novos padrões internacionais e regulamentos regionais traga insumos sobre descritores de intensidade subjetiva e valores de referência atinentes aos produzidos pelo algoritmo ITU-R BS.1770, subsidiando de tal forma potencial revisão da regulamentação brasileira de *loudness* para radiodifusão digital e propostas futuras de regulamentos para o SeAC e para conteúdo distribuído sobre IP (*streaming*).

1.6 Organização

Esta seção apresenta a estrutura da investigação desenvolvida na presente tese. O [Capítulo 2](#), a seguir, cobrirá o arcabouço teórico desenvolvido no campo da psicoacústica sobre intensidade sonora percebida, abrindo com uma descrição das propriedades essenciais do sistema auditivo relacionadas à percepção de intensidade, seguida por uma definição da sensação de *loudness* enquanto

² A Recomendação ITU-R BS.2127-0: *Audio Definition Model renderer for advanced sound systems*, que traz especificações de um renderizador para sistemas avançados de áudio, foi publicada somente em junho de 2019, coincidentemente com a conclusão desta pesquisa.

entidade subjetiva (seria a expressão “*loudness* percebido” uma tautologia?). A necessidade de se medir objetivamente a intensidade percebida levou a duas representações de *loudness* distintas: *loudness* calculado, expresso em *sones*, e nível de *loudness*, expresso em *phons*. Já a segunda metade do Capítulo cobrirá os efeitos na percepção de intensidade causados por frequência, duração e espacialidade. Dentre os efeitos espectrais estão o mascaramento em frequências e a dependência da largura de faixa, das quais derivam as bandas críticas do sistema auditivo. A seção de efeitos temporais elenca o mascaramento no tempo e a chamada integração temporal, referente ao tempo de subida entre o início da execução de um som de intensidade constante até a percepção desta mesma intensidade como sendo constante. Por fim, será descrita a dependência da direção do som e o efeito denominado “somatório biauricular de *loudness*”.

O Capítulo 3, por sua vez, detalhará a fortuna crítica de modelos de predição de *loudness*, classificados em modelos multifaixa e modelos de faixa única. Dentre os modelos multifaixa para sons estacionários descritos estão os métodos de Stevens (1961), de Zwicker *et al.* (1991) e de Moore, Glasberg e Baer (1997), dos quais evoluíram os principais modelos multifaixa para sons não estacionários: os modelos de Zwicker e Fastl (1999) e de Glasberg e Moore (2002). Este último será brevemente comparado com o modelo dinâmico de Chalupper e Fastl (2002) e, a partir desta comparação, serão definidos os conceitos de *loudness* integrado e de *loudness* de curta duração. A segunda metade do capítulo tratará dos modelos de faixa única historicamente mais relevantes: integrações L_{eq} nas suas diferentes ponderações em frequência e o algoritmo ITU-R BS.1770 (ITU-R, 2015b). Este último, por ser o padrão vigente para produção de áudio multicanal, será mais detalhado na sua concepção original e nas modificações aos quais foi submetido de 2006 a 2015, sendo boa parte delas propostas pela EBU (2014).

Já o Capítulo 4 traçará um panorama da pesquisa recente e elencará os experimentos preliminares para obtenção dos primeiros *insights* sobre as perguntas da pesquisa. O primeiro experimento refere-se à implementação de um controlador de intensidade de tempo-real com critérios baseados em descritores de *loudness* momentâneo e de *loudness* de curta duração, conceitos importados dos modelos multifaixa para sinais não estacionários e empregados no funcionamento do algoritmo ITU-R, cujos valores de referência poderiam somar-se às

métricas constantes da normativa brasileira. Neste mesmo controlador, é também proposto um classificador binário de conteúdo de formato curto a partir de um espaço de características extraídas somente do áudio. No segundo experimento, é discutida uma primeira proposta de aprimoramento do modelo ITU-R para conteúdo de áudio imersivo que consiste na substituição do filtro de sombreamento de cabeça por síntese binauricular de sinais provenientes de um arranjo virtual de alto-falantes, auralizado a partir de respostas ao impulso de salas de referência. A ponderação em frequência também conta com a adição do modelo do meato acústico proposto por [Moore, Glasberg e Baer \(1997\)](#). A primeira implementação será avaliada pela acurácia de classificação e pelo sucesso no processamento da dinâmica das faixas que excederam os critérios de referência. Já a segunda será comparada com os métodos descritos no [Capítulo 3](#) em razão das diferenças médias das predições de *loudness* de conteúdo de áudio imersivo em relação às predições em conteúdos de referência em cinco canais.

Por fim, o [Capítulo 5](#) analisará as perspectivas obtidas pelos resultados dos experimentos preliminares à luz da experiência internacional regulatória e de produção, ao fechar linhas de investigação em aberto e ao amadurecer os cursos de ação propostos no capítulo anterior. Os valores de referência usados na implementação do controlador de *loudness* que venham ao encontro das melhores práticas regulatórias internacionais, comporão a estratégia de modificação da norma brasileira a ser elaborada. Para tanto, realizei estágio doutoral no Instituto de Gravação de Som da Universidade de Surrey, no Reino Unido, de agosto de 2017 a julho de 2018, ao longo do qual foram realizados testes de escuta para se avaliar experimentalmente os efeitos desses parâmetros na sensação de *loudness* dos ouvintes, e se construir modificações no algoritmo BS.1770 baseadas nas respostas destes participantes. Assim sendo, este capítulo descreve a evolução da proposta preliminar até um modelo consolidado, perceptivamente motivado, baseado em parâmetros posicionais constantes dos metadados associados ao áudio medido. Este modelo será então comparado não somente com os métodos descritos no [Capítulo 3](#), como também com o próprio modelo preliminar apresentado no [Capítulo 4](#).

O [Capítulo 6](#) faz uma síntese da pesquisa tal como foi apresentada neste documento. Relata as principais descobertas do estudo, enumera suas contribuições

e propõe direções para sua continuidade em trabalhos futuros.

O [Apêndice A](#) contém uma relação das publicações em eventos de divulgação científica que progressivamente relataram os avanços deste trabalho. O [Apêndice A](#) traz links de referência geral e o [Apêndice B](#) contém um conjunto de documentos que deram suporte aos experimentos perceptivos realizados nesta pesquisa.

SENSAÇÃO DE LOUDNESS

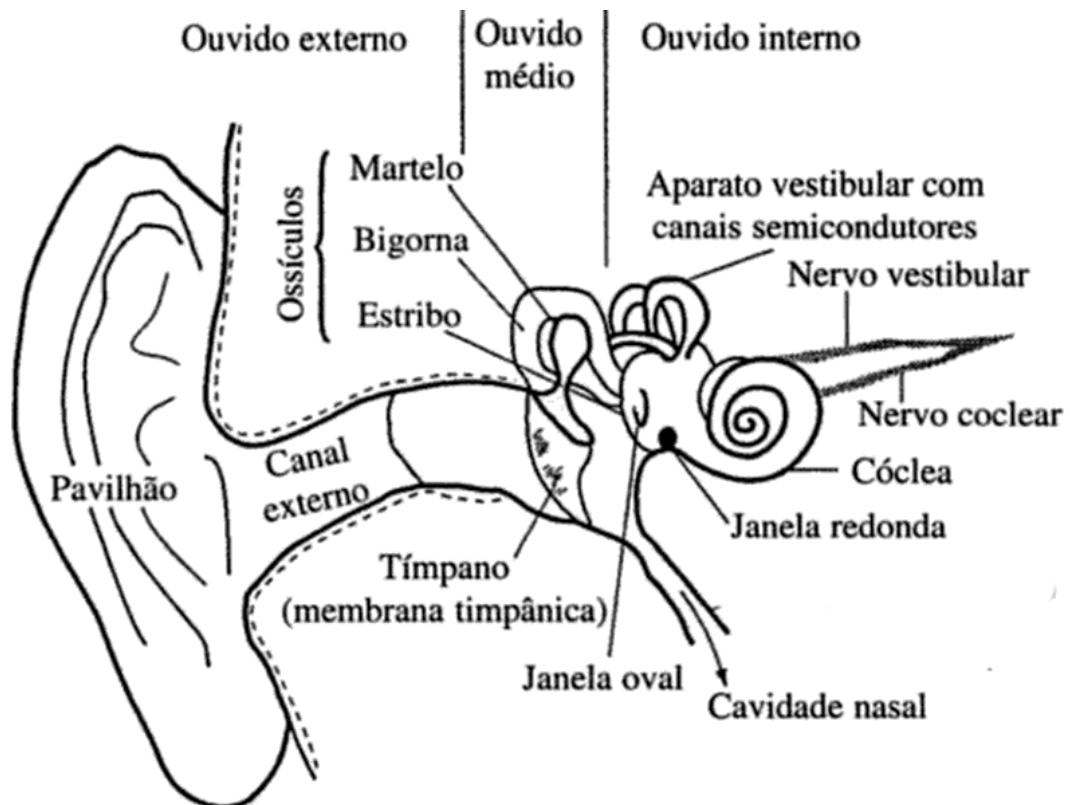
A sensação de *loudness* constitui uma das principais dimensões da percepção auditiva. É uma quantidade psicológica – não física – considerada por Florentine, Popper e Fay (2010) como sendo o “o correlato psicológico *primordial* do nível físico”. Definições encontradas na literatura abrangem desde afirmações diretas do tipo “o atributo do som que varia mais prontamente com a alteração da intensidade sonora” (SCHARF, 1969) até explicações crípticas como “percepção da intensidade do som ou dos sinais de áudio quando estes são reproduzidos acusticamente, tratando-se de uma função complexa, que pode ser medida objetivamente por meio de algoritmos definidos na Recomendação ITU-R BS.1770-2 e na Recomendação EBU R-128- 2011” (MC, 2012), passando por verdadeiras tautologias como “percepção de *loudness*” (ZEMACK, 2007). Logo, uma definição inequívoca deve ser feita de maneira cuidadosa.

Loudness é uma sensação de magnitude que corresponde à intensidade percebida de pressão sonora e que depende do nível de pressão, da frequência, da duração e da espacialidade dos sons. Logo, é correlacionada com o próprio funcionamento do sistema auditivo humano, descrito na seção 2.1. A seção 2.2 define *loudness* como sensação e como grandeza mensurável. Já as demais seções explorarão as características do som das quais a percepção de intensidade é dependente.

2.1 Anatomia e Fisiologia do Sistema Auditivo

O sistema auditivo humano é representado na [Figura 2.1](#) e está dividido em três seções anatômicas distintas denominadas ouvido externo, ouvido médio e ouvido interno ([FLANAGAN, 2013](#)).

Figura 2.1 – Estrutura do Ouvido Humano (externo, médio e interno)



Fonte: Adaptada de [Flanagan \(2013, p.87\)](#).

O ouvido externo é composto pelo *pavilhão auricular* (orelha) e um canal aproximadamente cilíndrico de 25 mm de comprimento por 10 mm de diâmetro denominado *meato* ou *canal auditivo*. E como demonstrado por Hermann von [Helmholtz \(1860\)](#), possui ressonância em frequência específica segundo a [Equação 2.1](#):

$$f = \frac{c}{2\pi} \sqrt{\frac{2r_0}{V}}, \quad (2.1)$$

onde r_0 é o raio do orifício, V é o volume da cavidade e c é a velocidade do som ([CHANAUD, 1994, p. 337](#)). Para as dimensões acima, o meato acústico ressoa aproximadamente em 3800 Hz ([CAMPBELL; GREATER, 1994, p. 41](#)).

O pavilhão auricular funciona como um coletor da energia sonora incidente, que a canaliza para uma área menor. Se pusermos uma das mãos atrás da orelha como uma concha, por exemplo, a área do pavilhão aumenta e os sons são percebidos com maior intensidade. A orelha também funciona como um estimador de direção dos sons, pois as ondas refletidas no meato acústico, a partir da captação proveniente de diferentes partes do pavilhão, viajarão por distâncias diferentes. O cérebro é então capaz de estimar a direção do som a partir destas diferenças de tempo de chegada.

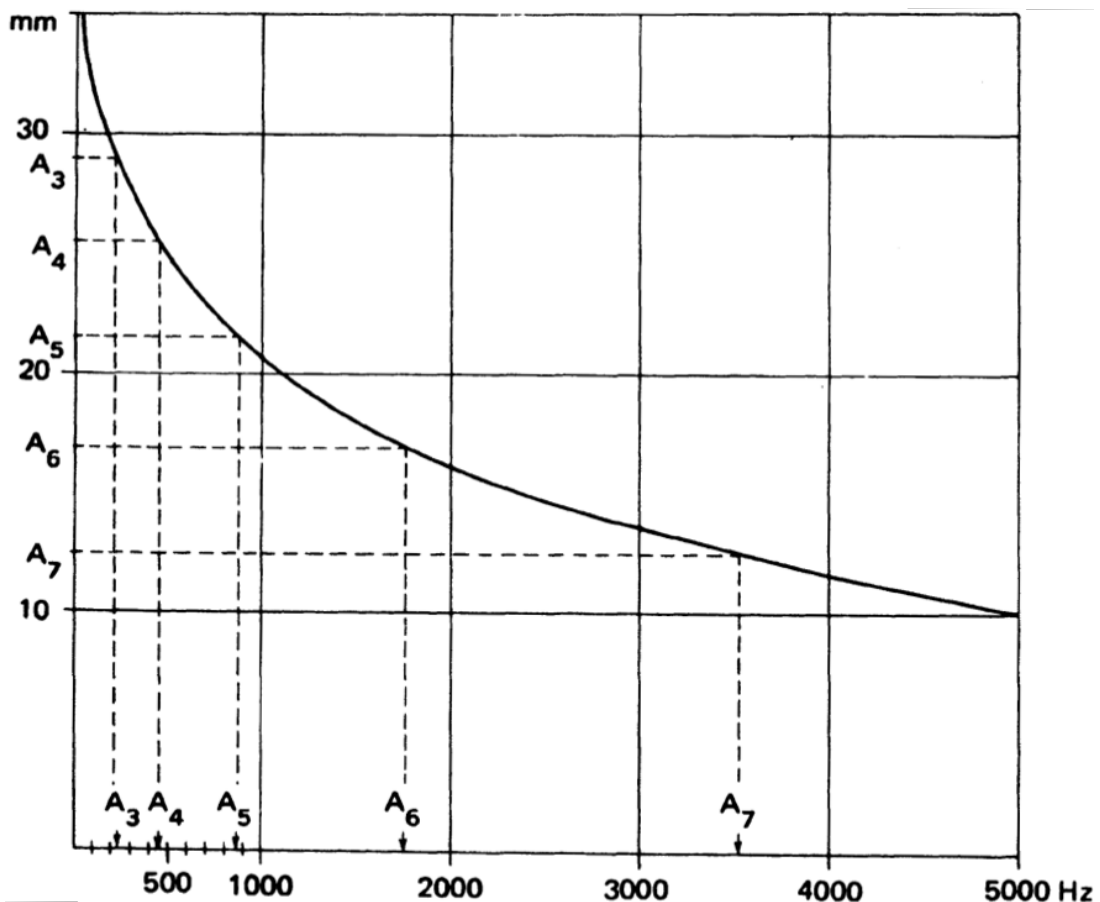
O ouvido médio compreende do tímpano à janela oval. Sua principal função é o casamento de impedâncias entre a propagação sonora no ar (baixa impedância) e a propagação fluídica no ouvido interno (alta impedância). Consegue-se este casamento devido à razão da área maior da membrana timpânica pela área menor de contato do estribo com a janela oval, e à alavanca formada desde a conexão longa da bigorna com o tímpano até a conexão curta do martelo com o estribo. A transformação total resulta num acréscimo de pressão num fator de até cinquenta vezes (CAMPBELL; GREATER, 1994, p. 46).

No ouvido interno, os *canais semicirculares* são responsáveis pelo senso de equilíbrio e a *cóclea* tem como função primordial a análise espectral de sons contendo várias componentes em frequência. A análise é feita por uma membrana situada entre as galerias superior e inferior da cóclea chamada *membrana basilar*. A Figura 2.2 mostra como as posições da membrana basilar de máxima sensibilidade variam com a frequência.

A Figura 2.2 mostra que a faixa de 20-4000 Hz cobre aproximadamente dois terços da extensão da membrana basilar enquanto a faixa superior a 4000 Hz cobre o terço restante. Note que quando a frequência de um tom é dobrada – ou quando a nota sobe em uma oitava – a região correspondente à máxima resposta é reduzida em aproximadamente 3,5 a 4 milímetros, independentemente dos valores absolutos de cada frequência e seu dobro. Quem determina o deslocamento da região ressonante ao longo da membrana basilar não são as diferenças entre as frequências, mas sim suas *razões*. Por isso nossa percepção auditiva é dita logarítmica (ROEDERER, 2008, p. 32).

O processamento de sinais no sistema auditivo pode ser descrito por uma sucessão de vários estágios de processamento que modelam o funcionamento

Figura 2.2 – Posições de máxima ressonância da membrana basilar para tons puros.



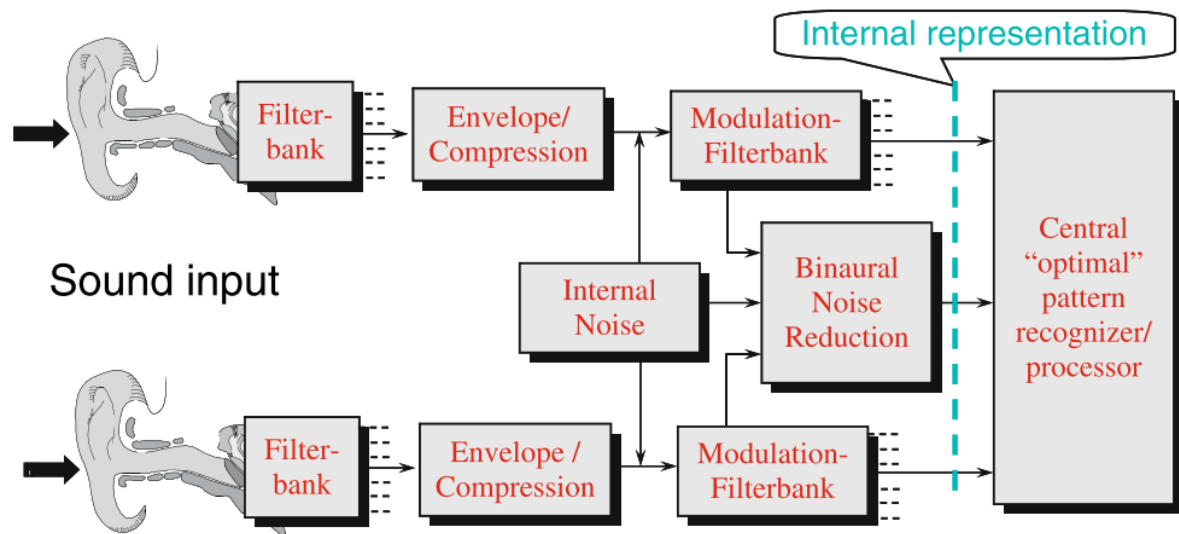
Fonte: Roederer (2008, p.32).

Nota – O eixo das ordenadas representa a extensão da membrana basilar excitada para cada tom.

do ouvido sem levar em consideração sua real implementação biológica (HA-VELOCK; KUWANO; VORLÄNDER, 2008, p. 156). Uma visão esquemática deste processamento é ilustrada na Figura 2.3.

O primeiro estágio consiste numa filtragem de banda larga correspondente à resposta em frequência pelo ouvido médio, seguido por um banco de filtros que distribui o som em diferentes faixas de frequência – ou canais – de modo semelhante ao feito pela membrana basilar (ZWICKER *et al.*, 1991). Em cada canal, a energia instantânea é obtida pela extração da envoltória temporal, classicamente feita por um retificador de meia-onda seguido de um filtro passa-baixas (VIEMEISTER, 1979). Esta etapa e a seguinte, que consiste num banco de filtros de modulação seguida de comparação binauricular, propõem-se a simular o funcionamento das células ciliadas e do nervo auditivo (HEWITT; MEDDIS, 1994). A saída destes estágios de processamento pode ser concebida como um padrão

Figura 2.3 – Visão esquemática do processamento de sinais no sistema auditivo.



Fonte: Havelock, Kuwano e Vorländer (2008, p.157).

de duas dimensões discreto no tempo, que representa frequência num eixo e flutuação de amplitude (modulação AM) no outro. Para cada combinação de frequência central e frequência modulante, há uma relação binauricular presente (KOLLMEIER; KOCH, 1994).

Quanto mais à direita se anda no diagrama da Figura 2.3, mais complexo é o modelo auditivo resultante, até se chegar ao limite da unidade cerebral de classificação de padrões auditivos, cujo funcionamento ainda é muito pouco conhecido, e a esta altura podemos somente postular os mecanismos usados pelo cérebro humano para processar som (RABINER; SCHAFER, 2007, p. 32). Ainda assim, uma boa carga de conhecimento sobre percepção sonora foi produzido por experimentos de estímulo do sistema auditivo por tons e ruídos específicos e controlados, especialmente no que tange à sensibilidade do sistema auditivo a propriedades acústicas tais como intensidade e frequência.

2.2 Intensidade Percebida

A noção de audição logarítmica levou à observação de que a função de transferência da magnitude física do estímulo sonoro para sua magnitude percebida é não-linear, noção que veio ao encontro das primeiras observações da psicofísica. Destacam-se aqui os trabalhos pioneiros de Gustav Fechner (1877),

nas suas próprias palavras:

A torre de Babel nunca foi terminada porque os trabalhadores não chegaram a um consenso sobre como deveriam tê-la construído; meu edifício psicofísico perdurará pois os trabalhadores nunca concordarão acerca de como destruí-lo (FECHNER, 1877, *Nachwort* (Epílogo), tradução minha)

O primeiro tijolo de seu “edifício psicofísico” foi assentado com a formulação da Lei de Weber (FECHNER, 1877) em homenagem a Ernst Weber, cujos experimentos tornaram-na possível (WEBER; ROSS; MURRAY, 1996, p. 9). A Lei foi originalmente enunciada da seguinte forma: “a diferença simples de sensibilidade é inversamente proporcional ao tamanho dos componentes desta diferença; a diferença relativa de sensibilidade permanece a mesma independentemente do tamanho dos componentes”, ou seja, a mudança percebida em um estímulo S é proporcional à mudança física ocorrida neste estímulo. A formulação da Equação 2.2 contém a definição de Diferença no Limite do Observável (JND), que é a menor mudança possível de ser percebida num estímulo físico. A JND é constante, independente das diferenças na sensação.

$$\frac{dS}{S} = \frac{JND}{S} = \text{constante}, \quad (2.2)$$

A Lei de Fechner (1907) foi uma extensão da Lei de Weber, com base em suas observações de que as sensações humanas aos estímulos físicos dependem do sentido estimulado, e de que a sensação provocada ψ é proporcional ao logaritmo da intensidade do estímulo S . A lei é escrita da forma

$$d\psi = k \frac{dS}{S}, \quad (2.3)$$

na qual k é uma constante de proporcionalidade. Integrando esta expressão, temos

$$\psi = k \ln S + C, \quad (2.4)$$

onde C é a constante de integração e \ln é o logaritmo neperiano. Uma solução para a constante C seria assumir que o estímulo percebido torna-se zero para um limiar de estímulo S_0 . Portanto, sendo $\psi = 0$ e $S = S_0$, temos:

$$C = -k \ln S_0. \quad (2.5)$$

A substituição de C na [Equação 2.4](#) resulta na expressão final da Lei de Fechner:

$$\psi = k \ln \frac{S}{S_0}. \quad (2.6)$$

A construção do “edifício psicofísico” de Fechner é alicerçada na constante k , que é determinada conforme o sentido e o tipo de estímulo ([FECHNER, 1907](#)).

No estreitamento do escopo da psicofísica para o sentido da audição, ou seja, na *psicoacústica*, uma aproximação notável para a percepção de intensidade é expressa na lei de potência de S. S. [Stevens \(1957\)](#), que relê a [Equação 2.6](#) na esteira de que razões iguais entre estímulos sonoros tendem a produzir razões iguais de sensação. O grande salto dado pela proposição de Stevens deu-se no estabelecimento de um segundo método dentre os dois comumente usados para o estabelecimento de uma escala de *loudness*. No primeiro método, usado por Fechner e denominado *estimação de magnitude*, sons de diferentes níveis são apresentados e o voluntário é solicitado a designar um número para cada, conforme a intensidade percebida. No segundo método, usado por Stevens e denominado *produção de magnitude*, o voluntário é solicitado a ajustar o nível de um som de teste até que este atinja um *loudness* específico; seja em termos absolutos, seja relativo a um padrão do tipo duas, quatro vezes mais intenso, por exemplo ([MOORE, 2012](#), p. 137). Com efeito, Stevens refuta Fechner:

A lei logarítmica de Fechner não é observada experimentalmente pela simples razão de que a Diferença no Limite do Observável (JND) não é uma constante em unidades psicológicas, mas sim uma aproximação grosseira da magnitude psicológica. Por esta razão, todos os procedimentos de extração Fechneriana, como o método de comparação de pares e técnicas afins, que buscam construir escalas a partir de medidas “unitizadas” de dispersão, não são métodos adequados para uma escala de magnitudes de continuidade Classe I¹ ou protética ([STEVENS, 1957](#), p.178, tradução minha).

Stevens conclui que uma sensação de *loudness* ψ é proporcional ao estímulo S elevado a uma potência n da forma

$$\psi = kS^n, \quad (2.7)$$

¹ Em seu artigo *On the Psychophysical Law*, Stevens define uma escala de Classe I como sendo, dentre outros critérios, quantitativa, na qual a JND incrementa em tamanho subjetivo conforme o incremento da magnitude psicológica.

onde n é uma potência empiricamente testada em relação à categoria do estímulo, sendo k uma constante de proporcionalidade. A potência n pode ser reescrita da forma:

$$n = \frac{\log \psi - \log k}{\log S}. \quad (2.8)$$

Esta correspondência entre magnitude física e magnitude percebida foi uma das motivações para o uso da escala *bel* de razões logarítmicas, ou *decibel*, ser adaptada para as grandezas de pressão sonora compreendidas entre os limiares de percepção (20 μPa ou 0 dB SPL) e de dor (20 Pa ou 120 dB SPL). A sensação de *loudness* para sons de intensidade superiores a 40 decibels referentes a um Nível de Pressão Sonora (SPL) de 20 microPascal (20 μPa), ou dB SPL, é uma relação de potência $n = 0,3$ (STEVENS, 1957). Um som dez vezes mais intenso fisicamente do que outro é percebido como tendo o dobro da intensidade, uma vez que $10^{0,3} \approx 2$. Portanto, há uma *compressão* da relação de estímulos para a relação de sensações.

Ao longo de seus experimentos, Stevens (1955) propôs uma escala de intensidade percebida denominada *sone*. A escala *sone* foi desenvolvida em testes nos quais os ouvintes eram solicitados a definir quando o *loudness* de um som dobrou de intensidade. A partir da referência de 1 *sone*, definido como a sensação de *loudness* produzida por um tom senoidal puro de 1 kHz a 40 dB SPL, outro som percebido como sendo duas vezes mais intenso teria uma sensação de *loudness* de 2 *sone*. O próximo nível seria passar do registro da sensação de *loudness* à medida de um nível de *loudness*, salto este relatado na [subseção 2.2.2](#).

2.2.1 Níveis sonoros de intensidade e pressão

Esta subseção faz um breve interlúdio para formalizar a relação entre nível de intensidade sonora e nível de pressão sonora.

A intensidade sonora medida num ponto específico e numa dada direção de propagação, é definida como sendo a taxa média na qual a energia sonora é transmitida através de uma unidade de área perpendicular à direção dada e no ponto especificado (BERANEK, 1954, p. 12). Sua unidade é W/m^2 e, para ondas

planas viajantes, pode ser escrita da forma:

$$I = \frac{p_{rms}^2}{\rho_0 c}, \quad (2.9)$$

onde p_{rms} é a pressão sonora efetiva² (N/m^2) e $\rho_0 c$ é a impedância característica do meio em *rayls* ($\text{N} \cdot \text{s}/\text{m}^3$) que é igual ao produto da densidade característica do meio (ρ_0) pela velocidade do som (c).

Já o nível de pressão sonora, em decibels referentes à pressão sonora de $20\mu\text{ Pa}$ (dBSPL), é vinte vezes o logaritmo na base dez da razão entre a pressão sonora efetiva em relação a uma pressão sonora efetiva de referência:

$$SPL = 20 \log_{10} \frac{p_{rms}}{p_{ref}}, \quad (2.10)$$

onde a pressão sonora de referência $p_{ref} = 20\mu\text{Pa}$ RMS.

O Nível de Intensidade (IL), em dBSPL, é de dez vezes o logaritmo da base dez da razão entre a intensidade do som e uma intensidade de referência, isto é:

$$IL = 10 \log_{10} \frac{I}{I_{ref}}, \quad (2.11)$$

onde a intensidade de referência $I_{ref} = 10^{-12} \text{ W}/\text{m}^2$.

Logo, a relação exata entre nível de intensidade sonora e nível de pressão sonora para ondas planas viajantes pode ser encontrada substituindo a [Equação 2.9](#) na [Equação 2.11](#):

$$IL = 10 \log_{10} \left(\frac{p_{rms}^2}{\rho_0 c I_{ref}} \right) \quad (2.12)$$

$$= 10 \log_{10} \left(\frac{p_{rms}^2}{p_{ref}^2} \cdot \frac{p_{ref}^2}{\rho_0 c I_{ref}} \right) \quad (2.13)$$

$$= 20 \log_{10} \frac{p_{rms}}{p_{ref}} + 10 \log_{10} \frac{p_{ref}^2}{\rho_0 c I_{ref}} \quad (2.14)$$

$$= SPL + 10 \log_{10} \frac{p_{ref}^2}{\rho_0 c I_{ref}}. \quad (2.15)$$

² Para ondas planas viajantes, a densidade média de energia é $\frac{p_{rms}^2}{\rho_0 c^2}$ [J/m^3] (BERANEK; MELLOW, 2012, p. 14)

Substituindo os valores de referência $p_{ref} = 20\mu\text{Pa RMS}$ e $I_{ref} = 10^{-12} \text{ W/m}^2$:

$$IL = SPL + 10 \log_{10} \frac{400}{\rho_0 c} \text{ dB.} \quad (2.16)$$

Para valores de temperatura e pressão $T = 22^\circ \text{ C}$ e $P_0 = 10^5 \text{ Pa}$, $\rho_0 c = 407 \text{ rayls}$ (BERANEK; MELLOW, 2012, p. 14). Nestas condições, a diferença entre o nível de pressão sonora e o nível de intensidade sonora é de 0,1 dB. Portanto, não havendo efeitos de refração entre os valores medidos e os valores de referência, os níveis de pressão sonora e de intensidade sonora serão considerados equivalentes ao longo do texto.

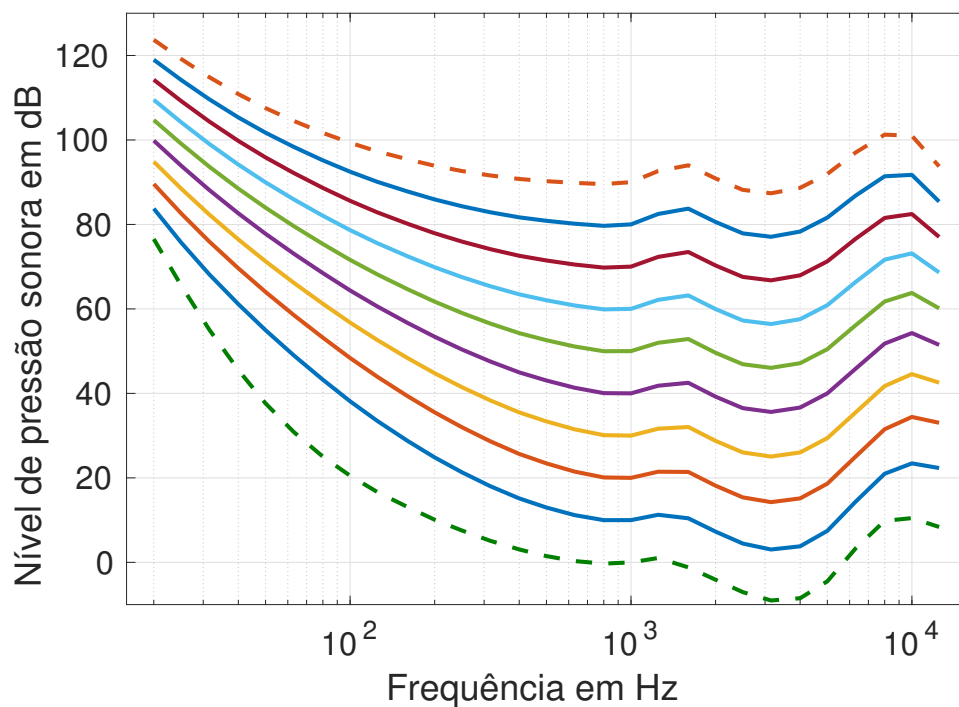
2.2.2 Nível de Loudness

O conceito de nível de *loudness* foi introduzido por H. Barkhausen (1926) ao propor um procedimento de medição de ruído baseado em características da audição humana. Para tanto, utilizou métodos psicofísicos de comparação de *loudness* que levaram à criação da escala *phon* (FASTL, 2010). O nível de *loudness* de um dado som é definido como sendo o nível de pressão sonora numa onda plana de incidência frontal gerada por um tom de referência de 1 kHz percebida de modo tão intenso quanto o som avaliado o é. Desta forma, o valor em *phon* é definido por uma curva no eixo de frequências que compreende as intensidades físicas de outros tons percebidos de modo tão intenso quanto o tom de referência de 1 kHz. Por exemplo, a curva de 20 *phon* é a curva de níveis de pressão sonora percebidos com a mesma intensidade de um tom de 1 kHz a 20 dBSPL.

Esta equalização auditiva foi catalogada nos trabalhos pioneiros de Fletcher e Munson (1933), responsáveis pelos primeiros testes subjetivos para registro das curvas *phon*, ou dos *contornos de mesmo loudness*. Estes contornos, reproduzidos na Figura 2.4, foram refinados ao longo dos anos por meio de levantamentos conduzidos por vários laboratórios que culminaram no padrão de número 226 da ISO (2014).

Num exame da Figura 2.4, quatro aspectos são dignos de nota:

1. Os contornos de níveis de *loudness* menos intensos são quase paralelos ao limiar de silêncio;

Figura 2.4 – Contornos de mesmo *loudness* para sons binauriculares de direção frontal.

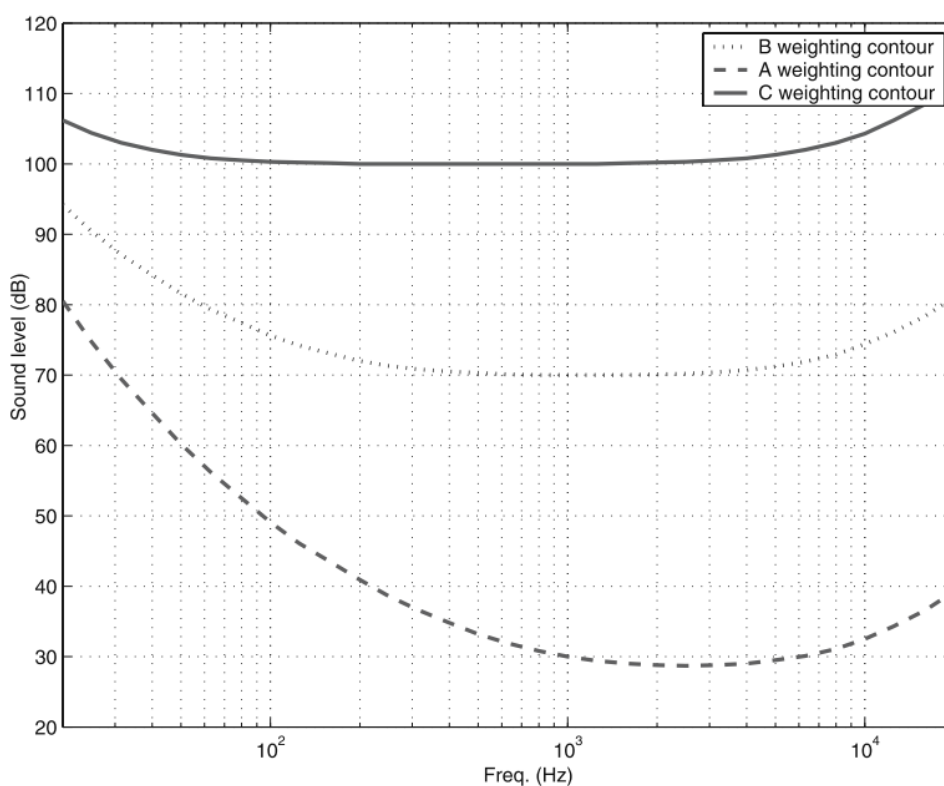
Fonte: Adaptada de ISO (2014, Interpolação da tabela de pares (frequência, nível de pressão sonora)).

2. Em baixas frequências, os contornos têm menores variações em dB SPL conforme o incremento dos níveis de *loudness*;
3. A diferença entre os valores em *phon*, numa única curva de mesmo *loudness*, e os valores em dB, no eixo das ordenadas, chega a 30 em níveis baixos e cai para 10 em níveis mais altos de *loudness* em *phons* (ZWICKER; FASTL, 2013);
4. A faixa de frequências compreendida entre 2 e 5 kHz, que corresponde a um vale em todos os contornos de mesmo *loudness*, contém as frequências de ressonância para cavidades com dimensões similares ao meato acústico, como visto na seção 2.1.

Os modelos de predição de *loudness* serão aprofundados no Capítulo 3, mas cabe aqui antecipar que as ditas *curvas de ponderação em frequência*, mencionadas na seção 1.2, são aproximações de contornos de mesmo *loudness*. Em baixos níveis sonoros, componentes de baixa frequência contribuem pouco para o *loudness* total de um som complexo e a ponderação pela curva “A”, uma aproximação do contorno de mesmo *loudness* de 40 phon, é utilizada para reduzir

a contribuição das baixas frequências na leitura final do medidor. Em níveis altos, todas as frequências contribuem de maneira mais equânime para a sensação de *loudness*, dado que os contornos de mesmo *loudness* tornam-se mais planos, e a ponderação pela curva “C”, uma aproximação do contorno de mesmo *loudness* de 100 phon, é utilizada. Já a curva “B” é utilizada para níveis intermediários, pois é uma aproximação do contorno de 70 phon (MOORE, 2012). Os contornos de ponderação equivalentes são ilustrados na Figura 2.5.

Figura 2.5 – As primeiras curvas de ponderação em frequência utilizadas em modelos de *loudness* de faixa única foram denominadas A, B e C.



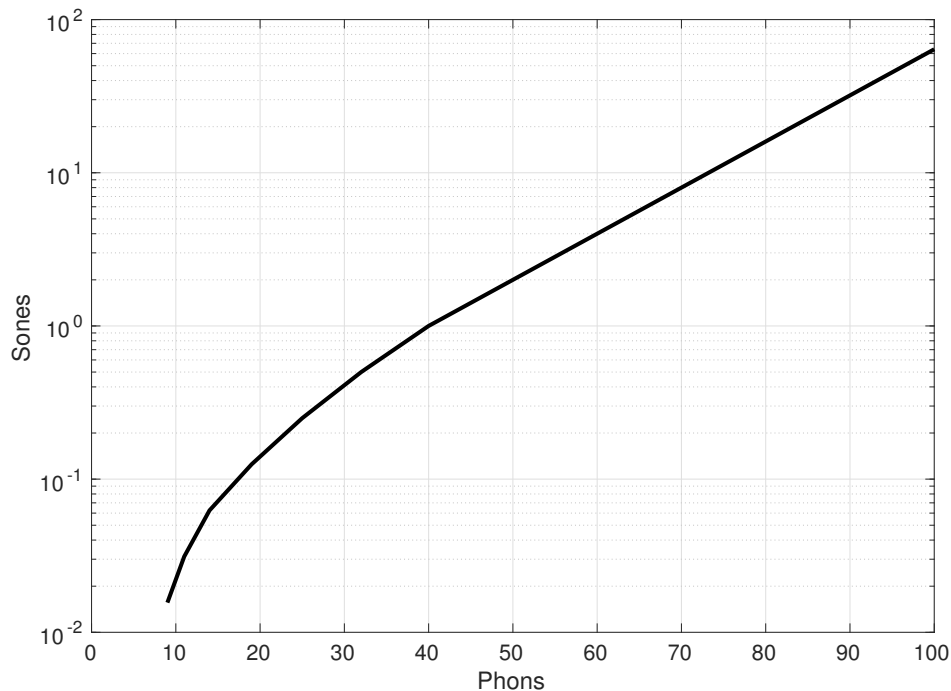
Fonte: Bharitkar e Kyriakakis (2008, p. 68).

Nota – Estes contornos de ponderação são aproximações dos contornos de mesmo *loudness* de 40, 70 e 100 phon, respectivamente.

Como em baixos níveis de *loudness* os contornos de mesmo *loudness* são mais côncavos, as variações de *loudness* conforme a variação do nível sonoro dão-se mais rapidamente do que a Lei de Potência de Stevens para $n = 0,3$. Muito embora as escalas *phon* e *sones* diverjam para níveis de pressão sonora inferiores a 40 dBSPL, como ilustrado na Figura 2.6, isso de forma alguma diminui a importância da Lei de Stevens para a psicoacústica, verificada em muitos dos experimentos de intensidade subjetiva que seguiram (ZWICKER;

FASTL, 2013). Por outro lado, a própria elaboração de escalas de *loudness* é sim discutida, pois as técnicas para sua obtenção são suscetíveis a vieses como as instruções passadas aos voluntários, a variedade de estímulos, as escolhas para resposta e treinamento/atenção dos ouvintes (POULTON, 1979).

Figura 2.6 – Equivalência das escalas sone e phon para um tom senoidal de 1 kHz.



Fonte: Adaptada de ISO (1996, Interpolação da tabela de pares (sone, phon)).

Nota – As escalas divergem em níveis de pressão sonora inferiores a 40 dB SPL

Indo mais longe, Moore (2012) pontua até mesmo o conceito de se pedir a um ouvinte para que julgue a magnitude de uma sensação. Pois se a estimativa de *loudness* de sons do cotidiano é afetada pela distância, pelo contexto e pelo seu significado, ou seja, pela tentativa humana em estimar a caracterização da própria fonte sonora, ater-se somente à magnitude de uma sensação é difícil por si só. Ponto de vista partilhado pelo próprio Helmholtz citado em Warren (1981), cujas palavras fecham diligentemente esta seção:

Nós somos excessivamente bem treinados em encontrar, pelas nossas próprias sensações, a natureza objetiva dos objetos ao nosso redor, porém somos completamente inaptos em observar estas sensações *per se*; e a prática de associá-las a coisas fora de nós mesmos, na verdade, impede que sejamos distintamente conscientes das sensações puras (WARREN, 1981, citação, tradução minha).

2.3 Efeitos Espectrais

A seção anterior tratou da dependência do nível de *loudness* para com a frequência, a partir de experimentos clássicos conduzidos com tons senoidais puros. Todavia, para o caso de sons complexos, aspectos de resolução da análise espectral realizada pela cóclea não devem ser ignorados, a exemplo de dois tons senoidais que, quando próximos em frequência, podem ter algum grau de sobreposição das suas áreas vibrantes na membrana basilar. Se considerarmos o aumento da distância entre as posições de máxima ressonância da membrana basilar com a frequência (ver [Figura 2.2](#)), é razoável presumir que a resolução em frequência também impacta a percepção de intensidade sonora.

Quanto maior a proximidade entre dois tons senoidais, maior é a sobreposição das áreas vibrantes até um dado limite de seletividade em frequência, no qual os tons não mais são distinguidos individualmente. Este efeito no qual o limiar de audibilidade de um som aumenta na presença de outro som é denominado *mascaramento* e é ilustrado na [Figura 2.7](#).

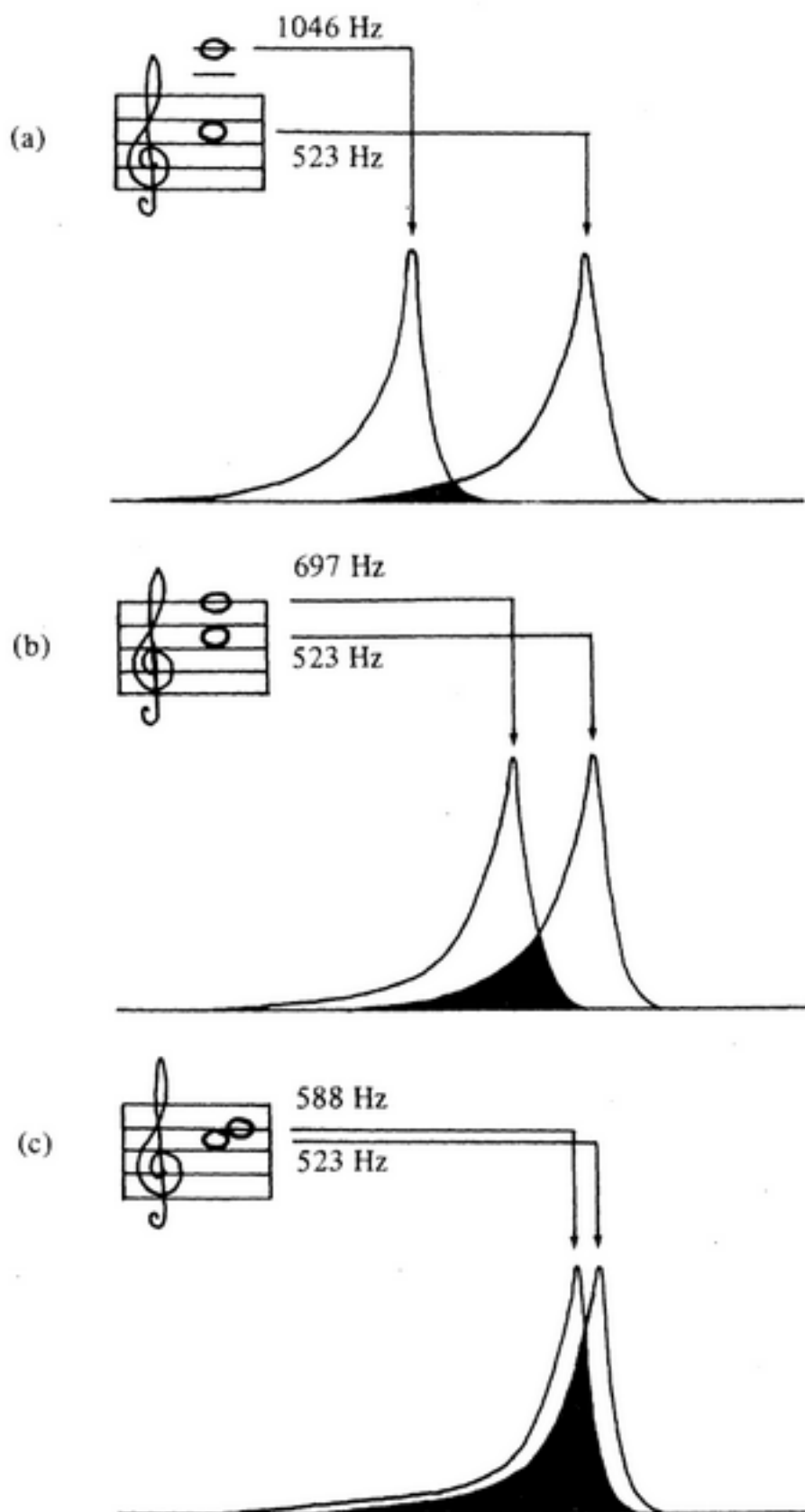
2.3.1 Acumulação espectral de loudness

Considerando não mais dois tons senoidais, e sim um tom e um Ruído de Excitação Uniforme (UEN), se a largura de banda deste ruído aumenta, o limiar de audibilidade do tom mascarado por este ruído também cresce. Contudo, esta taxa de variação tende a diminuir quando a banda do ruído é alargada para além de um dado limite, como se os limites de integração da potência deste ruído fossem restritos a esta largura de faixa “crítica” e a energia fora dela fosse descartada. Este efeito causado pelo alargamento da faixa de um UEN foi observado originalmente por Harvey [Fletcher \(1940\)](#), que formulou o seguinte postulado:

Para este tipo de ruído [limitado em banda], a largura de banda crítica em ciclos é numericamente igual à razão da intensidade do tom mascarado pela intensidade média por ciclo do ruído que produziu o mascaramento. ([FLETCHER, 1940](#), p. 56, tradução minha)

[Zwicker, Flottorp e Stevens \(1957\)](#) refinaram os experimentos de Fletcher para de fato medir as bandas críticas da audição, com resultados significativos

Figura 2.7 – Conforme dois tons se aproximam em frequência, cresce a sobreposição de suas áreas de vibração na membrana basilar.



Fonte: Campbell e Greated (1994, p. 58).

Nota – Este efeito no qual o limiar de audibilidade de um som aumenta na presença de outro som é denominado *mascaramento*.

para a percepção de intensidade: se a largura de banda do UEN aumenta com seu nível permanecendo constante, o *loudness* permanecerá constante até a largura de banda crítica e crescerá em seguida. Zwicker (1961) publicou – no formato de tabela – larguras médias de bandas críticas obtidas por diversos experimentos à época e, posteriormente, Zwicker e Terhardt (1980) publicaram uma aproximação numérica dos dados tal como na Equação 2.17.

$$\Delta f_c = 25 + 75 \left(1 + 1,4(f_c/1000)^2 \right)^{0,69}, \quad (2.17)$$

onde Δf_c é a largura de banda crítica associada à frequência central f_c .

As bandas críticas de Zwicker são de aproximadamente 100 Hz para frequências inferiores a 500 Hz e depois crescem linearmente numa escala de $0,2f$. Com base nesta percepção, Zwicker e Terhardt (1980) propuseram uma escala de 24 bandas críticas com as frequências centrais tabeladas em (ZWICKER, 1961) para cobrir toda a faixa de audição. A escala foi assim chamada de *bark* em homenagem aos experimentos de intensidade sonora feitos por H. Barkhausen (1926). A função de conversão de hertz para bark é descrita pela Equação 2.18:

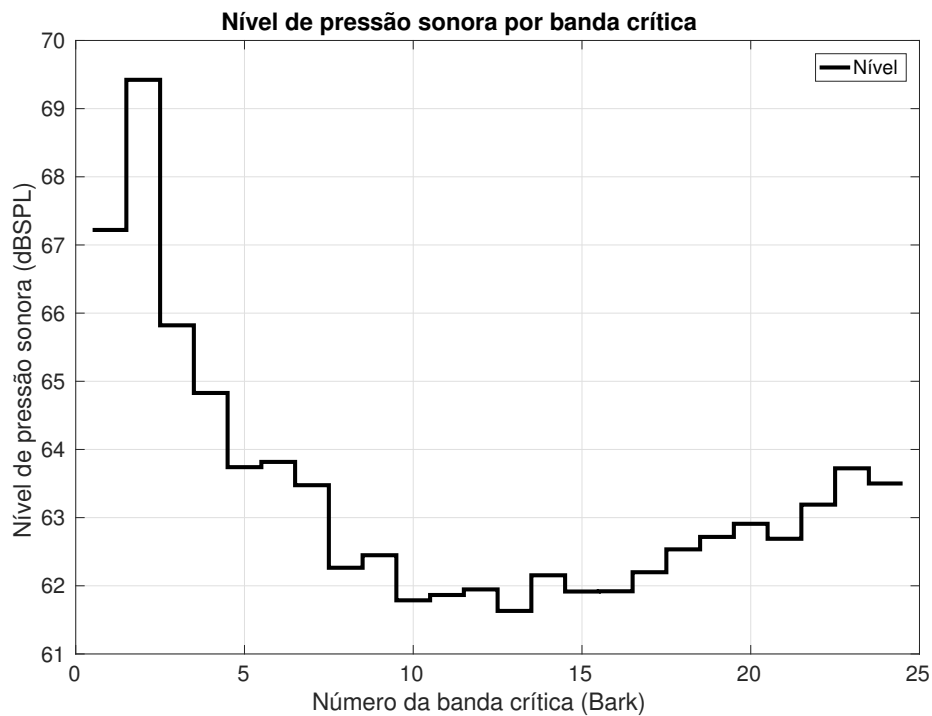
$$z(f)[bark] = 13 \arctg(0,76(f/1000)) + 3,5 \arctg\left(\frac{f/1000}{7,5}\right)^2, \quad (2.18)$$

onde f é a frequência em Hertz e $z(f)$ é a frequência em barks. Um exemplo de distribuição de níveis de pressão sonora pelas bandas críticas de Zwicker e Terhardt (1980) na escala bark é ilustrado na Figura 2.8.

Muito embora os experimentos de Fletcher (1940) tenham obtido sucesso em medir as larguras de banda crítica, o próprio tinha conhecimento de que os filtros auditivos não eram retangulares, e sim com um topo arredondado e decaimentos nas bordas (MOORE, 2012, p. 71). A detecção do formato dos filtros auditivos foi feita por Roy D. Patterson (1976) que utilizou a abordagem do ruído dentado (*notched*), ilustrada na Figura 2.9. Um método bastante sagaz, explicado pelas palavras do autor:

No modelo de detecção mais simples, quando o filtro é suposto como sendo centrado no tom e o mascarador é um ruído passa-baixas, a área na qual o ruído e o filtro se sobrepõem representa o ruído que efetivamente mascara o tom. Quando o mascarador é o mesmo ruído

Figura 2.8 – Distribuição do nível de pressão sonora pelas bandas críticas na escala Bark para um ruído rosa de 40 dB/Hz @ 1 kHz.



Fonte: Elaborada pelo autor.

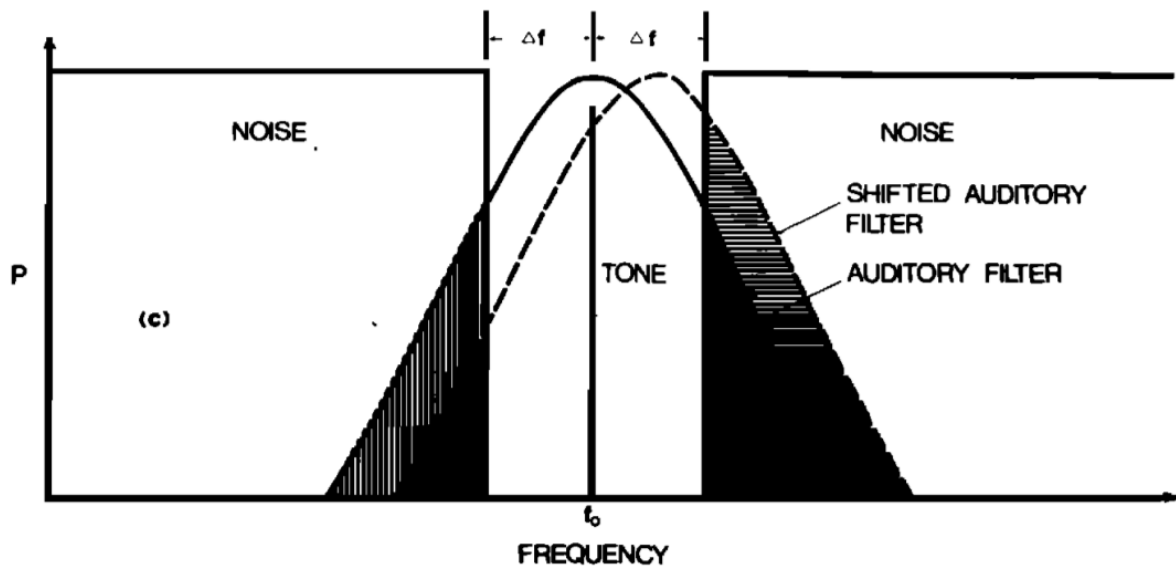
passa-baixas e o filtro é deslocado, a detecção melhora nos casos em que o deslocamento produz mais redução de ruído do que redução de sinal. Já quando o mascarador é um ruído dentado, deslocar o filtro produz um mínimo, senão nenhum, melhoramento na detecção, porque a redução de ruído de um lado do tom é acompanhada por um aumento de ruído do outro lado. (PATTERSON, 1976, p. 643, tradução minha)

A partir dos resultados de Patterson, Glasberg e Moore (1990) sugeriram uma aproximação para as larguras de banda dos filtros auditivos inteiramente baseada em mascaramentos por ruído dentado, denominada Largura de Faixa Retangular Equivalente (ERB). As expressões de largura de banda $BW_{ERB}(f)$ e de conversão de hertz para a escala $ERB_N(f)$ são escritas nas formas da Equação 2.19 e da Equação 2.20, respectivamente.

$$BW_{ERB}(f_c)[Hz] = 24,7(4,37(f_c/1000) + 1) \quad (2.19)$$

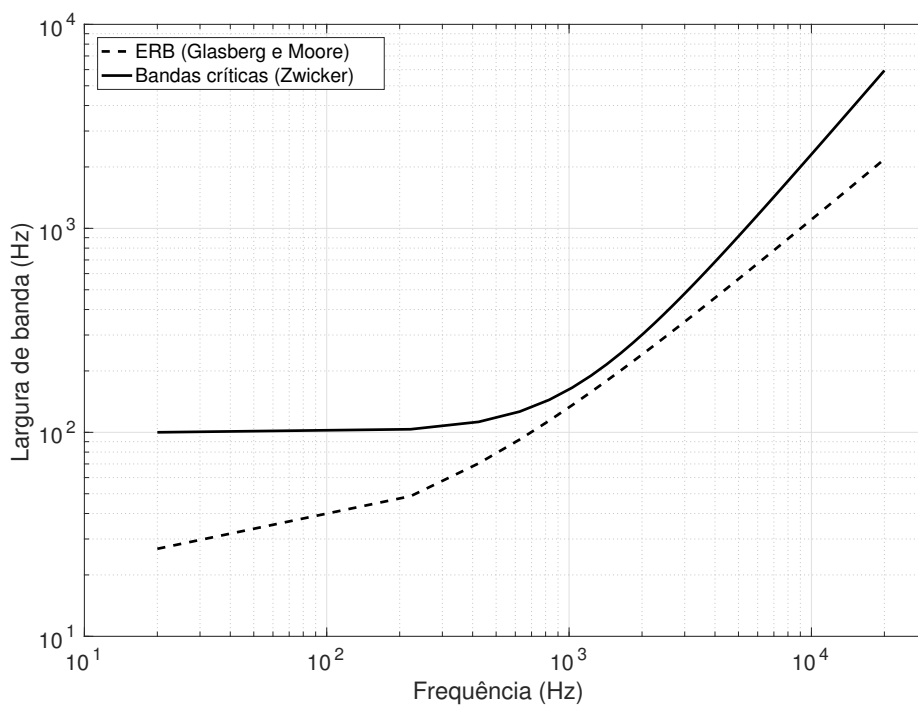
$$ERB_N(f)[ERB] = 21,4\log_{10}(4,37f + 1). \quad (2.20)$$

Figura 2.9 – Determinação do formato de um filtro auditivo pela abordagem do ruído dentado em [Patterson \(1976\)](#).



Fonte: [Patterson \(1976, Figura 2\(c\)\)](#).

Figura 2.10 – Comparação entre as bandas críticas de [Zwicker \(1961\)](#) e as ERBs de [Glasberg e Moore \(1990\)](#).



Fonte: Elaborada pelo autor.

A [Figura 2.10](#) compara as bandas críticas de [Zwicker \(1961\)](#) com as ERBs de [Glasberg e Moore \(1990\)](#). Nota-se que as últimas são mais estreitas em todas as frequências centrais e, ao contrário das primeiras, continuam estreitando em frequências inferiores a 500 Hz, ou seja, filtros auditivos baseados em ERBs possuem melhor resolução em baixas frequências do que os filtros auditivos baseados em bandas críticas.

A modelagem dos filtros auditivos foi de suma importância para o primeiro estágio de processamento auditivo ilustrado na [Figura 2.3](#). As curvas de ponderação construídas a partir de aproximações dos contornos de mesmo *loudness* e os bancos de filtros projetados a partir de larguras de banda crítica e de ERBs são as principais estratégias de tratamento dos efeitos de frequência aplicadas pelos modelos de *loudness* que serão apresentados no [Capítulo 3](#).

2.4 Efeitos Temporais

Assim como em função da frequência, o *loudness* em função do tempo não está limitado a uma percepção única, especialmente ao se considerar sons ditos *não-estacionários*, cujas características físicas de intensidade e frequência são variantes no tempo. Para um tom de 1 kHz cujo nível de pressão sonora varie por alguns segundos, por exemplo, como sua intensidade é percebida? Supondo variações não abruptas de nível físico, a sensação de *loudness* é *dinâmica* e também variante no tempo.

2.4.1 Integração temporal de loudness

Após [Fletcher e Munson \(1933\)](#) concluírem o levantamento das curvas de mesmo *loudness*, [H. Fletcher \(1940\)](#) concentrou-se nos efeitos de frequência e [W. A. Munson \(1947\)](#) investigou os efeitos temporais na percepção de intensidade, como se movidos por um dever para com a comunidade científica naqueles anos conturbados:

Com o objetivo de se obter mais dados para o estudo da dependência do *loudness* em relação à duração do estímulo, um programa de testes teve

início anos atrás³ mas, devido a interrupções, nunca foi concluído. Alguns resultados foram obtidos, contudo, e estes foram disponibilizados a projetos governamentais durante os anos da guerra. Embora novos testes nunca tenham sido conduzidos, os dados foram considerados de interesse geral para garantir uma publicação no presente formato (MUNSON, 1947, p. 585, tradução minha).

E, de fato, foram. Os experimentos utilizaram sons cuja duração variou de 5 a 200 ms com observações de que uma senóide com duração de 10 ms seria percebida de modo menos intenso do que outra senóide de mesmo nível de pressão sonora, mas com duração de 100 ms, por exemplo. Seus resultados levaram à intuição de que uma predição de *loudness* poderia seguir a relação:

$$N = N_{max}E_s(t), \quad (2.21)$$

onde N é o nível de *loudness* no instante de tempo em que o tom é encerrado, N_{max} é o nível máximo de *loudness* atingido se o tom persistisse indefinidamente, e $E_s(t)$ seria a “integral de sensação”, uma função para modelar o aumento na intensidade nos primeiros 200 ms que caracterizaria o fenômeno como uma *integração temporal de loudness*.

Décadas depois, uma extensa bateria de testes conduzidos por S. D. G. Stephens (1973) ratificou a intuição de Munson, concluindo que a integração temporal não era monotônica e o tempo de integração é dependente do nível físico, sendo maior para níveis intermediários:

No decurso destes experimentos, foram apresentados dois efeitos distintos de intensidade global num padrão de integração temporal para relações constantes de sinal/ruído. Um diz respeito ao efeito de sons de curtíssima duração nos quais sua detectabilidade relativa é menor em níveis intermediários do que em níveis altos ou baixos. O outro concerne a um aparente encurtamento da duração crítica conforme o aumento dos níveis de intensidade. (STEPHENS, 1973, p. 123, tradução minha)

A primeira afirmação de Stephens foi confirmada num estudo mais recente de Buus, Florentine e Poulsen (1997) que testou tons de 5 kHz com durações de 2 a 250 ms e com níveis físicos variando de 2 a 60 dB SPL. Os autores observaram

³ Munson se refere aqui aos experimentos para levantamento das curvas de mesmo *loudness* realizados no início dos anos 30, cujos voluntários foram recrutados das forças armadas norte-americanas (FLETCHER; MUNSON, 1933).

uma diferença de aproximadamente três vezes entre os tempos de integração de *loudness* de tons com níveis intermediários e os tempos de integração de tons com níveis baixos/altos:

A quantidade de integração temporal, definida como sendo a diferença de nível entre tons curtos e longos igualmente intensos, depende significativamente do nível. Com os tons muito breves usados no presente estudo, esta diferença variou em mais de 25 dB quando o tom de 250 ms era em torno de 22 dBSPL e o tom de 2 ms era em torno de 63 dBSPL. Para todos os pares de duração, a quantidade de integração temporal é por volta de três vezes maior em níveis moderados do que em níveis baixos ou altos. A JND do nível de *loudness* também varia com o nível e é maior em níveis moderados. (BUUS; FLORENTINE; POULSEN, 1997, p. 678, tradução minha)

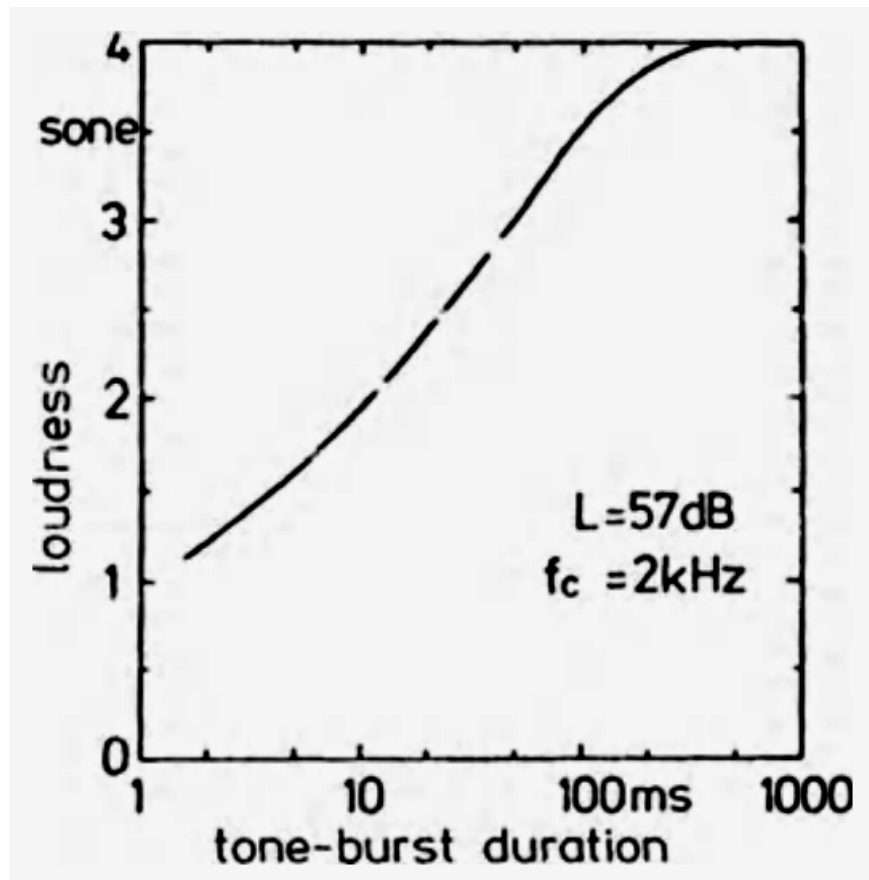
Já com relação ao encurtamento da duração crítica conforme os níveis de intensidade, é preciso cautela. Uma leitura pouco atenta da obra definitiva de Zwicker e Fastl (2013) pode levar ao erro de que 100 ms seria o ponto de ruptura a partir do qual a sensação de *loudness* é constante, especialmente no que se refere às interpretações da Figura 2.11 aqui reproduzida. Afirmações como “O *loudness* de um tom persistente decresce para durações inferiores a 100 ms. Para durações mais longas, o *loudness* é quase independente da duração” e “Acima de 100 ms, o nível de *loudness* é aproximadamente independente da duração” (ZWICKER; FASTL, 2013, p. 216–217, grifos meus), devem ser lidas num contexto de tons senoidais cujos níveis de pressão sonora não foram avaliados nos experimentos.

Tanto não há um mapeamento direto entre níveis de pressão sonora e tempos de integração de *loudness*, que Bertram Scharf (1978) fez um extenso levantamento dos principais estudos de integração temporal até aquele momento para demonstrar que não há consenso no que se refere ao efeito de nível na duração crítica, como pode ser visto na Tabela 2.1.

Na tabela, a coluna “Relação de Compromisso” refere-se à relação entre intensidade e tempo. Enquanto a duração de um estímulo cresce até atingir a duração crítica, para manter o *loudness* constante, a energia total do som ($I \times t$) foi encontrada como sendo constante⁴, decrescente ou crescente. Já a

⁴ Posteriormente verificou-se que esta relação do limiar de intensidade I com a duração t do tom está incorreta, pois desta forma o limiar dependeria somente da quantidade total de energia no estímulo e

Figura 2.11 – Loudness de um tom persistente de 2 kHz com 57 dB SPL como uma função da duração.



Fonte: Zwicker e Fastl (2013, Figura 8.12.).

coluna “Duração Crítica” está associada ao tempo em que, para uma sensação de *loudness* constante, a intensidade deve ser reduzida enquanto a duração aumenta (SCHARF, 1978).

Mesmo que ainda não se conheça uma função de integração temporal de *loudness*, a noção de que o sistema auditivo possui um retardo de percepção de intensidade leva a um problema de mascaramento distinto daquele visto na seção 2.3. Dados dois tons próximos no tempo, este *mascaramento temporal* ocorre quando o tom mais intenso impacta a percepção do tom menos intenso, seja o tom mascarado seguido do tom mascarador (*a posteriori*) ou precedido por ele (*a priori*). Zwicker e Fastl (2013) ilustraram este efeito temporal *a posteriori* na Figura 2.12.

A sensação de *loudness* do tom de teste da Figura 2.12 numa condição

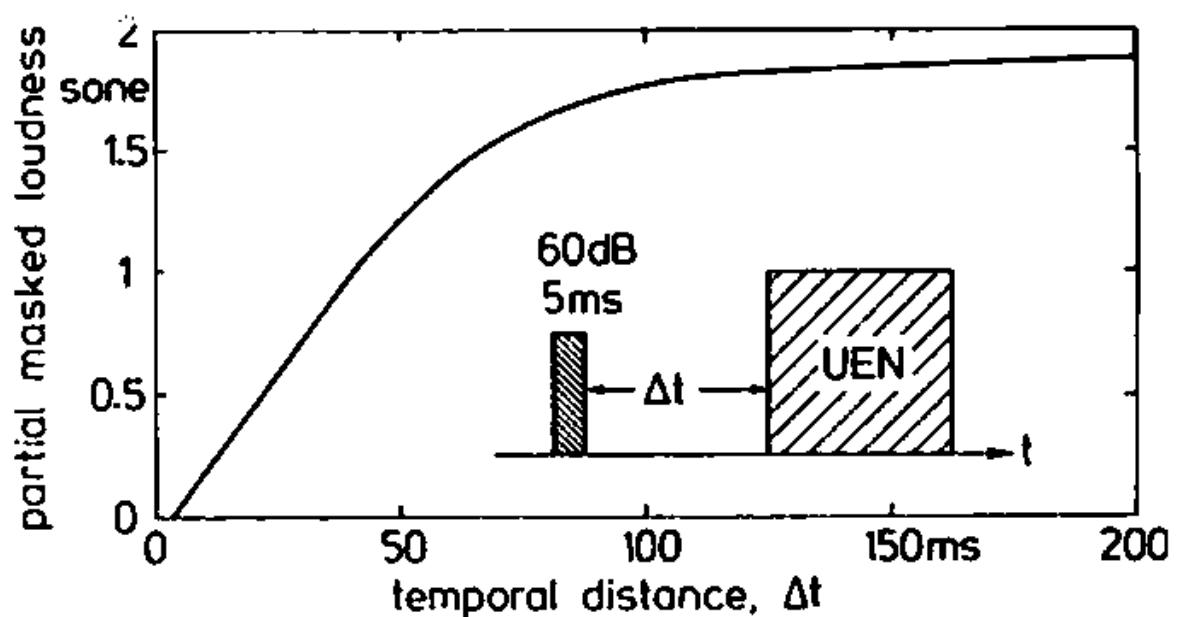
não de como essa energia estaria distribuída no tempo. Uma expressão alternativa para esta relação seria: $(I - I_L) \times t = \text{constante}$, sendo I_L o limiar de intensidade para um estímulo de longa duração (MOORE, 2012).

Tabela 2.1 – Levantamento de estudos sobre o efeito do nível *loudness* na Duração Crítica (DC).

Autor e ano	Participantes	Relação de Compromisso	Duração Crítica (ms)	Efeito de nível
Miller e Taylor (1948)	3	Energia ↑	60–140	DC ↓ Nível ↑
Pollack (1958)	7–10	Energia constante.	100	–
Small, Brandt e Cox (1962)	12	Energia ↓	15–50	DC ↓ Nível ↑
Stevens e Hall (1966)	12	Energia ↓	150	nenhum
Zwislocki (1966)	83	Energia cte.	200–400	–
	74	Energia constante.	200–400	–
Békésy (1929)	–	Energia ↑	120–180	DC ↓ em níveis altos
Ekman e Berglund (1966)	10	Energia ↑	>500	DC ↓ em níveis altos
Garner (1949)	6	Energia ↑	500	DC ↓ em níveis altos
Munson (1947)	–	Energia ↓	200	DC ↓ em níveis altos
Niese (1956)	12	Energia constante.	65	–
Niese (1959)	10	Energia ↑	100	nenhum
Pedersen e Lyregaar (1972)	300	Energia constante.	160–320	nenhum
Reichardt e Niese (1970)	50	Energia constante.	100	–
Port (1963)	8	Energia constante.	70	nenhum

Fonte: Adaptada de Scharf (1978, p. 205–206).

Figura 2.12 – Mascaramento temporal de *loudness* de um tom de 2 kHz, 60 dB SPL e 5 ms, ocorrido antes de um ruído de excitação uniforme (UEN) com uma diferença de tempo Δt .



Fonte: Zwicker e Fastl (2013, Figura 8.15.).

sem mascaramento é de 1,9 *sones*, que é atingida em valores de Δt próximos de 200 ms. O impacto na percepção de intensidade é mais pronunciado quando $\Delta t < 100$ ms, e a sensação cai para zero quando $\Delta t = 5$ ms.

Os efeitos de *mascaramento temporal* e de *integração temporal de loudness* aqui apresentados são levados em consideração pelos modelos de predição de *loudness* existentes para percepções globais, de curta duração ou momentâneas. Isso é feito pela escolha de limites de integração não inferiores às durações críticas encontradas experimentalmente na literatura. Esta discussão sobre modelos de *loudness* para sons variantes no tempo continuará no [Capítulo 3](#). Por ora, o fechamento desta seção dar-se-á na fala de E. Zwicker:

(...) o efeito [temporal] indica que nosso sistema auditivo precisa de algum tempo para desenvolver a sensação de *loudness* (...) O desenvolvimento de sensações segue uma regra bem conhecida em muitos outros sistemas: a causa e suas consequências são frequentemente não simultâneas! (ZWICKER; FASTL, 2013, p. 220, tradução minha)

2.5 Efeitos Espaciais

Tanto os aspectos espectrais quanto os temporais de *loudness* detalhados anteriormente têm sentido amplo em sons ambientais ou musicais, cujas características específicas de frequência e nível mudam ao longo do tempo. É preciso considerar que os ouvintes possuem duas orelhas recebendo sinais diferentes, e que cenas sonoras são normalmente compostas por vários objetos sonoros (em movimento, muitas das vezes).

Dessas considerações, surgem duas observações: em primeiro lugar, ondas sonoras provenientes de fontes estáticas ou em movimento que não estejam posicionadas diante do ouvinte induzem diferentes sensações de *loudness* em cada ouvido. E em segundo lugar, ao se tratar de paisagens sonoras compostas por muitos objetos, o *loudness* tanto pode estar relacionado a um objeto, em especial, ou à paisagem como um todo. As considerações levantadas neste parágrafo serão abordadas nas subseções a seguir.

2.5.1 Somatório biauricular de loudness

Para tratar a indução de diferentes sensações de *loudness* em cada ouvido, o sistema auditivo faz um “somatório biauricular de loudness”. É o nome dado ao efeito psicoacústico de um som incidente em ambos os ouvidos ser percebido de modo mais intenso do que se tivesse incidência monoauricular (EPSTEIN; FLORENTINE, 2009). A grande empreitada de Fletcher e Munson (1933) também contemplou o primeiro estudo de *loudness* biauricular ao experimentarem casamentos de tons monóticos (em apenas um ouvido) com tons dióticos (em ambos os ouvidos). Eles encontraram diferenças de 5 dB para níveis de *loudness* de 20 phons e diferenças de 10 dB para níveis de *loudness* superiores a 50 phons, demonstrando a ocorrência de um somatório biauricular. Estas diferenças são conhecidas como *ganho biauricular*. Contudo, os pesquisadores apenas supuseram que a sensação de *loudness* biauricular seria duas vezes mais intensa do que a sensação monoauricular, informação que carecia de mais experimentos para ser alçada ao status de afirmação comprovada.

A ratificação da suposição de Fletcher e Munson deu-se com os estudos de Larry Marks (1978) sobre somatório biauricular do *loudness* de tons puros, nas palavras do autor:

A concordância entre os presentes resultados e os de Fletcher e Munson é muito boa. Vale ressaltar, entretanto, que os dados apresentados fornecem uma evidência empírica mais forte no quesito aditividade do que as reunidas por Fletcher e Munson. (MARKS, 1978, p. 111, tradução minha)

Marks teve a Lei de Potência de Stevens (1957) como ponto de partida, reescrevendo as funções de *loudness* nas formas mono (m) e biauricular (b):

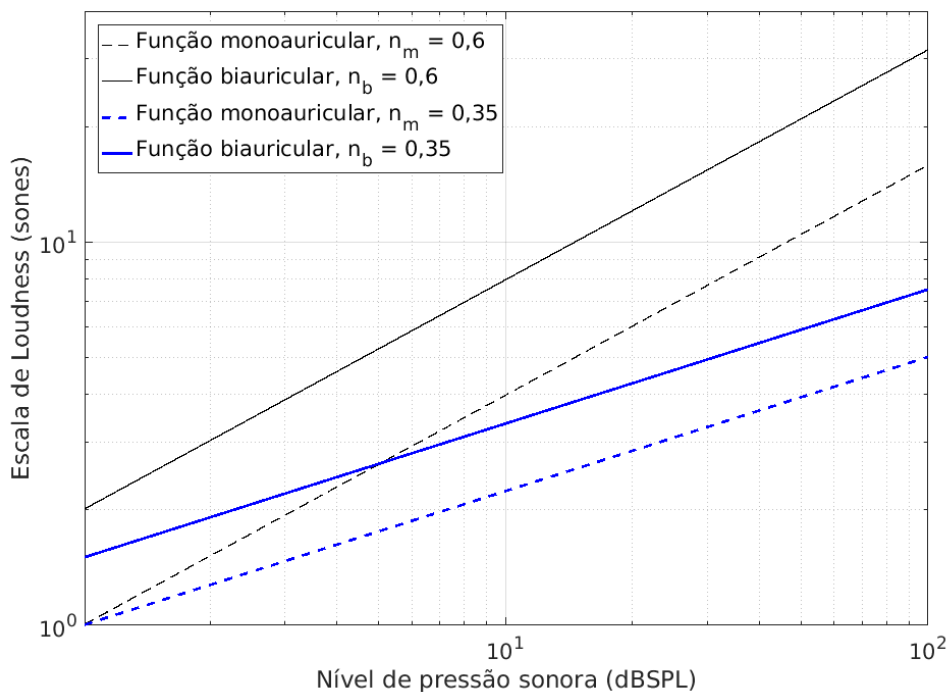
$$\psi_m = k_m S_m^{n_m} \quad (2.22)$$

$$\psi_b = k_b S_b^{n_b} \quad (2.23)$$

Dadas estas equações, o somatório de *loudness* é perfeito para quaisquer níveis de pressão sonora quando $n_m = n_b$ e $k_b/k_m = 2$. Entretanto, isto não foi observado em seus experimentos como ilustrado na Figura 2.13 para potências

$n = 0,6$ e $n = 0,35$. Note que o espaçamento vertical de 10 dB é mantido entre as funções mono e biauriculares de $n = 0,6$, atestando uma relação de dobro da sensação ($\psi_b/\psi_m = 2$) como visto na seção 2.2. E apesar de o espaçamento horizontal entre níveis de pressão sonora ser o mesmo em ambos os pares de curvas, ou seja, o ganho biauricular ser preservado nas curvas inferiores em que $n = 0,35$, nestas a relação entre sensações de *loudness* mono e biauricular é de 1,5 ($\psi_b/\psi_m = 1,5$).

Figura 2.13 – Funções teóricas de *loudness* mono e biauricular em somatórios biauriculares para potências $n = 0,6$ e $n = 0,35$.



Fonte: Adaptada de Marks (1978, Reprodução parcial da Fig. 1).

Para funções de *loudness* em somatórios biauriculares não perfeitos, a potência é calculada em função da relação entre as sensações e do ganho biauricular:

$$n = \frac{20 \log_{10} (\psi_b / \psi_m)}{g}, \quad (2.24)$$

onde g é o ganho biauricular (MARKS, 1978).

Testes subsequentes também com fones de ouvido seguiram esta linha de raciocínio, fixando um dos parâmetros e medindo cuidadosamente os outros dois (EPSTEIN; FLORENTINE, 2009). Em avaliações com alto-falantes, o somatório

biauricular depende da separação dos sons em cada alto-falante no tempo e na frequência. Para separação alguma nem em frequência e nem em tempo, dois alto-falantes produzem uma sensação de *loudness* mais intensa do que apenas um, correspondendo a diferenças de nível de 4 dB. Para separações em frequência superiores a 4 kHz, as diferenças de nível de *loudness* correspondem a 12 dB para baixos níveis de *loudness* (45 phon) e por volta de 8 dB para níveis altos de *loudness* (85 phon) (ZWICKER; FASTL, 2013).

A relação de sensações de *loudness* com e sem somatório biauricular é influenciada tanto pela diferença de fase entre os sinais incidentes quanto pela presença de sinais mascaradores, podendo ser o dobro no melhor caso:

Em suma, sob condições favoráveis, o *loudness* biauricular pode atingir quase duas vezes o valor do *loudness* monoauricular para o mesmo nível físico apresentado ao mesmo tempo (ZWICKER; FASTL, 2013, p. 313, tradução minha).

2.5.2 Loudness em campos sonoros

A radiação oriunda de fontes sonoras acontece idealmente em espaço *livre*, seja em espaços abertos ou em curtas distâncias entre fonte e receptor, no qual o som é recebido em visada direta. Contudo, ondas refletidas em obstáculos e contornos no ambiente circundante ao ouvinte viajam por distâncias superiores a do som direto. Portanto, o som em visada direta chega primeiro ao ouvinte, seguido das primeiras reflexões até um estado de energia reverberante do espaço acústico, composta pelo som refletido inúmeras vezes (BERANEK; MELLOW, 2012, p. 482). Quando o som direto e as primeiras reflexões chegam ao ouvinte por direções distintas, a reverberação é dita *difusa* por não se restringir a nenhuma direção específica.

Campos sonoros livres podem ser reproduzidos em câmaras anecoicas, nas quais a solução da equação da onda esférica para uma fonte isotrópica e uma propagação sem reflexões é dada por:

$$\tilde{p}(r) = \tilde{A}_+ \frac{e^{-jkr}}{r}, \quad (2.25)$$

onde $\tilde{p}(r)$ é a pressão sonora, \tilde{A}_+ é a amplitude complexa da pressão so-

nora irradiada da onda numa distância unitária ao centro da esfera radiante⁵, $k = \omega/c = 2\pi/\lambda$ é o número de onda, e r é a distância do receptor ao centro da esfera (BERANEK, 1954, p. 36). Note que a pressão sonora é inversamente proporcional à distância de propagação ($p \propto 1/r$) e a intensidade, por ser proporcional ao quadrado da pressão sonora (ver Equação 2.9), é inversamente proporcional ao quadrado da distância de propagação ($I \propto 1/r^2$).

Em situações ambientais de campo livre, quando as reflexões são muito menos intensas do que o som direto, a proporcionalidade inversa ao quadrado da distância pode ser considerada como uma aproximação da relação de intensidade sonora. Porém num campo difuso⁶, as reflexões são mais intensas que o som direto e a distribuição de energia é melhor caracterizada pelas reflexões do que pelas linhas de visada, como ilustrado no diagrama de captação da resposta ao impulso de uma sala de estar na Figura 2.14. Note que, nos primeiros milissegundos, quase toda a energia sonora está concentrada nas linhas de visada para os alto-falantes posicionados a $\pm 30^\circ$ à frente. Para tempos superiores a 50 milissegundos, a energia é predominantemente reverberante.

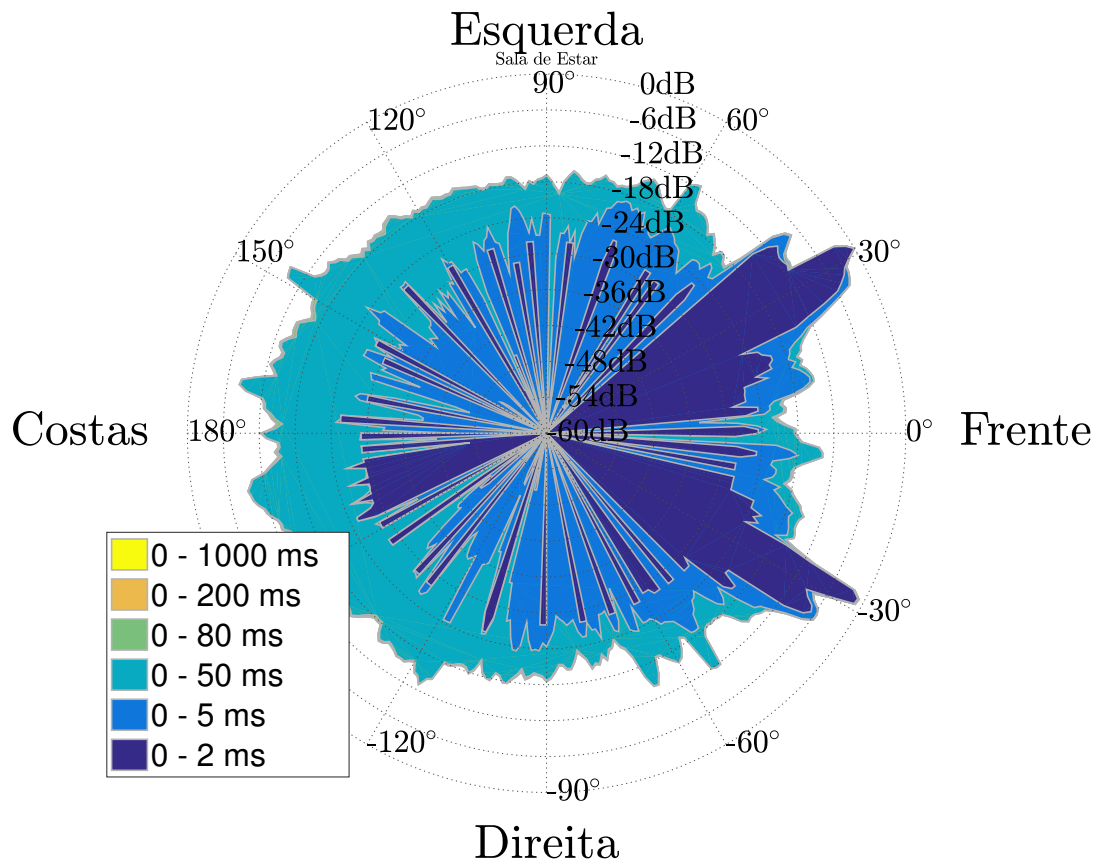
O levantamento dos contornos de mesmo *loudness* de Fletcher e Munson (1933) foi feito considerando uma apresentação em fones de ouvido que produzira o mesmo *loudness* se comparada a uma fonte sonora de frente para o ouvinte em campo livre. Sua primeira revisão – de uma série até se chegar ao padrão 226 da ISO (2014) – foi feita por Robinson e Dadson (1956), que levantaram as curvas utilizando alto-falantes numa câmara anecoica. Com a introdução progressiva de reflexões num trabalho subsequente (ROBINSON; WHITTLE; BOWSHER, 1961), foi possível levantar as atenuações necessárias para obtenção dos contornos de mesmo *loudness* em campo sonoro difuso.

A curva de atenuação a_D para campos difusos, ilustrada originalmente em (ZWICKER; FASTL, 2013, p. 205), foi adaptada na Figura 2.15(a). As medidas que levaram aos resultados expressos na curva a_D , não foram produzidas com tons senoidais, mas com ruídos de banda estreita. Logo, os pontos em frequência

⁵ A amplitude complexa em $r = 1$ é $\tilde{p}(1) = \tilde{A}_+ e^{-jk}$

⁶ Note que “campo” se refere a uma posição do espaço, e não a um único ponto. A rigor, um campo é dito “difuso” quando todas as direções de incidência sonora são equiprováveis em um volume do espaço. Tal condição só pode ser aproximadamente verificada em laboratório nas chamadas “câmaras reverberantes” (RAFAELY, 2000)

Figura 2.14 – Diagrama polar de captação de uma resposta ao impulso de uma sala de estar.



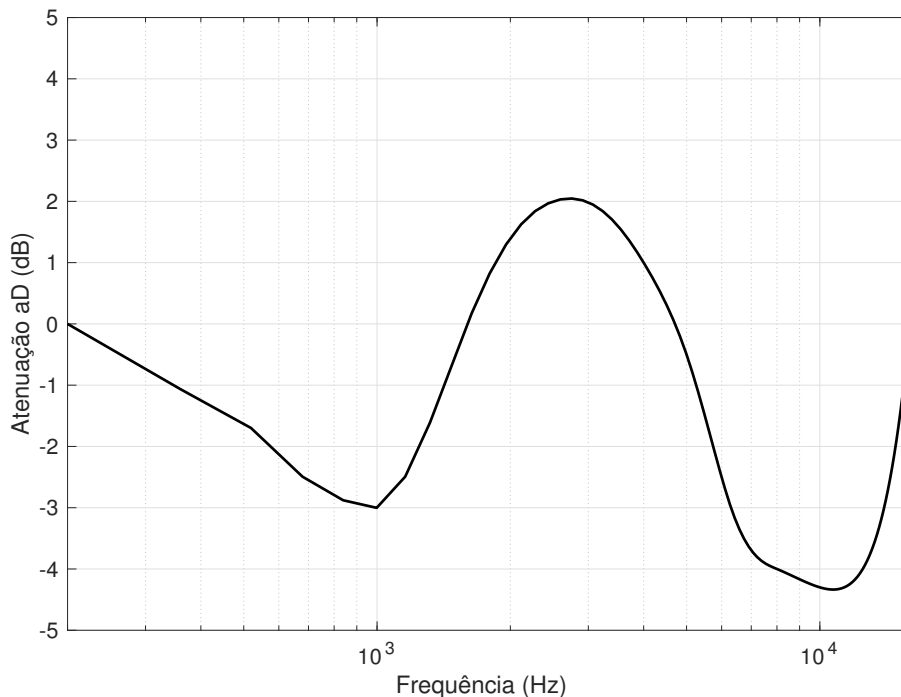
Fonte: Elaborada pelo autor.

Nota – Resposta ao impulso de uma sala de estar capturada com uma sonda de intensidade vetorial G.R.A.S. 50VI posicionada a dois metros de dois alto-falantes ativos, localizados a azimutes de $\pm 30^\circ$ em relação à posição da sonda. Esta representação será explorada na [subseção 4.3.3](#).

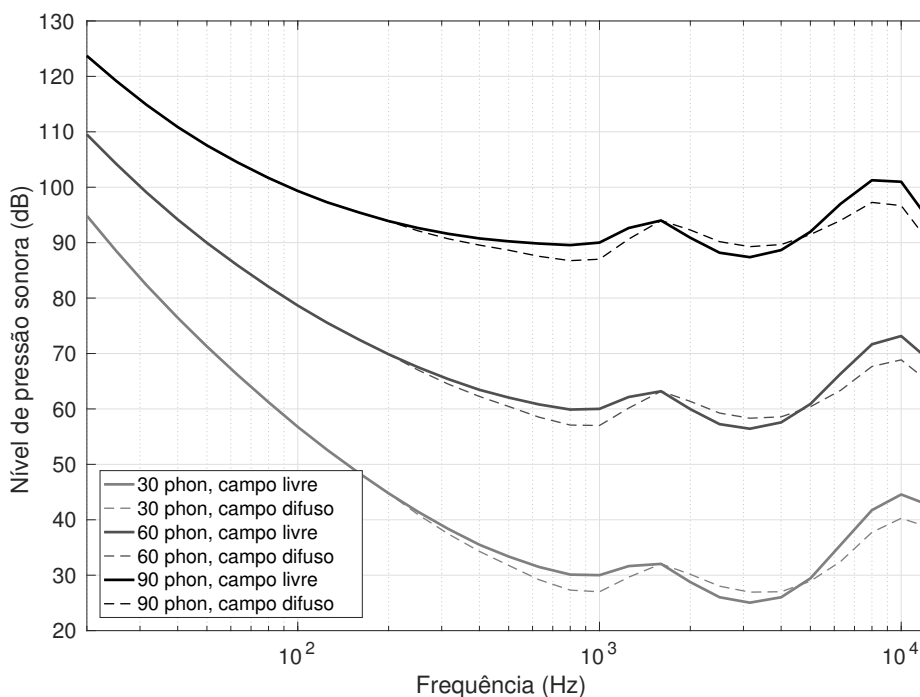
na abcissa são na verdade frequências centrais de UENs (ZWICKER; FASTL, 2013). Nas frequências médias, a atenuação a_D cresce e atinge 2 dB em 2,5 kHz, e em frequências mais altas a atenuação decresce. Os efeitos desta atenuação podem ser observados nos contornos de mesmo *loudness* de 30, 60 e 90 phon em campo livre e em campo difuso ilustrados na [Figura 2.15\(b\)](#). É possível perceber que a diferença entre os contornos em cada campo sonoro aumenta com a frequência, sendo desprezível nas baixas frequências.

Para os casos intermediários em que a energia sonora está distribuída entre a linha de visada e as primeiras reflexões, ou para os casos nos quais há mais de um objeto sonoro, mesmo que as fontes estejam equidistantes do ouvinte e irradiando a mesma pressão sonora, o ângulo de incidência faz com que a pressão sonora nos ouvidos varie consideravelmente. Esta variação de

Figura 2.15 – (a) Atenuação a_D , necessária para produção das curvas de mesmo *loudness* de um tom puro em campo livre para o mesmo tom em campo difuso, em função da frequência do tom. (b) Contornos de mesmo *loudness* de 30, 60 e 90 phon em campo livre e em campo difuso.



Fonte: Adaptada de Zwicker e Fastl (2013, Reprodução da Fig. 8.18, por meio da interpolação das medidas de Robinson e Dadson (1956) nas frequências centrais correspondentes de 1/3 de oitava).



Fonte: Adaptada de Fletcher e Munson (1933, Reprodução dos contornos originais de mesmo *loudness* descontados da atenuação a_D nas frequências centrais correspondentes de 1/3 de oitava).

intensidade conforme a direção da fonte pode ser estimada pelo uso das Funções de Transferência Relativas à Cabeça (HRTFs). As HRTFs representam, para um dado ângulo de incidência, a transferência do som numa linha de visada entre a fonte sonora e um dos ouvidos e modelam a chamada Diferença de Tempo Interaural (ITD) (CHENG; WAKEFIELD, 1999). A cabeça, o torso e o pavilhão auricular atuam como um filtro de resposta dependente da posição da fonte sonora. Pode-se assim determinar um sistema linear invariante no tempo que relaciona um sinal sonoro emitido na ausência do ouvinte com o sinal filtrado pelo ouvinte, medido no interior do canal auditivo (LARA; PASQUAL, 2014). A HRTF é a resposta em frequência desse sistema e sua resposta ao impulso no domínio do tempo é denominada Resposta ao Impulso Relativa à Cabeça (HRIR).

Conjuntos de dados de HRTFs podem ser medidos empiricamente em humanos e manequins, ou calculados a partir de um modelo matemático de aproximação da cabeça via uma solução da equação da onda acústica plana incidente numa esfera rígida. Tal modelo fornece a pressão resultante produzida na superfície da esfera ou, mais precisamente, nos pontos da superfície da esfera que correspondem às localizações dos ouvidos (CHENG; WAKEFIELD, 1999). A aproximação mais conhecida foi a modelada por Brown e Duda (1998)⁷ e sua análise auxilia na compreensão dos fatores envolvidos na intensidade espacialmente percebida.

O modelo de Brown e Duda (1998) é estruturalmente dividido em três partes: sombreamento de cabeça e Diferença Temporal Interauricular (ITD), eco dos ombros, e reflexões do pavilhão auricular. Na primeira parte, a ITD em segundos para um ângulo de chegada de azimute θ , sendo $\theta = 0^\circ$ a incidência frontal, é modelada da forma

$$\text{ITD}(\theta) = \begin{cases} -\frac{a}{c} \cos(\theta), & 0 \leq |\theta| \leq \frac{\pi}{2}, \\ \frac{a}{c} \left(|\theta| - \frac{\pi}{2} \right), & \frac{\pi}{2} \leq |\theta| \leq \pi, \end{cases} \quad (2.26)$$

onde a é o raio da cabeça e c é a velocidade do som. Já o sombreamento da cabeça é modelado por uma função de transferência de único zero e único polo, derivada da solução de Rayleigh para difração da onda plana por uma

⁷ O mesmo Richard O. Duda do livro *Pattern Classification* (DUDA; HART; STORK, 2012).

esfera rígida. Assumindo uma taxa de amostragem f_s , a função de transferência para um ângulo de incidência θ é dada pela função de tempo discreto abaixo, mapeada pela Transformada- z :

$$\text{HS}(z, \theta) = \frac{\left(\frac{c}{a} + \alpha(\theta)f_s\right) + \left(\frac{c}{a} - \alpha(\theta)f_s\right)z^{-1}}{\left(\frac{c}{a} + f_s\right) + \left(\frac{c}{a} - f_s\right)z^{-1}}, \quad (2.27)$$

com a localização do zero da função configurada como uma função de θ segundo

$$\alpha(\theta) = 1,05 + 0,95 \cos(1,2\theta). \quad (2.28)$$

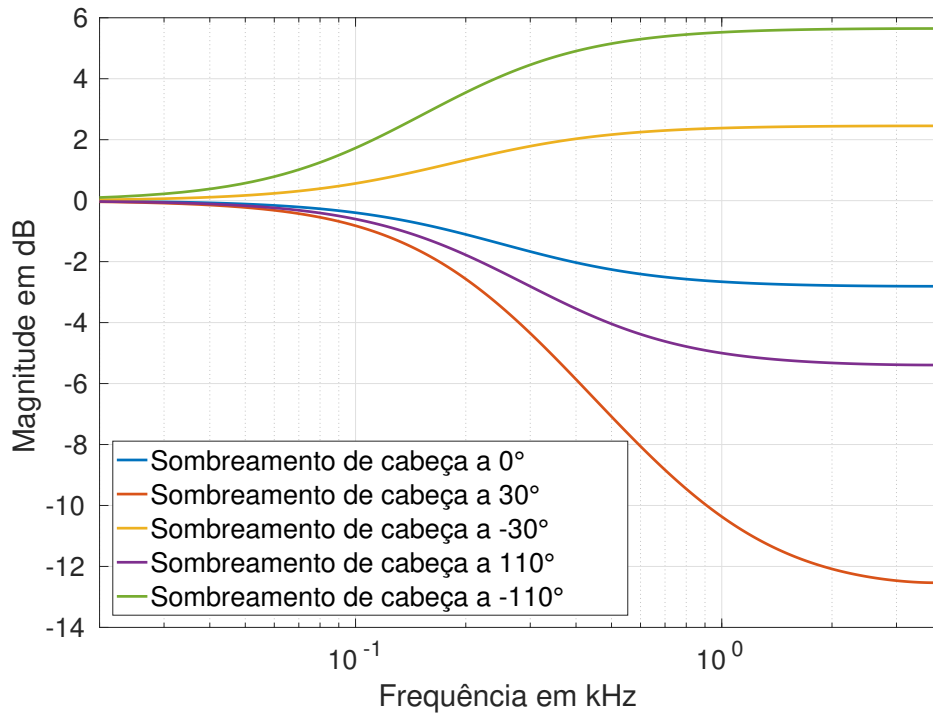
Combinando a ITD com a função de sombreamento de cabeça, tem-se então uma primeira aproximação da HRTF, da forma:

$$\text{HRTF}(z, \theta) = z^{-f_s \text{ITD}(\theta)} \text{HS}(z, \theta). \quad (2.29)$$

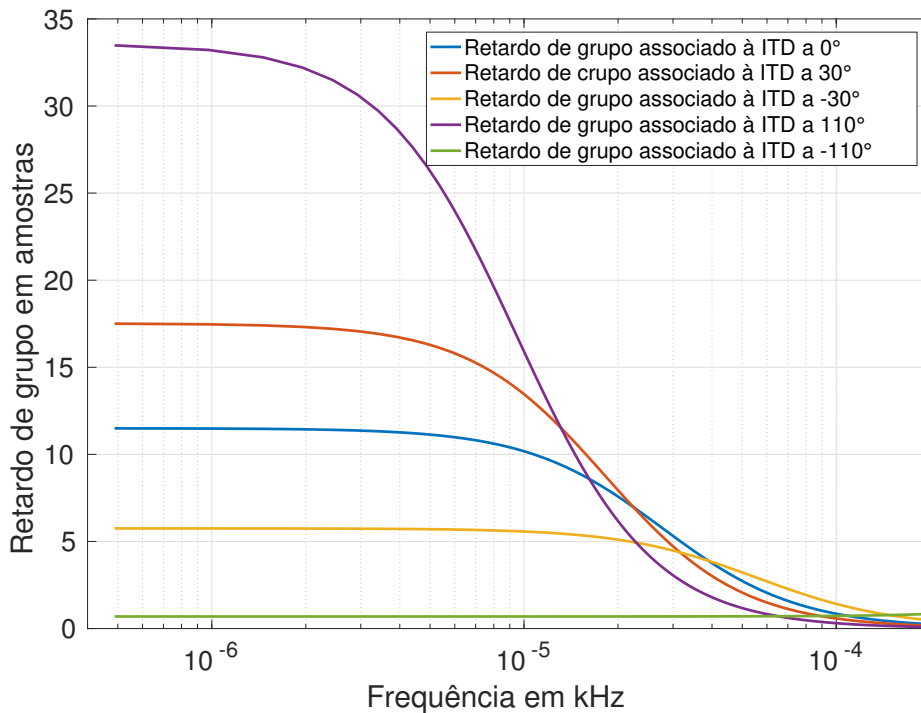
A diferença de tempo interaural $\text{ITD}(\theta)$, interpretada como um retardo, pode ser obtida por um filtro passa-tudo de primeira ordem. O sombreamento de cabeça $\text{HS}(z, \theta)$, por sua vez, é implementado por um filtro de Resposta ao Impulso Infinita (IIR) de um pólo e de um zero (ZÖLZER, 2011). As respostas em frequência de $\text{HS}(z, \theta)$ e os retardos de grupo de $\text{ITD}(\theta)$ no ouvido direito para fontes sonoras distribuídas como num sistema de reprodução de cinco canais (5.1) são ilustradas na Figura 2.16 considerando uma taxa de 48000 amostras por segundo. Na Figura 2.16(a), note que as respostas em frequência com ganho referem-se às fontes diretamente incidentes, correspondendo aos canais direito e *surround* direito, e as incidências associadas aos canais esquerdo e *surround* esquerdo são penalizadas pelo sombreamento. Na Figura 2.16(b), os maiores retardos de grupo estão associados às incidências contralaterais e os retardos menores correspondem às incidências ipsilaterais.

Uma melhor aproximação das HRTFs consiste na incorporação do retardo causado por torso e ombros e pelas reflexões no pavilhão auricular, sendo ambos dependentes da elevação. A partir dos levantamentos de Respostas ao Impulso Relativas à Cabeça (HRIRs) ipsi- e contralaterais feitos por Brown e Duda (1998), uma fórmula para o retardo de ombros e torso dependente do azimute e da elevação (ϕ), sendo $\phi = 0^\circ$ a incidência frontal, foi escrita por Zölzer (2011)

Figura 2.16 – (a) Respostas em frequência de filtros de sombreamento de cabeça . (b) Retardos de grupo associados às diferenças de tempo interaural.



Fonte: Adaptada de [Brown e Duda \(1998](#), com base na implementação de [Zölzer \(2011\)](#)).



Fonte: Adaptada de [Brown e Duda \(1998](#), com base na implementação de [Zölzer \(2011\)](#)).

Nota – As curvas do ouvido direito consideram $f_s = 48000$ amostras/segundo e dizem respeito a fontes sonoras distribuídas tal como num sistema de reprodução de cinco canais.

da forma:

$$\tau_{sh} = 1,2 \frac{180 - \theta}{180} \left(1 - 0,00004 \left((\phi - 80) \frac{180}{180 + \theta} \right)^2 \right), \quad (2.30)$$

com θ e ϕ em graus.

Já os efeitos do pavilhão auricular dão-se na forma de múltiplas reflexões obtidas por uma linha de retardo com derivações. [Brown e Duda \(1998\)](#) escreveram o retardo da forma:

$$\tau_{p_n} = A_n \cos(\theta/2) \text{sen}(D_n(90 - \phi)) + B_n, \quad (2.31)$$

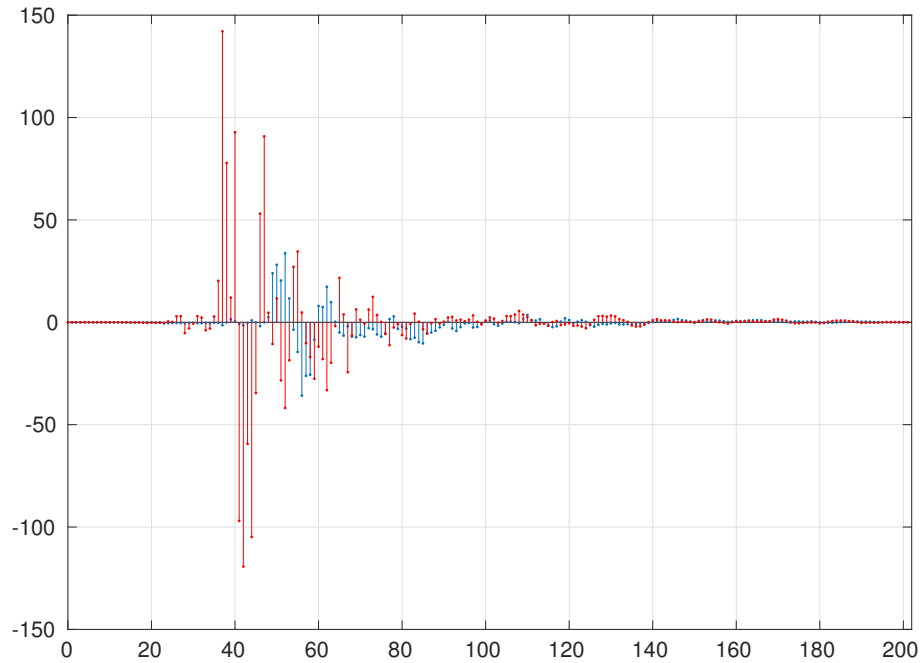
onde A_n é a amplitude da n -ésima reflexão, B_n é um deslocamento inicial (*offset*) e D_n é um parâmetro de escala usado para ajustar o retardo a um formato individual de pavilhão auricular.

Dados de HRIRs medidos experimentalmente são obtidos medindo-se a pressão sonora em microfones colocados no pavilhão auricular de seres humanos ou manequins. Há bases de dados com respostas ao impulso medidas em laboratório para vários azimutes e elevações. A [Figura 2.17](#) ilustra respostas ao impulso constantes da base de dados compilada por [Algazi et al. \(2001\)](#) de ambos os ouvidos para uma mesma fonte sonora e funções de transferência para cinco fontes posicionadas como num sistema *surround*.

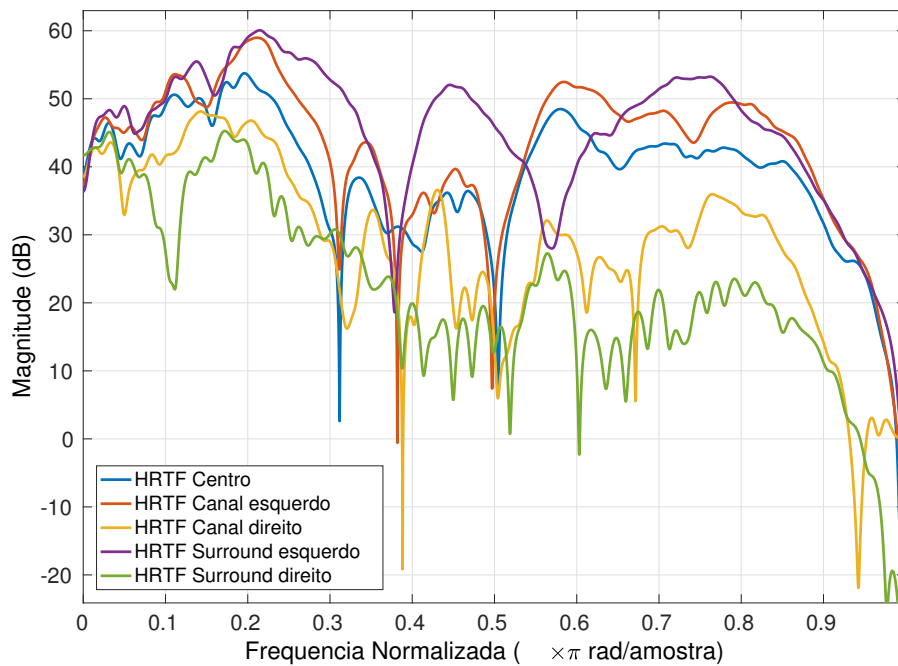
A [Figura 2.17\(a\)](#) ilustra HRIRs nos ouvidos direito (em vermelho) e esquerdo (em azul) para uma fonte sonora situada 30 graus à direita do ouvinte e a zero grau de elevação. Note que a resposta do ouvido direito possui amplitudes significativamente maiores do que a resposta do ouvido esquerdo, que recebe o som incidente sombreado pela cabeça e retardado pelo torso. Já a [Figura 2.17\(b\)](#) mostra HRTFs no ouvido esquerdo para fontes sonoras posicionadas como um sistema de reprodução *surround* de cinco canais. É possível verificar que as respostas em frequência com maiores magnitudes são as das fontes ipsilaterais (canal esquerdo e *surround* esquerdo), sendo as respostas em frequência contralaterais (canal direito e *surround* direito) penalizadas pelo sombreado causado pela cabeça do ouvinte.

Se, por um lado, um modelo de HRTFs paramétricas é tão bom quanto melhor for sua aproximação em relação aos valores medidos, conjuntos de

Figura 2.17 – (a) Respostas ao impulso relativas à cabeça (HRIRs) nos ouvidos direito (em vermelho) e esquerdo (em azul) para uma fonte sonora situada 30 graus à direita do ouvinte. (b) Funções de transferência relativas à cabeça (HRTFs) no ouvido esquerdo para 5 fontes posicionadas como um sistema *surround*.



Fonte: Elaborada pelo autor.



Fonte: Elaborada pelo autor.

Nota – Estas HRIRs são encontradas na base de dados de [Algazi et al. \(2001\)](#).

dados de HRTFs medidos experimentalmente podem apresentar variabilidade considerável entre voluntários/manequins (CHENG; WAKEFIELD, 1999).

O *somatório biauricular de loudness* e a sensação de *loudness* em campos sonoros são efeitos de audição espacial cujos impactos na percepção de intensidade não devem ser desconsiderados. Entretanto, a dependência do *loudness* para com a espacialidade foi recentemente mais investigada em áreas específicas de pesquisa auditiva – tais como áudio espacial e síntese biauricular – do que na própria psicoacústica tradicional. Consequentemente, apenas um dos principais modelos de *loudness* que serão apresentados no [Capítulo 3](#) contempla aspectos espaciais da percepção de intensidade sonora. As noções adquiridas com esta revisão bibliográfica sugerem que o desenvolvimento de um novo algoritmo de predição de *loudness* passa necessariamente por refinar a modelagem da espacialidade sonora, o que é tido como o norte desta pesquisa.

MODELOS DE LOUDNESS

Um dilema clássico de qualquer modelagem reside na escolha entre descrever a estrutura interna do sistema – no caso em questão, o sistema auditivo – ou modelar as características de transferência das variáveis de entrada às variáveis de saída. O peso desta escolha é determinado pelo conhecimento adquirido sobre o sistema, com o qual é possível dividi-lo em subsistemas mais simples e matematicamente tratáveis. Modelos auditivos complexos podem cobrir desde o ouvido externo até as células ciliares, como visto na [seção 2.1](#).

No que tange aos modelos de *loudness*, se apenas algumas categorias de sinais de entrada forem consideradas, os modelos podem ser bastante simplificados. Sendo as entradas restritas somente a tons puros, por exemplo, determinar uma função de *loudness* por meio de experimentos de escuta e posterior cálculo dos parâmetros do modelo pode ser feito de forma direta por uma abordagem “fechnerista” (ver [seção 2.2](#)), e este mesmo modelo simplificado poderia alternar entre funções distintas para o tratamento de sinais com diferentes conjuntos de características, como um aplicável a tons puros e outro a ruídos. Por conseguinte, a complexidade dos modelos de *loudness* aumentou progressivamente conforme o leque de categorização de sinais de entrada. Nesta ordem, surgem os modelos para sons caracterizados somente por seus espectros de frequência, modelos para sons de características flutuantes e os primeiros métodos para fala e música.

No [Capítulo 2](#), a descrição dos efeitos mais importantes para o desenvolvimento de modelos de *loudness* deu-se aliada à ordem cronológica dos

experimentos pioneiros que observaram esses mesmos efeitos. Por razões históricas, a relação da percepção de intensidade com a frequência foi a mais investigada dentre as dependências do *loudness* (ver [subseção 2.2.2](#) e [seção 2.3](#)), e esta mesma relação é o pilar de sustentação dos principais modelos multifaixa para sons estacionários, ao mesmo tempo em que os aspectos temporais ganham força na modelagem de faixa única e são posteriormente incorporados nas evoluções dos modelos multifaixa para sons não estacionários. E, ainda que de forma tímida, os efeitos espaciais no *loudness* foram modelados somente com o advento do algoritmo ITU-R BS.1770, característica que contribuiu com sua consolidação em previsões de intensidade percebida do áudio digital multicanal.

Em outras palavras, ao apresentar a fortuna crítica de modelos de *loudness* dentro de uma arborescência de classificações, este capítulo procurará fazê-lo dentro de uma perspectiva histórica para que seja possível compreender a metodologia de previsão de *loudness* por meio da sua própria evolução, contribuindo assim para situar historicamente esta pesquisa e sua relevância.

3.1 O Modelo de Detecção de Energia

Na segunda metade dos anos 1950, a psicofísica ganha força com o advento da teoria geral de detecção de sinal que, se aplicada a seus experimentos, permitiria a análise da estrutura do processo decisório do observador em tarefas psicofísicas. Em 1954, matemáticos e engenheiros das Universidades de Michigan ([PETERSON; BIRDSALL; FOX, 1954](#)), de Harvard e do *Massachusetts Institute of Technology* (MIT) ([METER; MIDDLETON, 1954](#)) apresentaram a teoria baseada em conceitos de estatística e de telecomunicações. A parte baseada em estatística deriva diretamente da teoria da decisão e de testes de hipóteses, enquanto que a parte derivada das telecomunicações especifica o receptor/detector ideal e o conceito de sensibilidade como sendo função de parâmetros mensuráveis do sinal e do ruído interferente.

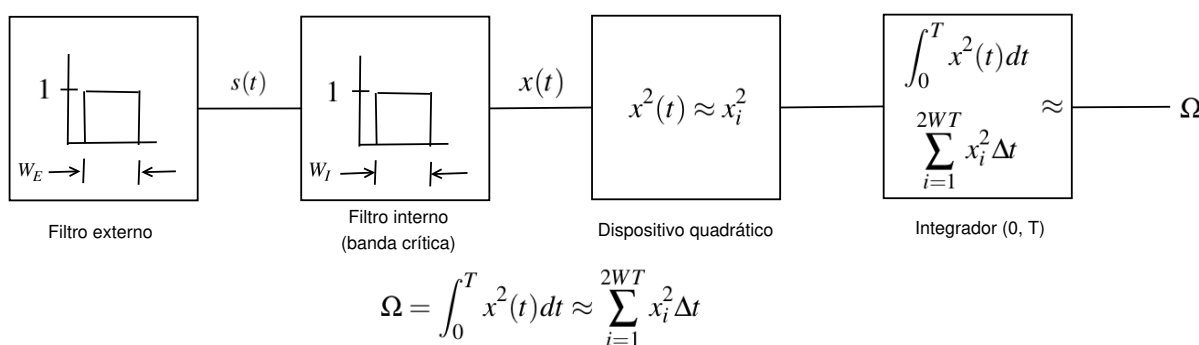
A partir dos primeiros experimentos isolados em psicofísica com detecção de sinal, David [Green \(1960b\)](#) fez um trabalho de referência da teoria aplicada à psicoacústica objetivando padronizar procedimentos experimentais e citando os mesmos problemas envolvidos na definição das escalas de *loudness* e elencados

na subseção 2.2.2 como uma motivação alarmista:

Uma segunda característica do campo [da psicoacústica] é a falta de qualquer estrutura integradora a partir da qual se observar a literatura experimental em expansão. Havendo alguma estrutura teórica básica, novos dados seriam facilmente integrados a antigos. A psicoacústica, contudo, não possui qualquer teoria completa e abrangente. Um reflexo deste déficit está na falta de consenso sobre metodologias. Frequentemente, mesmo quando há um consenso geral numa determinada área, um novo artigo poderá demandar um re-exame completo de todo o procedimento de medida. Um exemplo recente pode ser encontrado nas contribuições de Garner¹ e Stevens² sobre a escala quantitativa de loudness. (GREEN, 1960b, p. 1189)

No mesmo ano, Green (1960a) propôs um modelo de detetor auditivo que alicerçou todos os modelos de loudness já desenvolvidos. O sistema é composto de três partes distintas: um primeiro estágio de filtragem pré-deteção, um dispositivo quadrático (ou um circuito detetor, tal como no jargão das radiocomunicações) e um integrador (ou filtro pós-deteção). A introdução do pré-filtro é sugerida pelo próprio conceito de banda crítica dos experimentos de Fletcher (1940). O dispositivo quadrático serve para fornecer uma quantidade cujo valor médio esteja relacionado à intensidade do sinal, valor então fornecido pelo integrador. Seu diagrama em blocos é ilustrado na Figura 3.1.

Figura 3.1 – Diagrama em blocos do modelo de detecção de energia



Fonte: Adaptada de Green e Swets (1966, p.211).

No diagrama, o sinal senoidal de entrada $s(t)$ passa por um filtro interno de ganho unitário e largura de banda W_I . O sinal filtrado $x(t)$ é limitado em faixa numa dada largura W . Para casos de filtragem única, a largura de faixa

¹ W. R. Garner, J. Acoust. Soc. Am. 30, p. 1005 (1958)

² S.S. Stevens, J. Acoust. Soc. Am. 31, p. 995 (1959).

W refere-se diretamente à largura W_I . Mas para o experimento de banda crítica no qual o autor se baseou, a limitação em banda é feita por um filtro externo também de ganho unitário mas de largura W_E , sendo W_I a largura de banda crítica avaliada. A largura efetiva W pode ser então W_I ou W_E , a que for menor. O sinal $x(t)$ possui um intervalo de amostragem $\Delta t = 1/2W$ segundos com variância σ_n^2 igual à potência média de $x(t)$, N_0W , onde N_0 é a densidade espectral de potência do ruído gaussiano branco. Este é então elevado ao quadrado e integrado no intervalo de duração do sinal, de 0 a T . A estatística de saída Ω pode então ser escrita como:

$$\Omega = \int_0^T x^2(t)dt \approx \sum_{i=1}^{2WT} x_i^2 \Delta t, \quad (3.1)$$

sendo a energia do sinal de entrada escrita da forma:

$$E_s = \int_0^T s^2(t)dt \approx \sum_{i=1}^{2WT} s_i^2 \Delta t. \quad (3.2)$$

No experimento de detecção de um tom senoidal, uma das hipóteses abaixo é verdadeira:

$$\begin{cases} x(t) \text{ é } n(t), & \text{somente ruído,} \\ x(t) \text{ é } n(t) + s(t), & \text{sinal adicionado ao ruído.} \end{cases} \quad (3.3)$$

Dado somente ruído, a estatística Ω_n é escrita como:

$$\Omega_n = \int_0^T n^2(t)dt \approx \sum_{i=1}^{2WT} n_i^2 \Delta t. \quad (3.4)$$

A variável aleatória n_i é gaussiana e independente dos demais termos do somatório. Ao normalizá-la, e se considerarmos que sua variância $\sigma_n^2 = N_0W$, onde N_0 é o valor do espectro de potência do ruído branco, tem-se:

$$\frac{\Omega_n}{\Delta t N_0 W} = \sum_{i=1}^{2WT} \left(\frac{n_i}{\sigma_n} \right)^2, \quad (3.5)$$

Substituindo $\Delta t = 1/2W$ na [Equação 3.5](#), tem-se a variável aleatória $2\Omega_n/N_0$ para a hipótese nula (somente ruído) e, analogamente, $2\Omega_{s+n}/N_0$ para a hipótese alternativa (sinal mais ruído). A variável aleatória h correspondente à probabilidade de acerto do teste de detecção é então formulada como sendo a

diferença das quantidades detetadas em cada hipótese, como desenvolvida por Green e Swets (1966):

$$h = \frac{2}{N_0}(\Omega_{s+n} - \Omega_n) = \frac{E_s/N_0}{(2WT + 2E_s/N_0)^{1/2}} \quad (3.6)$$

Além de se querer conhecer o comportamento do detetor conforme a variação da intensidade do sinal de entrada, a preocupação à época estava em se saber a razão sinal-ruído para uma dada probabilidade de detecção. Se $2WT \gg 2E_s/N_0$, a Equação 3.6 pode ser aproximada desprezando-se o termo $2E_s/N_0$ no denominador:

$$E_s/N_0 \approx h \times (2WT)^{1/2} \quad (3.7)$$

ou, em decibels:

$$\varepsilon - \eta_0 = 10 \log E_s/N_0 \approx 10 \log h + 10 \log (2WT)^{1/2}. \quad (3.8)$$

Green e Swets (1966) propuseram um cenário exemplo de um experimento de escolha forçada de duas alternativas, no qual a variável aleatória h é dada por:

$$h = Z(AUC) \quad (3.9)$$

onde $Z(\cdot)$ é o inverso da função de distribuição cumulativa da distribuição Gaussiana, e AUC é a área sob a curva Característica de Operação do Receptor (ROC) do detetor de energia. Para uma probabilidade de detecção correta de 76%, $h = 1/\sqrt{2}$. Portanto:

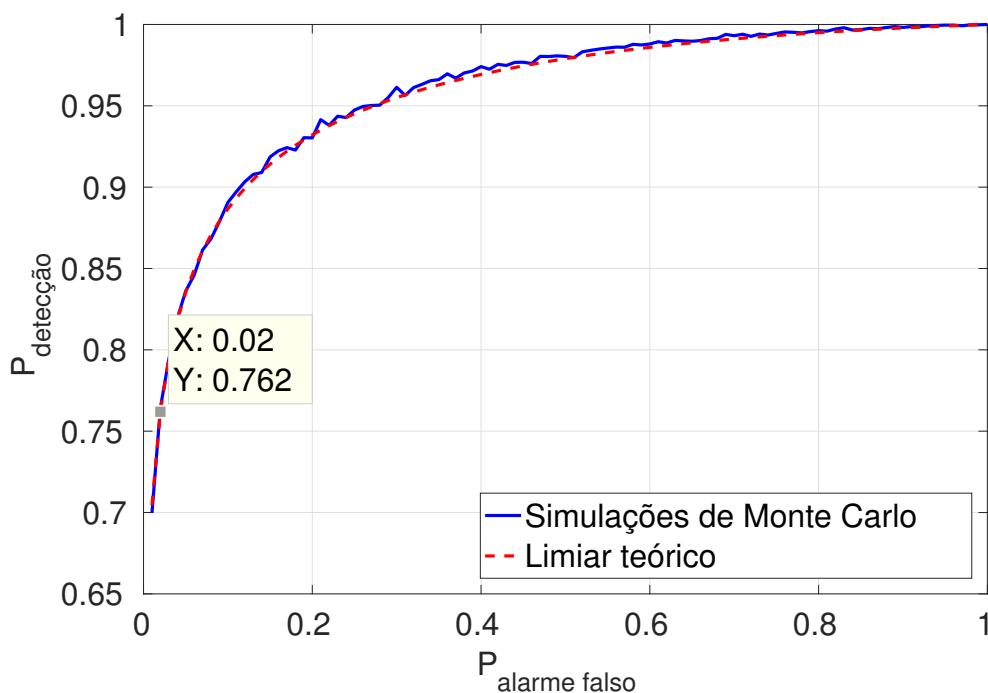
$$\varepsilon - \eta_0 \approx 10 \log (WT)^{1/2} \quad (3.10)$$

ou

$$E_s/N_0 \approx (WT)^{1/2}. \quad (3.11)$$

Na Equação 3.11, o tempo de integração é suposto sendo igual à duração T do sinal. Para um tom de 1 kHz com duração de 100 ms na entrada do mesmo detetor, uma probabilidade de detecção de 76% é obtida com uma razão energia-ruído $E_s/N_0 = 10$, ou $S_0/N_0 = 0,107$ como medida no experimento original (GREEN, 1960a, p. 125-126). A curva ROC de um detetor de energia ao lidar com uma razão sinal/ruído de aproximadamente -10 dB ilustrada na Figura 3.2

Figura 3.2 – Curva ROC para detecção da energia de um tom de 1 kHz com duração de 100 ms e com uma razão sinal-ruído de -10 dB.



Fonte: Elaborada pelo autor.

Nota – A probabilidade de detecção foi estimada por simulações de Monte Carlo (linha cheia em azul) e calculada como em (LIANG *et al.*, 2008, Eqs. 12-14) (linha tracejada em vermelho).

indica que, para uma probabilidade de detecção de 76%, a probabilidade de alarme falso – ou de detecção incorreta – é de aproximadamente 2%.

Apresentado todo o equacionamento, chega-se a um ponto importante para esta narrativa. Fletcher (1940) definiu a ocorrência do mascaramento em frequência quando a potência do sinal – de duração ilimitada em seu experimento – se igualava à potência do ruído limitado à largura de banda crítica (ver subseção 2.3.1). Resolvendo a Equação 3.11 para as configurações do parágrafo anterior, tem-se $WT \approx 100$ ou, sendo $T = 100$ ms, $W = 1000$ Hz. Esta estimativa da banda crítica é cerca de 16 vezes mais larga que a suposta por Fletcher, 60 Hz para $f_c = 1$ kHz, e 5 vezes maior que a obtida nos experimentos de Zwicker, Flottorp e Stevens (1957), 160 Hz para $f_c = 1$ kHz. Contudo, é razoável assumir que o tempo de integração seja, na prática, superior à duração do sinal em alguma medida. Se, por exemplo, o limite superior de integração for de 200 ms (um pouco mais distante do limiar de integração temporal de loudness para tons senoidais visto na seção 2.4), a banda crítica estimada pelo detetor de energia

passa a ser de 500 Hz, menos inflacionada em relação aos resultados obtidos experimentalmente (GREEN; SWETS, 1966). Esta descoberta sugere que limites de integração adequadamente dimensionados podem levar a predições melhores de intensidade percebida.

Dadas as aproximações da Equação 3.6, o modelo de detecção de energia prediz uma relação entre a largura de banda W_E limitada externamente e E_s/N_0 para uma mesma probabilidade de detecção (GREEN; SWETS, 1966):

$$\begin{cases} \frac{E_s}{N_0} = h \left(2W_E T + \frac{2E_s}{N_0} \right)^{1/2}, & W_E < W_I \\ \frac{E_s}{N_0} = h \left(2W_I T + \frac{2E_s}{N_0} \right)^{1/2}, & W_I < W_E \end{cases} \quad (3.12)$$

Se $W_E < W_I$ e há um alargamento de W_E , a potência média entregue ao detetor ($W_E N_0$) aumenta de acordo. Todavia, isso não altera a eficiência do detetor para uma mesma razão E_s/N_0 . A importância do alargamento de W_E está no aumento do número de amostras $2W_E T$, que aumenta a razão média por desvio padrão de Ω resultando em medidas mais precisas de Ω . Aumentar a largura de banda limitante de $x(t)$ – ou aumentar a taxa de amostragem – leva a uma melhor resolução das predições de intensidade percebida. Esta observação pode parecer elementar à luz do moderno processamento digital de sinais, mas é preciso pensar em sua importância para a psicoacústica numa época de filtragem analógica e eletrônica discreta.

Esta seção se encerra com a constatação de que uma definição rápida do modelo de detecção de energia como sendo um modelo RMS seria um reducionismo frente à importância deste para todos os métodos de predição de *loudness* que se seguiram.

3.2 Modelos Multifaixa

A viabilidade de um modelo de detecção de energia acústica filtrada em banda crítica, somada às descobertas experimentais de Zwicker (1961) e Glasberg e Moore (1990) sobre as larguras críticas e retangulares equivalentes (ver Figura 2.10), culminou num desenvolvimento natural de métodos de predição de *loudness*, projetados a partir da inclusão no detetor de novas etapas de pro-

cessamento ligadas aos efeitos descritos no [Capítulo 2](#) dos quais o *loudness* é dependente.

Mas, principalmente, os modelos multifaixa nasceram da reinterpretação do bloco integrador do detetor de energia à luz da descoberta da acumulação espectral de *loudness*, tal como descrita na [subseção 2.3.1](#), e de sua quantificação nas palavras dos próprios E. Zwicker e H. Fastl:

Por termos visto que dois tons influenciam um ao outro na criação da sensação total de *loudness*, muito embora estejam separados espectralmente, pode ser útil tratar o *loudness* total como a integral de um valor a ser encontrado, mas que possa ser desenhado como uma função das bandas críticas. Se tal integral resultar num *loudness* dado em *sones*, o valor requerido deverá ser expresso na unidade *sones/bark*. Neste caso, o *loudness* total, dado pela integral deste valor ao longo da escala de bandas críticas, será expresso em *sones*. (FASTL; ZWICKER, 2007, p. 220-221)

A explicação é sintetizada na equação abaixo

$$N = \int_1^{24 \text{ bark}} N' dz, \quad (3.13)$$

onde N é o *loudness* total em *sones*, N' é o *loudness* específico em *sones/bark* e dz é o incremento infinitesimal da frequência $z(f)$ em *bark* calculada pela [Equação 2.18](#).

3.2.1 Modelos para sons estacionários

Os dois pioneiros foram S. S. Stevens (1961) e E. Zwicker (1960). Seus modelos gráficos compuseram a norma ISO (1975) número 532, intitulada “Métodos para o cálculo de nível de *loudness*”.

O modelo de Stevens é identificado na norma como “Método A”, sendo mais adequado a sinais de faixa larga sem picos espectrais pronunciados com separação larga em frequência, e assume um campo sonoro difuso e um som estacionário. A acumulação espectral de *loudness* é contabilizada pela integração em faixas de oitava, embora o método possa ser adaptado para operar com faixas

de meia oitava e faixas de um terço de oitava³. O método é ancorado numa tabela de índices de *loudness* para um tom 1 kHz a vários níveis de pressão sonora e numa aproximação das curvas de Fletcher e Munson (1933) em *sones*. O procedimento foi descrito por Stevens (1961) como a seguir:

1. Entrar com a média geométrica das frequências de cada faixa na tabela ou na abcissa da Figura 3.3. Então para cada nível, determinar o índice de *loudness* de cada faixa.
2. Encontrar o *loudness* total S_t pela fórmula

$$S_t = S_m + F \left(\sum S - S_m \right), \quad (3.14)$$

onde S_m é o maior dos índices de *loudness* e $\sum S$ é a soma dos índices de *loudness* em todas as faixas. O valor do fator F depende da largura de faixa usada na análise de ruído, como na Tabela 3.1.

Tabela 3.1 – Valor do fator F conforme larguras utilizadas no banco de filtros do modelo de Stevens para sons estacionários (ISO 532-A).

Largura de faixa	Fator F
Terça de oitava	0,15
Meia oitava	0,2
Oitava	0,3

Fonte: Stevens (1961, p. 1.578).

3. O *loudness* total poderá então ser convertido em nível de *loudness* calculado pela fórmula

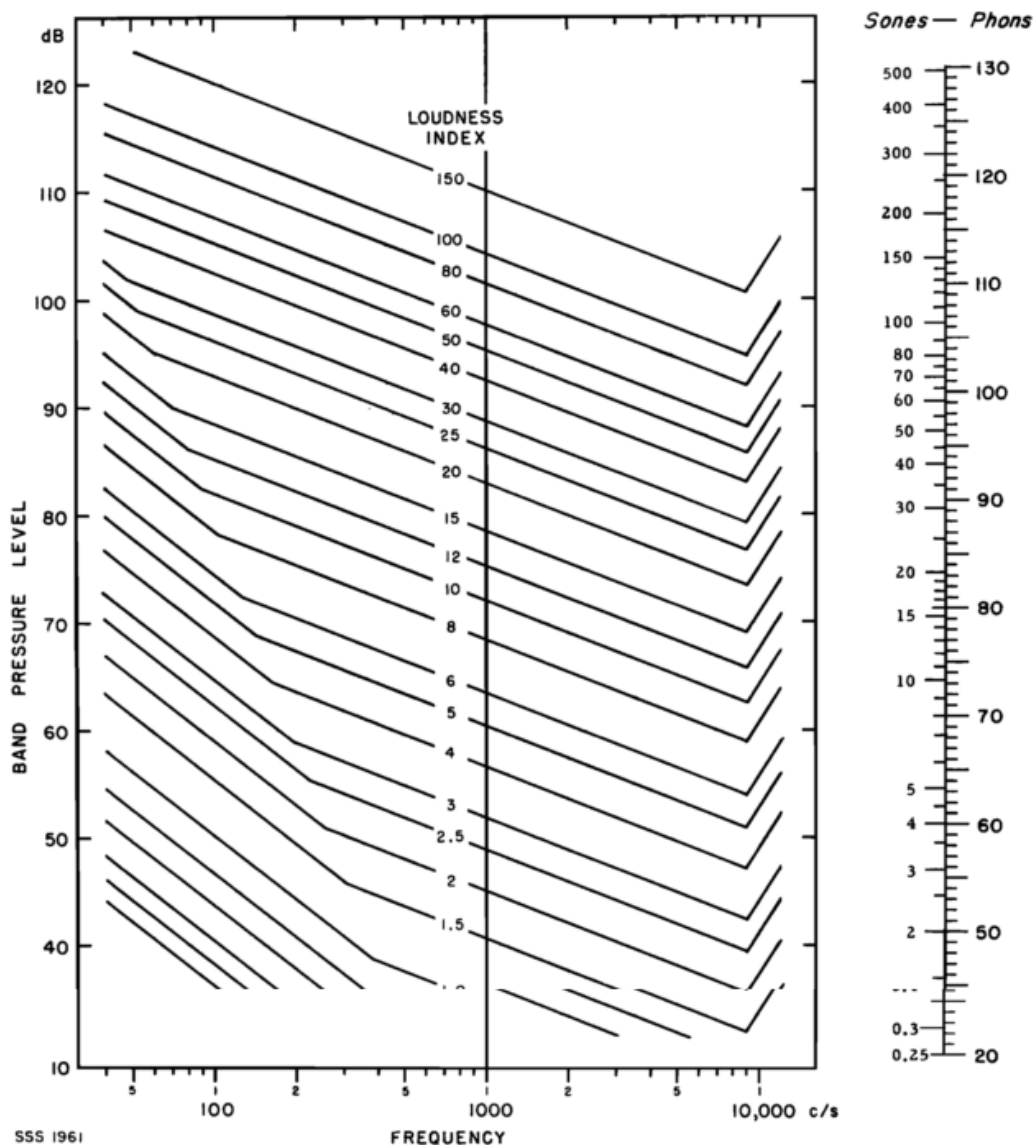
$$S_t = 2^{(P_t - 40)/10} \quad (3.15)$$

ou

$$P_t = 40 + 10 \log_2 S_t. \quad (3.16)$$

Um nomograma desta relação encontra-se à direita da Figura 3.3.

³ O espectro auditivo pode ser dividido em aproximadamente 11 faixas de oitava, 21 faixas de meia oitava e 31 faixas de um terço de oitava. Ao centrar-se a frequência central da 7ª faixa de oitava, da 13ª faixa de meia oitava e da 19ª faixa de um terço de oitava a 1 kHz, as frequências centrais inferiores são obtidas pelas relações $f_{n-1} = f_n/2$, $f_{n-1} = f_n/2^{1/2}$ e $f_{n-1} = f_n/2^{1/3}$, e as frequências centrais superiores são obtidas pelas relações $f_{n+1} = 2f_n$, $f_{n+1} = 2^{1/2}f_n$ e $f_{n+1} = 2^{1/3}f_n$, respectivamente.

Figura 3.3 – Contornos de mesmo índice de *loudness* do modelo multifaixa de Stevens (1961).

Fonte: Stevens (1961, Fig. 1, p.1.579).

Todavia, o método apresentava um descolamento entre a função do índice de *loudness* e a lei de potência do próprio Stevens (1957) entre os níveis de pressão sonora 20 e 40 dB SPL, problema que o autor chamou de “protuberância de nível médio”. É possível que o autor a considerasse como um problema menor, ao decidir prosseguir com o desenvolvimento de um modelo de *loudness* sem observar a descoberta das bandas críticas da audição, estudo do qual o próprio Stevens participou:

Nos anos que se seguiram após esta observação, encontramos outras instâncias impressionantes da protuberância de nível médio. O efeito é particularmente claro nos resultados reportados por Zwicker, Flottorp

e Stevens (1957), que demonstraram que o incremento de *loudness* em níveis médios ocorre somente quando o estímulo excede a “banda crítica”. Uma vez que a protuberância é característica do ruído de faixa larga, deve-se esperá-la quando o equilíbrio de *loudness* é atingido entre um ruído branco e um tom puro que, de fato, foi encontrado por vários experimentos (ver Fig. 7 de Stevens (1955) e Figs. 7 e 11 de Zwicker (1958)) (STEVENS, 1961, p. 1.580).

Já o método de Zwicker (1960) é descrito na mesma norma, mas referido como sendo o “Método B”. Neste, a análise de frequências é feita em terços de oitava, largura considerada como sendo uma aproximação razoável das bandas críticas. No trabalho original foi utilizado um banco de filtros com as exatas larguras das bandas críticas, mas por questões práticas de implementação, o modelo foi normatizado com filtros de terços de oitava. Poderia a inviabilidade técnica à época ser uma razão pela qual Stevens desconsiderou as larguras de banda crítica em seu modelo? Talvez. Anteriormente à publicação da ISO 532:1975, Os trabalhos de Bauer *et al.* (1967) realizados no laboratório da emissora norte-americana Sistema Columbia de Radiodifusão (CBS), resultaram num monitor baseado em filtros de oitava, o primeiro construído para medir *loudness* em conteúdo tipicamente encontrado em emissoras de rádio e TV. Durante a fase de estudos preliminares para o projeto do equipamento, Bauer e Torick (1966) tomaram o modelo gráfico de Stevens como ponto de partida para se levar em conta a acumulação espectral de *loudness*.

Muito embora o método de Stevens e, conseqüentemente, o monitor CBS não tenham sobrevivido ao teste do tempo, cumpre notar que a divulgação dos estudos de Bauer e Torick (1966) foi a primeira apresentação do problema à comunidade de engenharia de radiodifusão. Por mais que os abusos de processamento dinâmico na última década tenham expandido as discussões de *loudness* para além da imprensa especializada, é interessante perceber como as perguntas que fazemos não são assim tão jovens:

“O que pode ser tão difícil sobre medir *loudness*?”, você pode perguntar. “Pode um medidor de Unidade de Volume (VU) ser usado?” ou “Que tal usar um medidor de nível sonoro?” Infelizmente, a solução não é tão simples. Para uma dada leitura de VU, tons de diferentes frequências podem diferir em níveis de *loudness* da ordem de 20 a 30 dB, e o medidor VU por si só não responde propriamente ao *loudness* sensorial de tons combinados (BAUER; TORICK, 1966, p. 141).

Se comparado ao método de Stevens (1961), o modelo de Zwicker (1960) contempla uma acumulação espectral de *loudness* mais sofisticada devido ao uso das larguras de banda crítica em seu banco de filtros. Vale tanto para campos livres quanto para campos difusos e é indicado tanto para faixa estreita como para faixa larga (ISO, 1975). Sua consolidação deu-se nos anos vindouros, quando deixou de ser um método gráfico e teve uma implementação na linguagem BASIC publicada por Zwicker, Fastl e Dallmayr (1984), e quando foi incluído no padrão alemão de cálculo de *loudness* DIN 45631 (ZWICKER *et al.*, 1991), na forma que servirá de referência pelo restante deste texto.

Com o passar dos anos, muitos pesquisadores propuseram refinamentos do modelo original de Zwicker. Em razão da própria formulação das ERBs por Glasberg e Moore (1990) (ver Equação 2.19 e Equação 2.20), estes autores se destacaram pelo aprimoramento do modelo de referência e desenvolvimento de um próprio. Moore e Glasberg (1996) mapearam a estrutura do modelo de Zwicker *et al.* (1991) como sendo baseada no cálculo da excitação da membrana basilar pelas larguras críticas, levando em consideração um modelo de transmissão do sinal acústico pelo ouvido externo e médio, o efeito de mascaramento em frequência no cálculo do *loudness* específico, e a acumulação espectral de *loudness* via integração do *loudness* específico ao longo das 24 bandas críticas. O modelo de Moore, Glasberg e Baer (1997) manteve esta estrutura, diferindo-se no cálculo dos filtros, nas correções de campo livre/difuso e no cálculo do padrão de excitação. As principais etapas de cálculo de *loudness* em ambos os modelos são ilustradas na Tabela 3.2.

O primeiro estágio do método de Zwicker *et al.* (1991) faz uso do fator de transmissão a_0 , que leva em conta a transformação entre o campo livre e o ouvido interno e é adicionado ao sinal original. Esta função de transmissão é similar ao limiar absoluto de audição porém invertido para além de 1 kHz, pois o ouvido interno é considerado igualmente sensível nessa faixa, e é constante para aquém de 1 kHz, pois a elevação do limiar absoluto nas baixas frequências seria somente explicada pelo ruído interno do ouvido (ZWICKER; FASTL, 2013). O fator de transmissão a_0 é ilustrado na Figura 3.4.

Já o modelo de Moore, Glasberg e Baer (1997) trabalha com duas funções de transferência: uma para o meato acústico e outra para o ouvido médio. Com

Tabela 3.2 – Ilustração dos principais estágios dos modelos multifaixa de Zwicker *et al.* (1991) e Moore, Glasberg e Baer (1997) para sons estacionários.

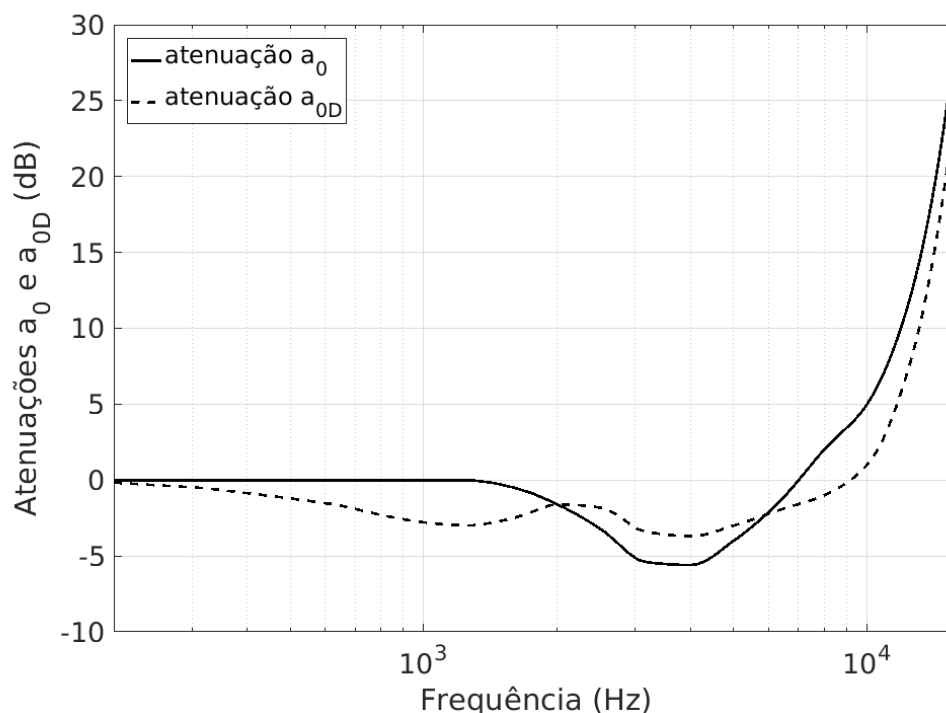
Estímulo
Estágio 1: Filtragem de correção dos efeitos de transferência do ouvido externo e médio
Estágio 2: Filtragem em Larguras Críticas (<i>Bark</i>) ou em Larguras Retangulares Equivalentes (<i>ERB</i>)
Estágio 3: Transformação de espectro para padrão de excitação
Estágio 4: Transformação de padrão de excitação para <i>loudness</i> específico
Estágio 5: Integração do <i>loudness</i> específico ao longo da escala de frequências (<i>Bark</i> ou <i>ERB</i>)

Fonte: Adaptada de Moore e Glasberg (1996, Fig.1, p. 336).

base em medidas da função de transferência do ouvido médio em cadáveres (PURIA; PEAKE; ROSOWSKI, 1997), os autores não associaram a elevação do limiar absoluto nas baixas frequências somente à presença de ruído interno. Portanto, a função de transferência para além de 1 kHz é similar ao limiar da audição invertido, assim como no modelo de Zwicker, mas para frequências aquém de 1 kHz há uma compensação equivalente à curva invertida de 100 phon. Já a função de transferência do meato acústico foi baseada em medidas feitas na ausência do ouvinte em campo livre por Shaw (1974). As respostas em frequência correspondentes são ilustradas na Figura 3.5 b) e a), respectivamente.

O estágio 2 corresponde ao modelo e cálculo dos filtros auditivos, tópico explorado na subseção 2.3.1. O modelo de Zwicker *et al.* (1991) utiliza as larguras críticas calculadas pela Equação 2.17 e pela Equação 2.18, e o modelo de Moore, Glasberg e Baer (1997) utiliza as larguras retangulares equivalentes (ERBs) calculadas pela Equação 2.19 e pela Equação 2.20. Uma comparação entre frequências centrais e larguras de faixa dos dois bancos de filtros é ilustrada na Figura 2.10.

Figura 3.4 – Atenuação correspondente ao fator de transmissão necessário em condição de campo livre (a_0) ou para campo difuso (a_{0D}).



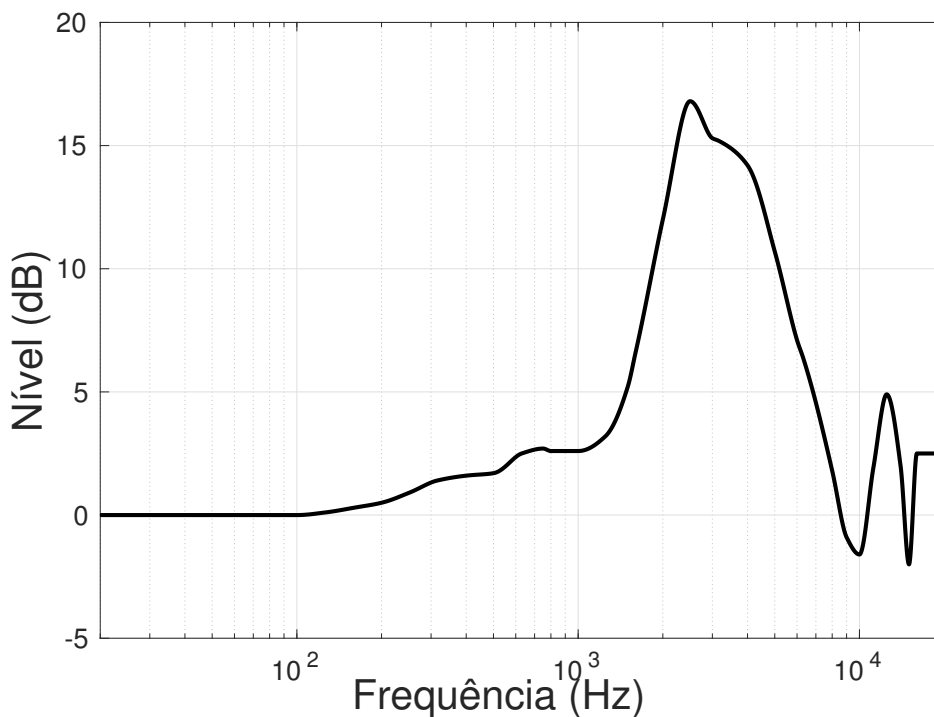
Fonte: Adaptada de Zwicker e Fastl (2013, Fig. 8.18, p. 226).

Nota – A atenuação a_{0D} é a atenuação a_0 descontada da atenuação para campos difusos a_D , apresentada na Figura 2.15.

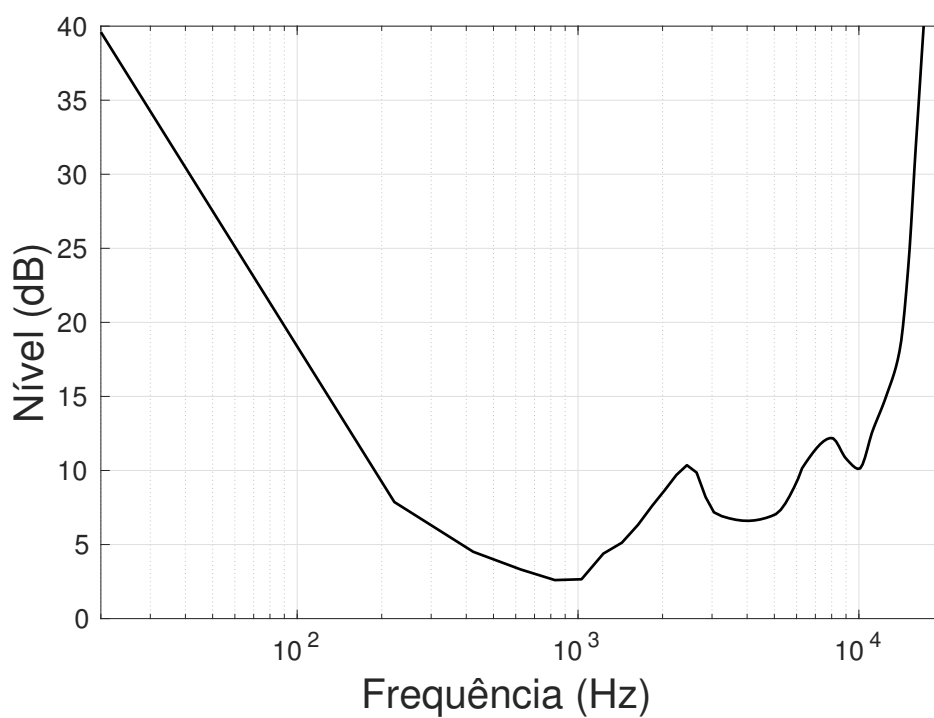
No terceiro estágio, acontece o cálculo do chamado *padrão de excitação*. Assim como as bandas críticas / ERBs se referem às larguras de faixa dos filtros auditivos, o padrão de excitação da membrana basilar é dado pelas magnitudes na saída desses filtros. O cálculo do padrão é uma diferença marcante entre os dois métodos.

Segundo Zwicker, o padrão de excitação reflete o padrão de mascaramento de um tom puro por um UEN. O autor pressupõe que os padrões de excitação e as curvas de limiar de detecção mascarado são idênticos. As curvas de mascaramento evidenciam a seletividade do ouvido e dependem da frequência central e do nível de intensidade sonora. Já de acordo com Moore, o padrão de excitação é calculado pela saída dos filtros auditivos centrados nas frequências que compõem o som. Para o exemplo de um tom puro de 1 kHz, a Figura 3.6(a) ilustra as respostas em frequência dos filtros centrados no entorno de 1 kHz com contribuições desta frequência, representada pelos cruzamentos das curvas de

Figura 3.5 – (a) Função de transferência do nível sonoro de campo livre para o nível sonoro timpânico. (b) Curva de atenuação do ouvido médio.



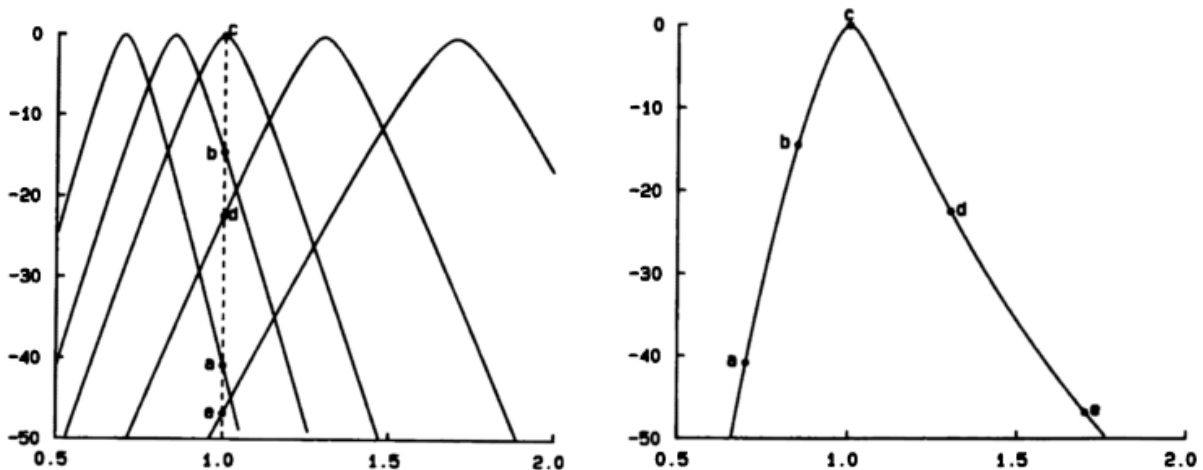
Fonte: Adaptada de Moore, Glasberg e Baer (1997, Fig. 2, p. 225).



Fonte: Adaptada de Moore, Glasberg e Baer (1997, Fig. 3, p. 226).

resposta com a linha vertical. Já a Figura 3.6(b) representa o padrão de excitação do tom puro. As abcissas correspondem às frequências centrais dos filtros na vizinhança do tom e as ordenadas representam a resposta em frequência a 1 kHz destes filtros. O padrão de excitação é construído pela curva que liga estes pontos.

Figura 3.6 – Padrão de excitação de um tom de 1 kHz calculado pelas saídas dos filtros auditivos em função de suas frequências centrais.



Fonte: Moore (2012, Fig. 3.11, p. 92).

Nota – À esquerda, respostas em frequência dos filtros centrados no entorno de 1 kHz com contribuições desta mesma frequência. À direita, o padrão de excitação resultante para o tom de 1 kHz

O quarto estágio calcula o *loudness* específico (N'), que pode ser interpretado a partir da Lei de Potência de Stevens (1957) ao se reescrever a Equação 2.7 da forma

$$N' = kE^n, \tag{3.17}$$

onde o *loudness* específico N' é a *sensação* associada ao estímulo do padrão de excitação E .

Zwicker calcula o *loudness* específico reescrevendo a Lei de Potência usando diferenças (ZWICKER; FASTL, 2013):

$$\frac{\Delta N'}{N'} = k \frac{\Delta E}{E}, \text{ ou } \frac{\Delta N'}{N' + N'_{gr}} = k \frac{\Delta E}{E + E_{gr}} \tag{3.18}$$

sendo N'_{gr} e E_{gr} os pisos de ruído a valores bem baixos de N' e E , respectivamente. E_{gr} é calculada ao se produzir o limiar de silêncio pelo tom de teste E_{TQ} , usando

a equação

$$E_{gr} = E_{TQ}/s, \quad (3.19)$$

onde s é a razão entre a intensidade do tom de teste no limite do audível e a intensidade do ruído interno na largura crítica que contém o tom de teste.

A equação [Equação 3.18](#) é então resolvida usando a condição de contorno “Se $E = 0$, $N' = 0$ ”:

$$N' = N'_{gr} \left[\left(1 + \frac{sE}{E_{TQ}} \right)^k - 1 \right], \quad (3.20)$$

onde N'_{gr} é substituído por um loudness específico de referência

$$N' = N'_0 \left(\frac{E_{TQ}}{sE_0} \right)^k \left[\left(1 + \frac{sE}{E_{TQ}} \right)^k - 1 \right], \quad (3.21)$$

onde N'_0 e E_0 são o *loudness* específico e a excitação correspondentes à intensidade de referência $I_0 = 10^{-12} \text{W/m}^2$, respectivamente.

Quando foi dito na [seção 2.2](#) que a constante k seria o alicerce do “edifício psicofísico” de Fechner, vislumbrou-se este ponto no texto ao qual chegamos. O valor de k pode ser estimado pela dependência do loudness de um UEN como função de seu nível. A [Equação 3.21](#) para grandes valores de E , nos quais a influência do limiar do audível é desprezível, pode ser aproximada como

$$N' \approx \left(\frac{E}{E_0} \right)^k, \quad (3.22)$$

em que um valor adequado de k seria de 0,23 ([ZWICKER; FASTL, 2013](#)).

Em frequências próximas a 1 kHz, para as quais o fator de limiar s seja igual a 0,5, e com a condição de contorno adicional de que um tom de 1 kHz a 40 dB SPL produz exatamente 1 *sone* de *loudness* total, a [Equação 3.21](#) é reescrita da forma

$$N' = 0,08 \left(\frac{E_{TQ}}{E_0} \right)^{0,23} \left[\left(0,5 + 0,5 \frac{E}{E_{TQ}} \right)^{0,23} - 1 \right] \frac{\text{sone}_G}{\text{Bark}}. \quad (3.23)$$

A [Equação 3.23](#) é o formato final do cálculo quantitativo do *loudness* específico. Nela, E_{TQ} é a excitação no limiar do audível. O valor “1” entre

parênteses da [Equação 3.21](#) foi substituído pelo valor $(1 - s)$ para que o *loudness* específico se aproxime assintoticamente do valor $N' = 0$ para pequenos valores de E . O índice G na unidade *sone* indica que o *loudness* foi calculado usando-se os níveis das larguras de faixa críticas ([ZWICKER; FASTL, 2013](#)).

A União Internacional de Telecomunicações usou a [Equação 3.23](#) para o cálculo do *loudness* específico nas Recomendações P.862 do [ITU-T \(2001\)](#) e BS.1387-1 do [ITU-R \(2001\)](#), nos métodos de Avaliação Perceptiva da Qualidade da Fala (PESQ) e de Avaliação Perceptiva da Qualidade do Áudio (PEAQ), respectivamente. A notação utilizada no método PESQ é da forma :

$$LX(f)_n = S_l \left(\frac{P_0(f)}{0,5} \right)^\gamma \cdot \left[\left(0,5 + 0,5 \cdot \frac{PPX(f)_n}{P_0(f)} \right)^\gamma - 1 \right] \quad (3.24)$$

onde $LX(f)_n$ é o *loudness* específico na frequência f para o n -ésimo quadro do cálculo do *loudness* instantâneo (ver [Equação 3.34](#) mais à frente), S_l é um fator de escala, $P_0(f)$ é E_{TQ} na frequência f , $PPX(f)_n$ é E na frequência f para o quadro n , e γ é a própria constante de proporcionalidade k . Na recomendação [ITU-T \(2001\)](#), os valores $S_l = 1/240,5$ e $\gamma = 0,001$.

Já o modelo de [Zwicker et al. \(1991\)](#) utilizado no PEAQ sofreu algumas modificações do ponto de vista de processamento de sinais. O espectro de potência de cada quadro é ponderado pela resposta em frequência do ouvido externo e médio, algo mais próximo do modelo de [Moore, Glasberg e Baer \(1997\)](#). Os espectros são então espaçados de 0,25 bark e agrupados nas larguras críticas a menos de um desconto (*offset*) para compensar o ruído interno do ouvido. Uma função de transferência triangular de espalhamento (em dB) é utilizada para eliminar as bandas de guarda de 0,25 bark, resultando no padrão de excitação espalhado \tilde{E}_{SR} . A [Equação 3.23](#) é escrita mantendo o fator de limiar s da forma:

$$LX(f)_n = c \left(\frac{E_t(f)}{s(f)E_0} \right)^{0,23} \cdot \left[\left(1 - s(f) + \frac{s(f)\tilde{E}_{SR}(f)_n}{E_t(f)} \right)^{0,23} - 1 \right], \quad (3.25)$$

onde $c = 1,07664$ e o fator de limiar em função da frequência é dado por

$$s_{dB}(f) = -2 - 2,05 \tan^{-1} \left(\frac{f}{4000} \right) - 0,75 \tan^{-1} \left(\left(\frac{f}{1600} \right)^2 \right). \quad (3.26)$$

Para níveis sonoros próximos ao limiar de audibilidade, a função de transferência da membrana basilar fica cada vez mais inclinada e se aproxima da linearidade (MOORE; GLASBERG, 1996). No modelo de Moore, Glasberg e Baer (1997), os autores repensaram a Lei de Potência de Stevens (1957) ao presumir que a função que relaciona a sensação de *loudness* específico e a excitação também incrementa a inclinação. A Equação 3.17 é então reescrita da forma

$$N' = k[(E_{\text{signal}} + A)^n - A^n], \quad (3.27)$$

onde A é uma “constante que pode depender da frequência”. Para frequências superiores a 500 Hz, os autores consideram A sendo constante e igual a duas vezes E_{TQ} .

Os autores observaram que o ganho aplicado pela amplificação coclear decresce em baixas frequências, que levaria a um aumento de E_{TQ} , uma elevação da inclinação da função de transferência da membrana basilar e o consequente aumento da potência n (MOORE; GLASBERG; BAER, 1997). Daí foi introduzido o termo G , que representa o ganho relativo de amplificação coclear numa frequência específica, relativo à constante A para frequências superiores a 500 Hz, de tal forma que o produto $G \cdot E_{TQ}$ seja constante. A Equação 3.27 torna-se:

$$N' = k[(GE_{\text{signal}} + A)^n - A^n]. \quad (3.28)$$

O *loudness* específico de pico de um tom puro no limiar do audível é suposto independente da frequência, ou seja, quando $E_{\text{signal}} = E_{TQ}$, a Equação 3.28 resulta numa constante

$$N'_{\text{limiar}} = k[(GE_{TQ} + A)^n - A^n] = 0,000537. \quad (3.29)$$

Contudo, os limiares de sinais de faixa larga encontrados pela Equação 3.28 foram considerados baixos demais, o que levou os autores a presumirem que quando E_{signal} decresce para valores próximos a E_{TQ} , o *loudness* específico decresce mais rapidamente do que como descrito na Equação 3.28. A equação foi então corrigida pela introdução de um termo para quando $E_{\text{signal}} < E_{TQ}$:

$$N' = k \left(\frac{2E_{\text{signal}}}{E_{\text{signal}} + E_{TQ}} \right)^{1,5} [(GE_{\text{signal}} + A)^n - A^n], \quad (3.30)$$

e há uma correção análoga para valores de E_{signal} próximos a 100 dB SPL da forma

$$N' = k \left(\frac{E_{\text{signal}}}{1,04 \times 10^6} \right)^{0,5} . \quad (3.31)$$

O cômputo do *loudness* específico total é dado pelas contribuições das excitações do sinal e do ruído somadas. Portanto, o *loudness* específico do sinal descontado do *loudness* específico do ruído pode ser calculado da forma:

$$N'_{\text{signal}} = N'_{\text{total}} - N'_{\text{ruído}} \quad (3.32)$$

$$= k \{ [(E_{\text{signal}} + E_{\text{ruído}})G + A]^n - A^n \} - k [(E_{\text{ruído}}G + A)^n - A^n], \quad (3.33)$$

com as correções da [Equação 3.29](#), da [Equação 3.30](#) e da [Equação 3.31](#), quando aplicáveis.

Por fim, o quinto estágio da [Tabela 3.2](#) refere-se à integração do *loudness* específico ao longo da escala de frequências como na [Equação 3.13](#), ao longo das larguras críticas no modelo de [Zwicker et al. \(1991\)](#), ou ao longo das ERBs no modelo de [Moore, Glasberg e Baer \(1997\)](#), fazendo uso do mesmo bloco integrador do modelo de detecção de energia descrito na [seção 3.1](#).

3.2.2 Modelos para sons não estacionários

Os métodos de predição de *loudness* de sinais não estacionários foram desenvolvidos de modo a capturarem os efeitos temporais na percepção de intensidade descritos na [seção 2.4](#). Para tanto, é necessário distinguir as sensações momentâneas daquilo que pode ser avaliado ao final da duração do áudio. [Glasberg e Moore \(2002\)](#) introduziram os conceitos de *loudness momentâneo* e de *loudness global*. O *loudness* momentâneo tem a ver com a sensação instantânea, e todo o padrão temporal de *loudness* induzido por um dado som pode ser definido como um vetor de medidas de *loudness* momentâneo ao longo do tempo. Já o *loudness* global corresponde ao *loudness* médio percebido ao final da duração do áudio, quando todo o vetor de medidas de *loudness* momentâneo é agregado num único valor.

Os principais modelos evoluíram diretamente dos modelos de referência para sons estacionários. O modelo de [Glasberg e Moore \(2002\)](#) é uma evolução

direta do modelo estacionário de 1997, e do método de Zwicker surgiram os modelos de sinais não estacionários de Zwicker e Fastl (1999) e de Chalupper e Fastl (2002). Os estágios de cálculo são basicamente os mesmos: i) filtragem de compensação dos efeitos do ouvido externo e médio, ii) filtragem auditiva por banco de filtros na escala bark ou ERB de frequências, iii) transformação em níveis de excitação a partir da energia do sinal na saída de cada filtro, iv) cálculo do *loudness* específico e v) integração ao longo da escala de frequências (bark ou ERB).

Zwicker (1984) modelou a integração temporal de *loudness* como o quadripolo da Figura 3.7(a). Se a tensão de entrada for caracterizada por uma função degrau, $C_1 = 0,7\mu\text{F}$ é carregado imediatamente, $C_2 = 1\mu\text{F}$ é carregado numa constante de tempo dependente de $R_2 = 20\text{k}\Omega$ ($\tau_2 = C_2 \times R_2 = 20\text{ ms}$). C_2 é considerado 95% carregado em $3 \times \tau_2 = 60\text{ ms}$ e 99% carregado em $5 \times \tau_2 = 100\text{ ms}$. Se o sinal tem duração inferior a 100 ms, C_2 não carregou completamente e C_1 se descarrega por $R_1 = 35\text{k}\Omega$ e carrega C_2 via R_2 . Se o sinal tem duração superior a 100 ms, C_2 está carregado completamente, e tanto C_1 quanto C_2 se descarregam via R_1 , com uma constante de tempo $\tau_{1+2} = R_1 \times (C_1 + C_2) = 60\text{ ms}$, mais lentamente que no primeiro caso. Um exemplo de mascaramento temporal de rajadas de 5 kHz com 10 e 100 ms é ilustrado na Figura 3.7(b).

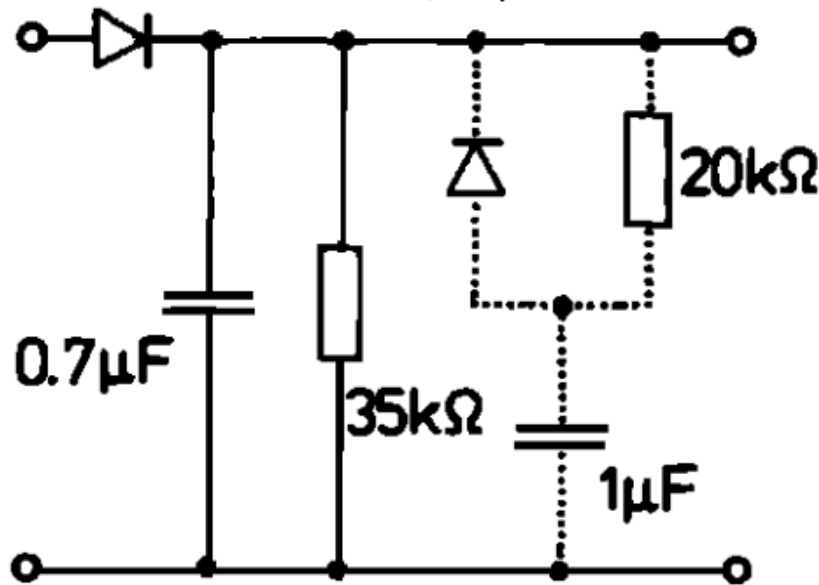
Em todos os modelos, a saída do quinto estágio é chamada de “*loudness* instantâneo”, definido da seguinte forma pelos autores:

Nós supusemos o *loudness* instantâneo como uma variável interferente e indisponível para a percepção consciente. Pode corresponder, por exemplo, à atividade total no nervo auditivo, medido ao longo de um intervalo de tempo muito curto, a exemplo de 1 ms. (GLASBERG; MOORE, 2002, p. 334, tradução minha.)

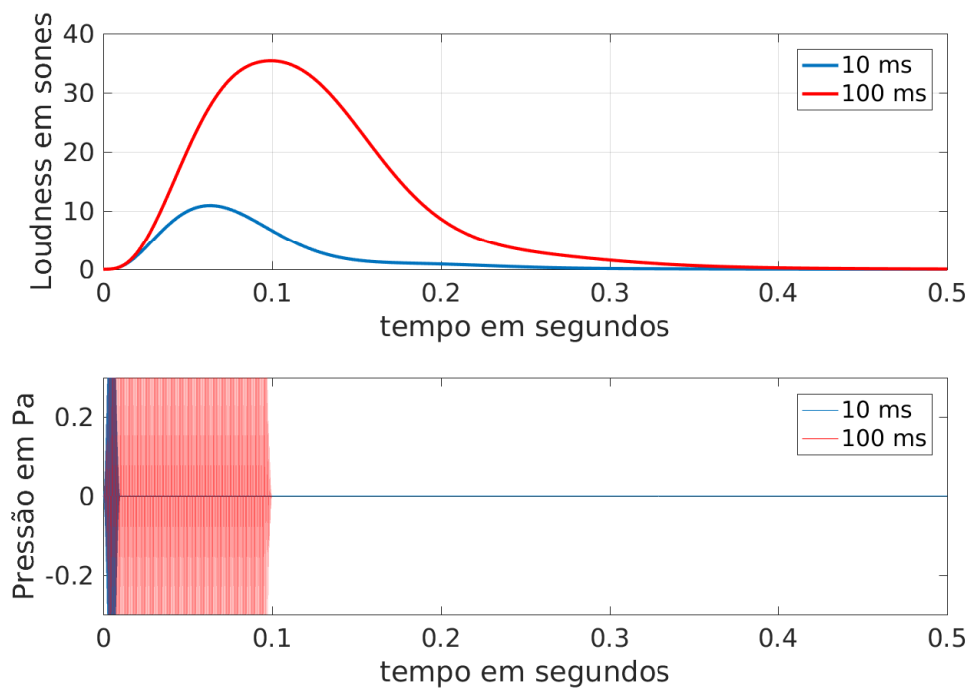
É, na verdade, um vetor temporal que contém valores de *loudness* computados em janela deslizante por períodos de tempo muito curtos: a cada 1 ms no modelo de Glasberg e Moore (2002), a cada 2 ms nos modelo de Zwicker e Fastl (1999) e a cada 4 ms nos modelo de Chalupper e Fastl (2002).

Os indicadores de *loudness* em função do tempo no modelo de Zwicker e Fastl (1999) são da forma Nx , que representa o valor de *loudness* excedido em x por cento do tempo pelos elementos do vetor de *loudness* instantâneo,

Figura 3.7 – (a) Circuito RC simulando a duração do decaimento do loudness específico (b) Processamento de loudness de rajadas de 5 kHz com durações de 10 e 100 ms.



Fonte: Zwicker (1984, Fig. 7a, p. 223).



Fonte: Adaptada de Zwicker e Fastl (2013, Fig. 8.20a e c, p. 229).

sendo $N7$ recomendado para fala, $N5$ para sons ambientais e $N4$ para tráfego automotivo. Já o modelo de Chalupper e Fastl (2002) também é baseado na integração temporal de Zwicker (1984), fornece valores de loudness de curta duração suavizando o vetor de loudness instantâneo por um filtro passa-baixas com frequência de corte de 8 Hz.

O método de [Glasberg e Moore \(2002\)](#) usa dois estágios de integração temporal. O primeiro estágio fornece o *loudness* de curta duração a partir do vetor de medidas de *loudness* instantâneo com a equação de diferenças

$$STL(t_{n+1}) = (1 - \alpha)IL(t_{n+1}) + \alpha STL(t_n), \quad (3.34)$$

onde n é o número do quadro, STL é o indicador “*loudness* de curta duração” e IL é o indicador “*loudness* instantâneo”. O decaimento α é expresso da forma

$$\alpha = \exp\left(\frac{-dt}{\tau_\alpha}\right), \quad (3.35)$$

onde o termo $dt = t_{n+1} - t_n$ é o passo de tamanho igual a 1 ms. Quando o *loudness* incrementa ($IL(t_{n+1}) > STL(t_n)$), $\tau_\alpha = 22$ ms. Quando o *loudness* decrementa ($IL(t_{n+1}) \leq STL(t_n)$), $\tau_\alpha = 50$ ms. Já o segundo estágio de integração calcula o chamado “*loudness* de longa duração” a partir do *loudness* de curta duração pela equação diferença:

$$LTL(t_{n+1}) = (1 - \beta)STL(t_{n+1}) + \beta LTL(t_n), \quad (3.36)$$

onde LTL é o indicador “*loudness* de longa duração”. O decaimento β é expresso da forma

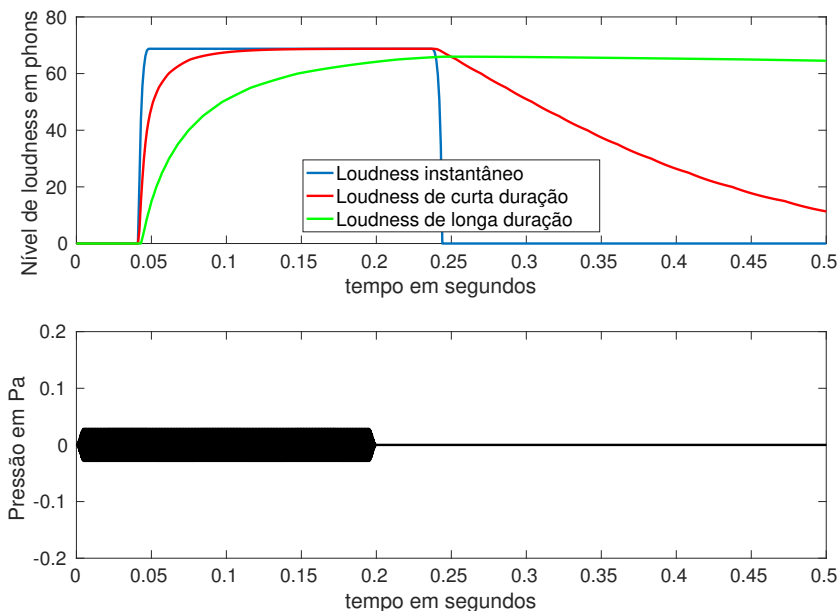
$$\beta = \exp\left(\frac{-dt}{\tau_\beta}\right). \quad (3.37)$$

Quando o *loudness* de curta duração incrementa ($STL(t_{n+1}) > LTL(t_n)$), $\tau_\beta = 100$ ms. Quando o *loudness* de curta duração decrementa ($STL(t_{n+1}) \leq LTL(t_n)$), $\tau_\beta = 2000$ ms. Exemplos da saída do modelo são ilustrados na [Figura 3.8](#).

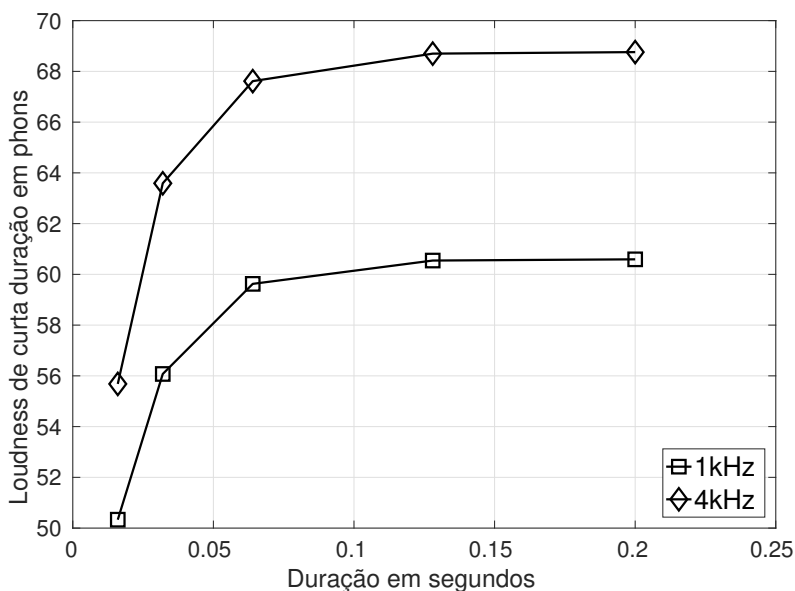
3.2.3 Outros modelos

Dentre os modelos multifaixa analisados nesta seção, foram considerados somente os modelos disponíveis para sons não-impulsivos e apresentados dioticamente, ou seja, com o mesmo sinal chegando aos dois ouvidos. Há modelos multifaixa para sinais dicóticos – com sinais diferentes incidindo em cada ouvido – a exemplo dos modelos de [Moore e Glasberg \(2007\)](#), [Sivonen e Ellermeier \(2008\)](#) e [Glasberg e Moore \(2010\)](#). Estes não foram utilizados como base de comparação com a proposta apresentada no [Capítulo 4](#) e há a pretensão de se estudá-los numa próxima etapa do trabalho.

Figura 3.8 – (a) Saída do modelo de Glasberg e Moore (2002) em resposta a um tom de 4 kHz com duração de 200 ms, ilustrando as curvas de loudness instantâneo, de curta e longa duração. (b) Nível de loudness de curta duração em função da duração de tons de 1 e 4 kHz.



Fonte: Adaptada de Glasberg e Moore (2002, Fig. 5, p. 337).



Fonte: Adaptada de Glasberg e Moore (2002, Fig. 6, p. 338).

O modelo de Chalupper e Fastl (2002) foi mencionado na subseção 3.2.2 por ser uma evolução do modelo de Zwicker *et al.* (1991) no que tange à integração temporal e ao emprego do indicador “loudness de curta de duração” também usado por Glasberg e Moore (2002). Contudo, é um modelo com métricas de perda auditiva voltado para avaliação de Aparelhos de Amplificação Sonora

Individual (AASI) e, portanto, fora do escopo desta pesquisa. Outros modelos não investigados pela mesma razão, porém dignos de nota, são os de [Boulet \(2005\)](#) para sons impulsivos e os modelos auditivos do tipo “fisiológicos”, cujos objetivos focam na compreensão do sistema auditivo periférico e nos quais a intensidade é uma das métricas de avaliação.

3.3 Modelos de Faixa Única

Os modelos de faixa única foram desenvolvidos a partir do entendimento de que se o som medido for de faixa larga, compreendendo quase todo o espectro da audição, não seriam mais necessárias várias linhas de filtragem em bandas críticas, mas sim uma única linha de filtragem, com resposta em frequência compatível com o inverso das curvas de [Fletcher e Munson \(1933\)](#) ilustradas na [Figura 2.4](#). Ruídos de faixa larga, sons ambientais e música dinamicamente comprimida são exemplos de sons cujas raias espectrais, na sua maioria, variam pouco em magnitude e poderiam ter suas intensidades medidas por esta adaptação mais simples do detetor de energia da [seção 3.1](#) denominada “nível sonoro contínuo equivalente”, ou L_{eq} .

3.3.1 Nível sonoro contínuo equivalente (L_{eq})

O valor L_{eq} é um valor médio da energia do sinal durante toda sua duração, ou durante um intervalo de tempo T suficientemente grande para caracterizá-lo. É definido matematicamente por ([BRIXEN, 2011](#)):

$$L_{eq}(W) = 10 \log_{10} \left(\frac{1}{T} \int_0^T \frac{x_W(t)^2}{x_{Ref}(t)^2} dt \right) \text{ dB} \quad (3.38)$$

$$= 20 \log_{10} \sqrt{\frac{1}{T} \int_0^T \left(\frac{x_W(t)}{x_{Ref}(t)} \right)^2 dt} \text{ dB}, \quad (3.39)$$

onde $x_W(t)$ é a pressão sonora ponderada em frequência do sinal medido no instante de tempo t , $x_{Ref}(t)$ é o sinal de referência e a segunda notação da igualdade é uma versão RMS da medida. A ponderação em frequência W corresponde à etapa de filtragem que precede o integrador no modelo de detecção de energia, porém em faixa larga e com resposta em frequência que enfatize

ou penalize partes do espectro de áudio, objetivando levar em consideração o comportamento do ouvido humano sob certas condições de intensidade.

3.3.2 Curvas de ponderação

De modo geral, filtros de ponderação em frequência penalizam as baixas frequências e enfatizam frequências superiores a 1 kHz. Em razão da dependência do *loudness* para com a frequência, seria razoável presumir que a região entre 1 e 4 kHz teria uma ênfase ligeiramente maior que as frequências acima desta faixa. Mas isso não acontece, porque as curvas de ponderação auditivas são consideradas modelos de resposta do ouvido interno, que é considerado igualmente sensível em frequências superiores a 1 kHz. As diferenças entre os contornos de mesmo *loudness* da [Figura 2.4](#) e as curvas de ponderação da [Figura 2.5](#) são compensadas por funções de transferência do ouvido externo/médio contabilizadas no Estágio 1 da [Tabela 3.2](#) para os modelos multifaixa, por exemplo.

A curva A especificada na norma de número 60.651 da [IEC \(1979\)](#) é uma aproximação invertida do contorno de baixa intensidade de 40 *phon* e é utilizada para medidas acústicas de ruído. Sua função de ponderação pode ser escrita em função da frequência como abaixo:

$$W_A(f) = \frac{(12194)^2 \cdot f^4}{(f^2 + (20,6)^2) \sqrt{(f^2 + (107,7)^2) (f^2 + (737,9)^2) (f^2 + (12194)^2)}}, \quad (3.40)$$

e a curva, em decibels, da forma

$$A(f) = 20 \log_{10}(W_A(f)) + 2,00, \quad (3.41)$$

onde o valor 2,00 é um deslocamento (*offset*) para ajustar uma resposta de 0 dB a 1 kHz. As medidas de intensidade ponderadas pela curva A possuem notação descrita com a letra “A” entre parênteses (dB(A) ou $L_{eq}(A)$) ou como índice (L_{Aeq}). Utilizado principalmente em dosímetros de ruído, o método $L_{eq}(A)$ também é usado no medidor de *loudness* LM100 dos Laboratórios [Dolby \(2011\)](#) para diálogos em programação de radiodifusão.

Como visto na [subseção 2.2.2](#), as curvas B e C também são aproximações invertidas dos contornos de 70 e 100 *phon*, para sinais moderados e intensos, respectivamente. A curva B caiu em desuso, mas foi recentemente resgatada com algumas modificações, como será visto mais adiante. Já a curva C continua sendo usada na medição acústica de níveis máximos de pressão sonora em sistemas de Endereçamento ao Público (PA) e em arranjos de alto-falantes para cinema. A norma IEC:60651 ainda cita uma curva D , voltada especificamente para medições de aeronaves. Suas funções de ponderação, curvas em dB e deslocamentos de ajuste são relacionados nas equações abaixo:

$$W_B(f) = \frac{(12194)^2 \cdot f^3}{(f^2 + (20,6)^2) \sqrt{(f^2 + (158,5)^2) (f^2 + (12194)^2)}}, \quad (3.42)$$

$$B(f) = 20 \log_{10}(W_B(f)) + 0,17, \quad (3.43)$$

$$W_C(f) = \frac{(12194)^2 \cdot f^2}{(f^2 + (20,6)^2) (f^2 + (12194)^2)}, \quad (3.44)$$

$$C(f) = 20 \log_{10}(W_C(f)) + 0,06, \quad (3.45)$$

e

$$W_D(f) = \frac{f}{6,8966888496476 \cdot 10^{-5}} \cdot \sqrt{\frac{g(f)}{(f^2 + 79919,29) (f^2 + 1345600)}}, \quad (3.46)$$

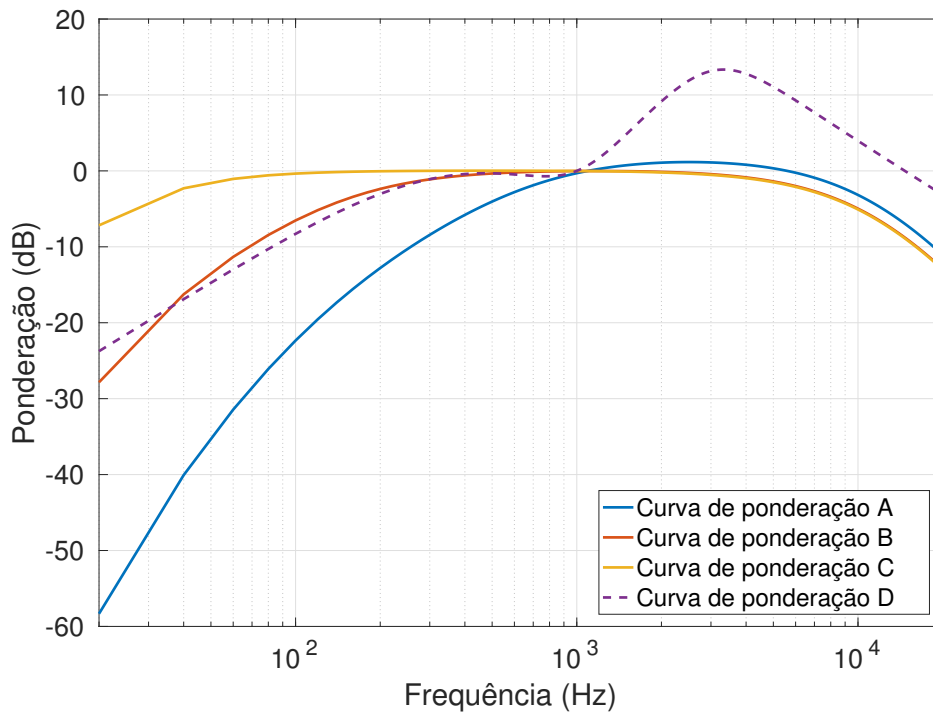
$$D(f) = 20 \log_{10}(W_D(f)), \quad (3.47)$$

onde

$$g(f) = \frac{(1037918,48 - f^2)^2 + 1080768,16f^2}{(9837328 - f^2)^2 + 11723776f^2}. \quad (3.48)$$

Na [Figura 2.5](#), os contornos A, B e C são ilustrados como aproximações de contornos de mesmo *loudness*. A representação em curvas de ponderação é ilustrada na [Figura 3.9](#).

Figura 3.9 – Curvas A, B, C e D de ponderação em frequência constantes da norma IEC:60651.



Fonte: Elaborada pelo autor.

A curva A, enquanto modelo do ouvido interno para um nível de *loudness* de 40 *phon*, não conseguiu captar a sensibilidade a ruído térmico de equipamentos eletrônicos evidenciada no entorno de 6 a 10 kHz (BRIXEN, 2011, p. 64). A ponderação de ruído na norma BS.468 do ITU-R (1986) foi desenvolvida especificamente para este fim. Sua função de ponderação, curva e deslocamento de ajuste, são escritos como nas equações abaixo.

$$W_{BS.468}(f) = \frac{1.246332637532143 \cdot 10^{-4} f}{\sqrt{(g_1(f))^2 + (g_2(f))^2}}, \quad (3.49)$$

$$BS.468(f) = 20 \log_{10}(W_{BS.468}(f)) + 18,2, \quad (3.50)$$

onde

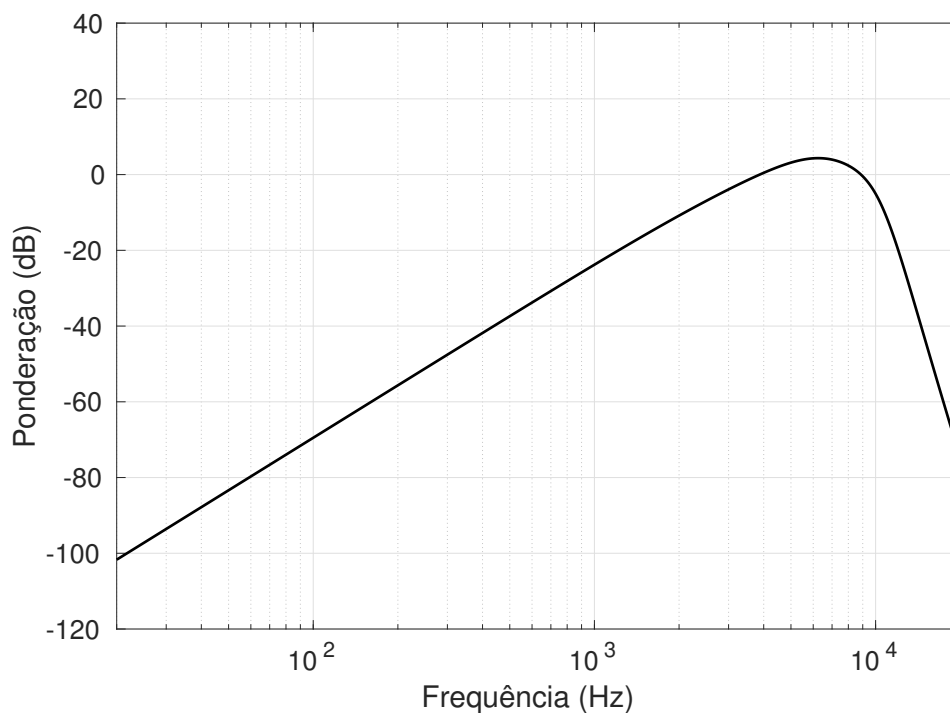
$$g_1(f) = -4,73733898137834 \cdot 10^{-24} f^6 + 2,043828333606125 \cdot 10^{-15} f^4 - 1,363894795463638 \cdot 10^{-7} f^2 + 1$$

e

$$g_2(f) = 1,306612257412824 \cdot 10^{-19} f^5 - 2,118150887518656 \cdot 10^{-11} f^3 + 5,55948023498642 \cdot 10^{-4} f.$$

A curva ITU-R BS.468 é ilustrada na [Figura 3.10](#). Esta foi adaptada no Modelo 737, o primeiro medidor de *loudness* para a indústria dos Laboratórios [Dolby \(1988\)](#). A adaptação consiste num deslocamento da curva em 5,6 dB, para um cruzamento de 0 dB não mais em 1 kHz, mas sim em 2 kHz. A medida com esta ponderação foi denominada posteriormente $L_{eq}(m)$ – onde “m” significa filme (*movie*) – e é usada para medir níveis sonoros (ou irritabilidade) em trilhas de cinema até a presente data.

Figura 3.10 – Curva ITU-R BS.468 de ponderação para medição de intensidade de ruído de equipamentos eletrônicos.



Fonte: Adaptada de [ITU-R \(1986\)](#).

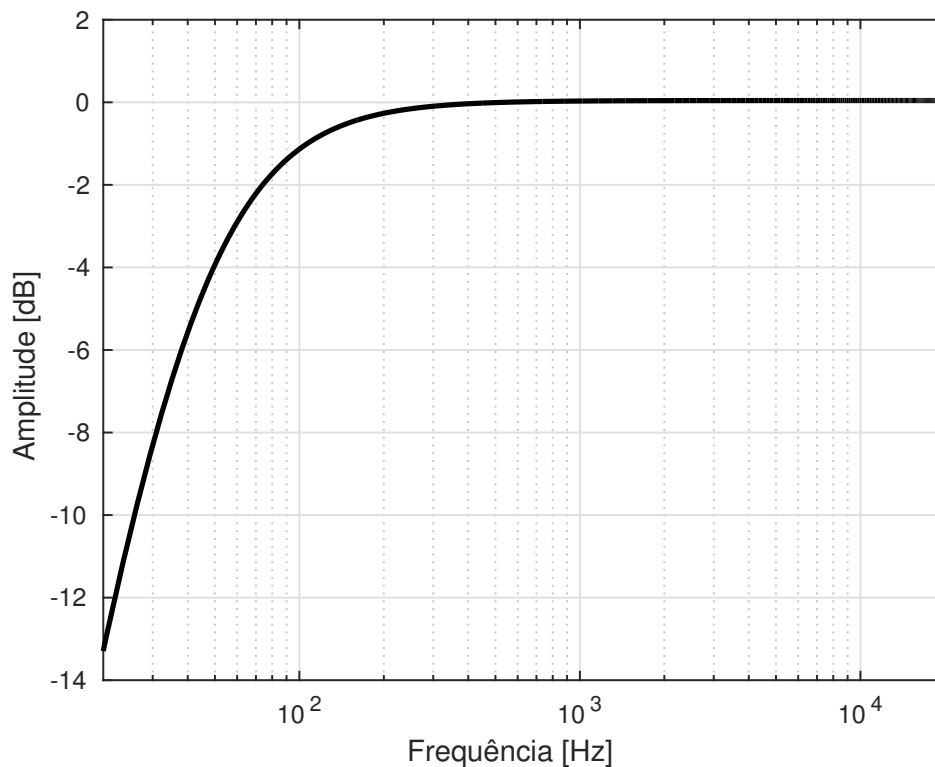
Como visto no [Capítulo 1](#), o método $L_{eq}(RLB)$ foi selecionado pelo setor de Radiocomunicação da ITU numa chamada de algoritmos de medida de *loudness* para radiodifusão em 2003. A curva *RLB* foi proposta por [Soulodre \(2004\)](#) e significa “curva *B* revisada em baixa frequência”. É a mesma curva *B* para valores abaixo de 1 kHz, porém não há qualquer penalização nas frequências

superiores a 1 kHz, comportando-se como um filtro passa-altas. A filtragem *RLB* é feita por um filtro IIR de segunda ordem descrito pela função de tempo discreto abaixo, para uma taxa de amostragem de 48000 amostras por segundo:

$$RLB(z) = \frac{1 - 2z^{-1} + z^{-2}}{1 - 1,99004745483398z^{-1} + 0,99007225036621z^{-2}}, \quad (3.51)$$

com resposta em frequência ilustrada na [Figura 3.11](#).

Figura 3.11 – Curva B revisada em baixa frequência (*RLB*) de ponderação para conteúdo de radiodifusão proposta por [Soulodre \(2004\)](#)



Fonte: Adaptada de [Soulodre e Lavoie \(2005\)](#).

O modelo definitivo de *loudness* adotado pelo [ITU-R \(2015b\)](#), que será detalhado mais à frente no texto, modifica a curva de ponderação *RLB* com a introdução de um pré-filtro de compensação dos efeitos acústicos da cabeça que funcionaria como uma compensação média das HRTFs paramétricas para cinco canais ilustradas na [Figura 2.16\(a\)](#) ([LYMAN; SEEFELDT, 2006](#)). O pré-filtro IIR de dois estágios é descrito pela função de tempo discreto abaixo, para uma

taxa de amostragem de 48000 amostras por segundo:

$$pre(z) = \frac{1,53512485958697 - 2,69169618940638z^{-1} + 1,19839281085285z^{-2}}{-1,69065929318241z^{-1} + 0.73248077421585z^{-2}} \quad (3.52)$$

com resposta em frequência ilustrada na [Figura 3.12\(a\)](#). A combinação das respostas em frequência do pré-filtro e da curva *RLB* foi denominada curva *K*, cuja resposta em frequência é mostrada na [Figura 3.12\(b\)](#).

Existem outras curvas de ponderação, ora mais especializadas, ora caindo no esquecimento. No primeiro grupo encontram-se as curvas psfométricas utilizadas para ponderação em frequência de medidas na telefonia fixa comutada e ponderações para medida de capacidade de alto-falantes. No segundo grupo encontram-se ponderações para medidas de tremores de terra e para reprodução de áudio por gramofones ([BRIXEN, 2011](#), p. 65).

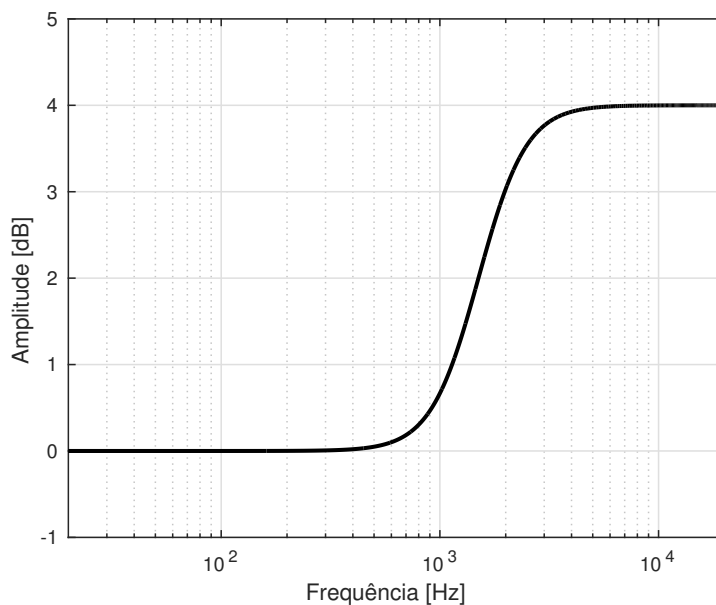
3.3.3 Recomendação ITU-R BS.1770

O modelo de *loudness* de faixa única BS.1770 do [ITU-R \(2015b\)](#) foi o primeiro modelo voltado especificamente para conteúdo multicanal típico de radiodifusão e o primeiro a levar em conta os efeitos espaciais vistos na [seção 2.5](#), dos quais a percepção de intensidade é dependente. Medidas objetivas feitas com este algoritmo basearam uma série de regulamentos regionais de controle de *loudness*, inclusive no Brasil ([MC, 2012](#)). Um diagrama em blocos da forma como o modelo foi originalmente apresentado em 2007 é disposto na [Figura 3.13](#).

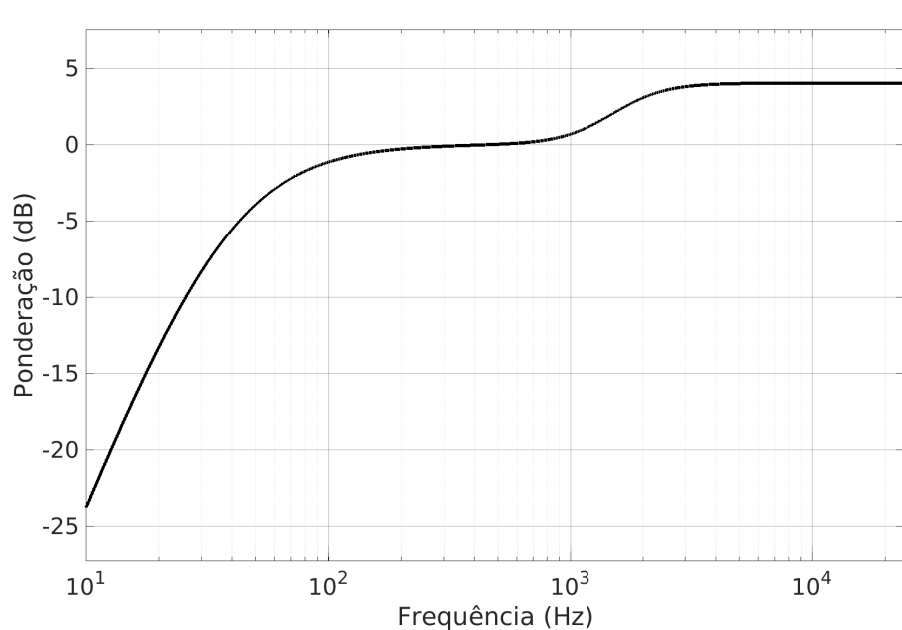
O algoritmo de medida de *loudness* multicanal é composto de quatro estágios:

1. Ponderação em frequência pela curva *K*;
2. Cálculo da média quadrática do sinal em cada canal.
3. Soma ponderada da energia do sinal nos canais, dentre os quais:
 - a) os canais *surround* têm maior peso;
 - b) o(s) canal(is) de Efeitos de Baixa Frequência (LFE) é(são) excluído(s) da soma.
4. Função portão (*gating*) na saída operando com blocos de 400 ms, com sobreposição de 75%, na qual são usados dois limiares:

Figura 3.12 – (a) Resposta em frequência do filtro de compensação dos efeitos acústicos da cabeça como uma média das respostas ao longo dos ângulos de incidência mais comuns num ambiente de escuta multicanal (b) Curva de ponderação K utilizada no modelo de *loudness* ITU-R BS.1770, resultado da combinação das respostas em frequência do pré-filtro e do filtro *RLB*.



Fonte: Adaptada de ITU-R (2015b, Fig. 2, p. 3).

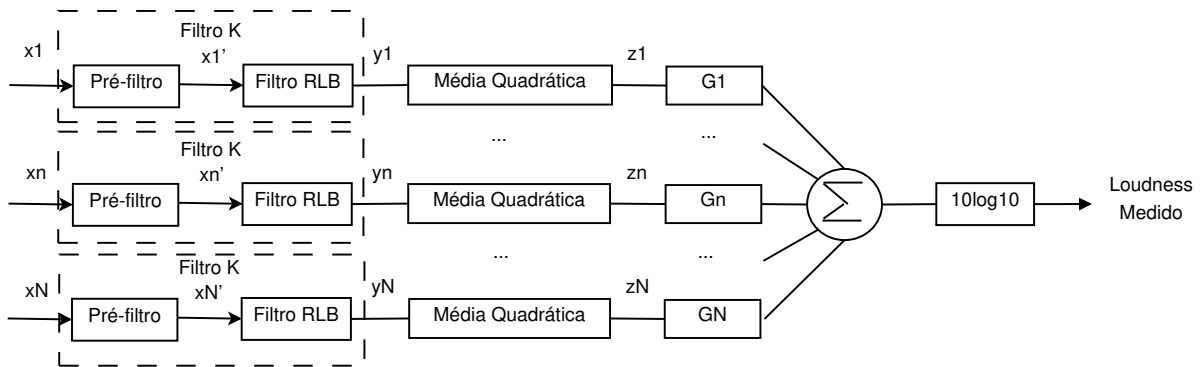


Fonte: Elaborada pelo autor.

- um absoluto a -70 LKFS;
- um relativo, 10 LU abaixo do nível de *loudness* medido após a aplicação do primeiro limiar.

Sobre as unidades de loudness vistas acima, assim como nas demais

Figura 3.13 – Diagrama em blocos do algoritmo de *loudness* multicanal ITU-R BS.1770



Fonte: Adaptada de ITU-R (2015b).

medidas ponderadas, o valor não é lido na medida absoluta em decibels referentes ao fundo de escala digital (dBFS), mas sim acompanhado de um intitutivo que identifica a ponderação em frequência: *Loudness* ponderado pela curva K , referente ao fundo de escala digital de 0 dBFS (LKFS). Medidas relativas são descritas com a *Loudness Unit* (Unidade de *Loudness*) (LU), para as quais $1 \text{ LU} = 1 \text{ dB}$.

No segundo estágio, o nível sonoro contínuo equivalente é medido em cada canal, após ponderação pela curva K no primeiro estágio ($L_{eq}(K)$), expresso pela Equação 3.38 reescrita em atenção aos blocos do diagrama:

$$z_n = L_{eq}(K) |_{\text{Linear}} = \frac{1}{T} \sum_{m=0}^{M-1} y_n^2[m] = \frac{1}{T} \sum_{m=0}^{M-1} x_{nK}^2[m] \quad (3.53)$$

onde $x_{nK}[m]$ é o sinal de entrada ponderado pela curva K e $n \in I = \{L, R, C, L_S, R_S\}$, que seria o conjunto de canais de entrada num exemplo de um sistema de reprodução 5.1⁴ tal como padronizado em (ITU-R, 2012b).

No terceiro estágio, os valores z_n são então ponderados pelos pesos G_n correspondentes aos ângulos de chegada dos sinais oriundos de cada canal, objetivando contabilizar a percepção mais intensa de ondas acústicas que incidam ipsilateralmente ao ouvinte. Caixas acústicas dispostas num azimuth θ tal que $\frac{\pi}{3} \leq |\theta| \leq \frac{2\pi}{3}$, a exemplo das caixas dos canais *surround*, têm pesos $G = \sqrt{2}$ (ou +1,5 dB) e, para os demais azimuthes, $G = 1$ (ou 0 dB). Caso o ângulo de elevação ϕ seja tal que $|\phi| > \frac{\pi}{6}$, todos os ganhos G são unitários.

⁴ Note que o canal de baixa frequência LFE $\notin I$. Argumenta-se que a exclusão se deve ao LFE não ser usado em sistemas domésticos para produção de estéreo a partir do sinal 5.1 (*downmixing*), embora isso esteja aberto à discussão. Mais detalhes em (NORCROSS; LAVOIE, 2009).

As contribuições individuais dos canais são então somadas linearmente para resultar numa medida composta de *loudness* L_K da forma:

$$L_K = -0,691 + 10 \log 10 \sum_n G_n \cdot z_n \text{ LKFS}, \quad (3.54)$$

na qual a constante $-0,691$ é um ganho de calibração aplicado para compensar os efeitos das filtragens, tal que um tom senoidal de teste a 1 kHz com excursão plena na escala digital corresponda a um nível fixo de referência nas curvas de mesmo *loudness*, como 100 *phons* ou 64 *sones* (ZWICKER; FASTL, 2013).

Para o quarto estágio, uma função portão de saída foi adicionada ao algoritmo em virtude da necessidade de um supressor de silêncio – e de níveis considerados muito baixos – no cálculo do *loudness* de programas e comerciais. Um anúncio com poucos segundos muito intensos é suficiente para irritar a audiência, mesmo que no restante do tempo seja fraco o suficiente tal que o nível de *loudness* médio do segmento de propaganda seja inferior a um valor de referência. A ideia da função portão foi proposta originalmente por Skovenborg e Lund (2009), e a estratégia de fechamento e abertura proposta por Grimm, Skovenborg e Spikofski (2010) foi incorporada na Recomendação em 2011.

No cálculo de um nível de *loudness* dito *entrecortado* (*gated loudness*), a duração T do áudio é dividida num conjunto de intervalos de blocos de fechamento com sobreposição. Cada bloco de amostras tem duração $T_g = 400$ ms. A sobreposição de cada bloco é de 75% de sua duração. A Equação 3.53 é então reescrita com os intervalos de integração limitados do início ao fim de cada bloco, sendo que blocos incompletos ao final do intervalo total de medição são descartados. O $L_{eq}(K)$ do j -ésimo bloco de fechamento no n -ésimo canal de entrada no intervalo T , na forma integral, é dado por (ITU-R, 2015b):

$$z_{jn} = \frac{1}{T_g} \int_{T_g \cdot j \cdot \text{passo}}^{T_g \cdot (j \cdot \text{passo} + 1)} y_n^2 dt, \quad (3.55)$$

onde $\text{passo} = 1 - \text{sobreposição}$ e

$$j \in \left\{ 0, 1, 2, \dots, \frac{T - T_g}{T_g \cdot \text{passo}} \right\}. \quad (3.56)$$

O *loudness* entrecortado do j -ésimo bloco l_j é calculado usando a Equação 3.54. Dado um limiar de fechamento Γ , há um conjunto de índices de blocos

$J_g = \{j | l_j > \Gamma\}$, ou seja, nos quais o *loudness* do bloco é superior ao limiar de fechamento.

Sendo o número de elementos em J_g igual a $|J_g|$, o *loudness* entrecortado L_{KG} no intervalo T de duração do áudio é dado por:

$$L_{KG} = -0,691 + 10 \log_{10} \sum_n G_n \cdot \left(\frac{1}{|J_g|} \cdot \sum_{J_g} z_{nj} \right) \text{ LKFS.} \quad (3.57)$$

O cálculo do *loudness* entrecortado é feito usando dois limiares: um absoluto e outro relativo (ITU-R, 2015b). O limiar relativo Γ_r é calculado após a medida de *loudness* entrecortado com o limiar absoluto, $\Gamma_a = -70$ LKFS, e subtraindo 10 LU do resultado, da forma:

$$\Gamma_r = -0,691 + 10 \log_{10} \sum_n G_n \cdot \left(\frac{1}{|J_g|} \cdot \sum_{J_g} z_{nj} \right) - 10 \text{ LKFS,} \quad (3.58)$$

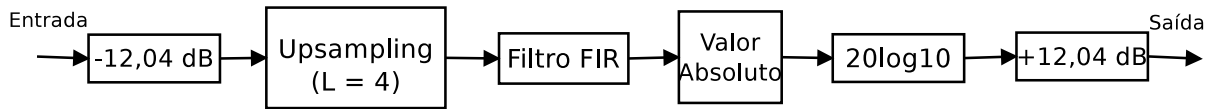
onde $J_g = \{j | l_j > \Gamma_a\}$.

A terceira versão da Recomendação ITU-R BS.1770 incorporou um algoritmo de medida de “pico verdadeiro” em resposta à necessidade de se abrir mão do controle da saturação pelos picos das amostras lidos em um Medidor de Picos de Programação (PPM), tradicional na radiodifusão. O problema foi apontado por Thomas Lund (2006), que afirmou não haver garantias de que um sinal com excursão dinâmica totalmente contida na escala digital teria seus níveis de pico adequadamente restaurados na reconstrução do sinal analógico. Após etapas de processamento que produzam alterações no sinal, amostras do sinal de saída poderão sofrer ceifamento (*clipping*) durante a reprodução mesmo que as amostras do sinal original estejam em escala.

De modo a evitar uma medida de pico sem efeito, seria necessário sobreamostrar o sinal digital com o objetivo de recriar o sinal original perdido entre as amostras, e então medi-lo. A implementação do medidor de pico verdadeiro por Pires (2014), utilizada mais à frente na seção 4.2, se orienta pelo diagrama em blocos da Figura 3.14. A unidade de medida é designada Pico verdadeiro em decibels relativos a 100% da escala digital (dBTP).

Para estimação da nova taxa de amostragem necessária, imagina-se o pior caso no qual a leitura de um pico no sinal sobreamostrado ocorra entre amostras

Figura 3.14 – Diagrama em blocos do medidor de pico verdadeiro conforme especificações no Anexo 2 da Rec. ITU-R BS.1770-3



Fonte: Adaptada de ITU-R (2015b, Anexo 2).

equidistantes da ocorrência do pico verdadeiro de uma senoide oscilando na frequência de Nyquist normalizada. Sejam L o fator de sobreamostragem, f_s a frequência de amostragem e f_n a frequência normalizada; o novo período de amostragem seria $\frac{1}{Lf_s}$ e o período referente à máxima frequência normalizada seria $\frac{1}{f_n f_s}$. O erro máximo de leitura seria então (DASH, 2014):

$$\text{Erro}_{\text{MAX}} = 20 \log \left(\cos \left(\frac{2\pi f_n f_s}{2Lf_s} \right) \right) = 20 \log \left(\cos \left(\frac{\pi f_n}{L} \right) \right), \quad (3.59)$$

onde o número 2 no denominador indica que o teto de erro é, pelo critério de Nyquist, o dobro do período de amostragem. A Equação 3.59 foi testada na terceira versão da Recomendação para diferentes valores de L e f_n . Na ocasião, verificou-se que para $L = 4$ e $f_n = 0,45$, esta resultaria em $\text{Erro}_{\text{MAX}} = 0,554$ dB e, para $f_n = 0,5$ (frequência de Nyquist), resultaria em $\text{Erro}_{\text{MAX}} = 0,688$ dB. Portanto, numa relação de compromisso entre a complexidade do medidor e erros de medida inferiores a 1 dB, padronizou-se um fator mínimo de sobreamostragem $L = 4$, podendo ser superior conforme decisão do projetista.

O valor de atenuação/ganho de 12,04 dB advém da razão sinal-ruído do quantizador de $(B + 1)$ bits (OPPENHEIM; SCHAFER, 2013, p. 196). Seja FS o valor total da escala (*full-scale*), a relação entre a variância do sinal σ_x^2 e a do ruído σ_n^2 é dada por:

$$\text{SNR} = 10 \log_{10} \left(\frac{\sigma_x^2}{\sigma_n^2} \right) = 6,02B + 10,8 - 20 \log_{10} \left(\frac{\text{FS}}{\sigma_x} \right). \quad (3.60)$$

Se o sinal possui uma boa excursão de faixa dinâmica, o termo $-20 \log_{10} \left(\frac{\text{FS}}{\sigma_x} \right)$ da Equação 3.60 é desprezível. Nestas condições, pode-se dizer que a razão sinal-ruído aumenta em aproximadamente 6,02 dB por bit. Portanto, uma atenuação de 12,04 dB para provimento de margem de excursão às etapas seguintes de processamento equivale a um deslocamento de dois bits para implementações

em aritmética de ponto fixo, sendo desnecessária para cálculos em aritmética de ponto flutuante.

Por fim, a quarta e mais recente versão da Recomendação ITU-R BS.1770 acomodou uma extensão do algoritmo para sistemas avançados de reprodução com mais de cinco canais. A proposta de reformulação dos pesos G_n feita por Komori *et al.* (2015) se baseou nos trabalhos de Sivonen e Ellermeier (2008) sobre *loudness* direcional, nos quais a intensidade percebida poderia ser aproximada somando-se 3 dB aos níveis de pressão sonora incidentes nos ouvidos esquerdo e direito.

3.3.4 Recomendação EBU R.128

A Recomendação BS.1770 do ITU-R (2015b) tratou especificamente do modelo de *loudness* médio sem sugerir valores de referência para normalização do áudio, ou definir descritores auxiliares de *loudness* para sinais não estacionários como feito nos modelos multifaixa.

Após participar ativamente na elaboração da medida do *loudness* entrecortado, e da sua conseqüente inclusão na segunda versão da Rec. ITU-R BS.1770 em 2011, a União Europeia de Radiodifusão (EBU) publicou a segunda versão de uma recomendação de *loudness* própria para seus membros designada R 128. Neste documento, a EBU recomenda a medida do “*Loudness* Médio” de uma programação para a normalização de sinais de áudio, e a medida de “Nível Máximo de Pico Verdadeiro” para conformidade com os limites técnicos de toda a cadeia de distribuição de sinal na radiodifusão. Adicionalmente, as medidas de “Faixa de *Loudness*”, de “Máximo *Loudness* Momentâneo” e de “Máximo *Loudness* de Curta Duração” podem ser usadas para uma melhor caracterização do sinal de áudio, assim como para satisfazer as necessidades estéticas de cada programa/estação dependendo dos gêneros veiculados, da audiência alvo e da plataforma de distribuição (EBU, 2014).

Um caderno suplementar da recomendação (EBU, 2016b) sobre especificações de um medidor de *loudness* define o *loudness* momentâneo como um vetor de medidas não fechadas (*ungated*) de *loudness* conduzidas numa janela retangular deslizante de 400 ms. O conceito é o mesmo definido por Glasberg e

Moore (2002) na subseção 3.2.2, e o tamanho da janela é ligeiramente superior aos intervalos de integração temporal de *loudness* no levantamento feito por Scharf (1978), sugerindo que o *loudness* medido com este tamanho de janela estaria livre dos efeitos de nível de pressão sonora nas durações críticas elencadas na Tabela 2.1. A recomendação BS.1771, publicada pelo ITU-R (2012a) com os mesmos objetivos da EBU R 128, prescreveu a introdução de um filtro IIR de primeira ordem na saída do medidor, descrito com equação de diferenças com a mesma formulação da Equação 3.34 e da Equação 3.36 como no modelo de Chalupper e Fastl (2002), porém com uma única constante de tempo de descida de 400 ms, tal que um medidor em tempo de execução pudesse se comportar de modo similar a um VU, cuja constante de tempo de descida é de 300 ms. O vetor de medidas de *loudness* momentâneo sem o filtro suavizador de saída é o mesmo *loudness* de bloco calculado pela Equação 3.54 e utilizado no cálculo do *loudness* entrecortado pela Equação 3.57.

O vetor de *loudness* de curta duração também é calculado pela Equação 3.54, porém numa janela deslizante retangular com duração de 3 segundos. Este tamanho de janela foi escolhido por ter sido o menor tamanho possível de suavização do vetor de *loudness* momentâneo de modo a casá-lo com testes subjetivos nos quais os participantes ajustaram controles de volume dinamicamente conforme a variação de *loudness* de um segmento de áudio (LAVOIE; SOULODRE, 2006). Então, se o vetor de *loudness* momentâneo é adequado a um display digital ou à emulação de um galvanômetro (a exemplo do VU), o vetor de *loudness* de curta duração é próprio para a construção de gráficos de variação de *loudness* ao longo do tempo, tal como no modelo de Glasberg e Moore (2002), ainda que com constantes de tempo diferentes.

Já o descritor Faixa de Loudness (LRA) foi definido em caderno suplementar específico (EBU, 2016c). A medida quantifica a variabilidade de uma medida de *loudness* feita num sinal não estacionário, e é calculada usando medidas de *loudness* entrecortado, em blocos de fechamento de mesmo tamanho $T_g = 3\text{ s}$ da janela deslizante do *loudness* de curta duração, sobrepostos em 66% (2 segundos), com um limiar absoluto de -70 LKFS e um limiar relativo de -20 LU. Os testes que resultaram no tamanho do bloco e no percentual de sobreposição foram feitos pelo Centro de Pesquisas em Comunicações do Canadá e relatados

por Esben Skovenborg (2012a) da TC Electronic, que disponibilizou a função MATLAB[®] utilizada no artigo, reproduzida no [Código-fonte 1](#).

Código-fonte 1: Código fonte cedido pela TC Electronic para o cálculo da LRA (SKOVENBORG, 2012a)

```

1 % A MATLAB FUNCTION TO COMPUTE LOUDNESS RANGE
2 % -----
3 function LRA = LoudnessRange( ShortTermLoudness )
4
5 % Input: ShortTermLoudness is a vector of loudness levels,
6 % computed as specified in ITU-R BS.1770 without gating, using
7 % a sliding analysis-window of length 3 s, overlap >= 2 s
8
9 % Constants
10 ABS_THRES = -70 ; % LUFS (= absolute measure)
11 REL_THRES = -20; % LU (= relative measure)
12 PRC_LOW = 10; % lower percentile
13 PRC_HIGH = 95; % upper percentile
14
15 % Apply the absolute-threshold gating
16 abs_gate_vec = (ShortTermLoudness >= ABS_THRES);
17 % abs_gate_vec is indices of loudness levels above
18 % absolute threshold
19 stl_absgated_vec = ShortTermLoudness(abs_gate_vec);
20 % only include loudness levels that are above gate threshold
21
22 % Apply the relative-threshold gating
23 % (non-recursive definition)
24 n = length(stl_absgated_vec);
25 stl_power = sum(10.^(stl_absgated_vec./10))/n;
26 % undo 10log10, and calculate mean
27 stl_integrated = 10*log10(stl_power); % LUFS
28 rel_gate_vec = (stl_absgated_vec >= stl_integrated + REL_THRES);
29 % rel_gate_vec is indices of loudness levels above
30 % relative threshold
31 stl_relgated_vec = stl_absgated_vec( rel_gate_vec );
32 % only include loudness levels that are above gate threshold
33
34 % Compute the high and low percentiles of the
35 % distribution of values in stl_relgated_vec
36 n = length(stl_relgated_vec);
37 stl_sorted_vec = sort(stl_relgated_vec);

```

```
38 % sort elements in ascending order
39 stl_perc_low = stl_sorted_vec(round((n-1)*PRC_LOW/100 + 1));
40 stl_perc_high = stl_sorted_vec(round((n-1)*PRC_HIGH/100 + 1));
41
42 % Compute the Loudness Range measure
43 LRA = stl_perc_high - stl_perc_low; % in LU
```

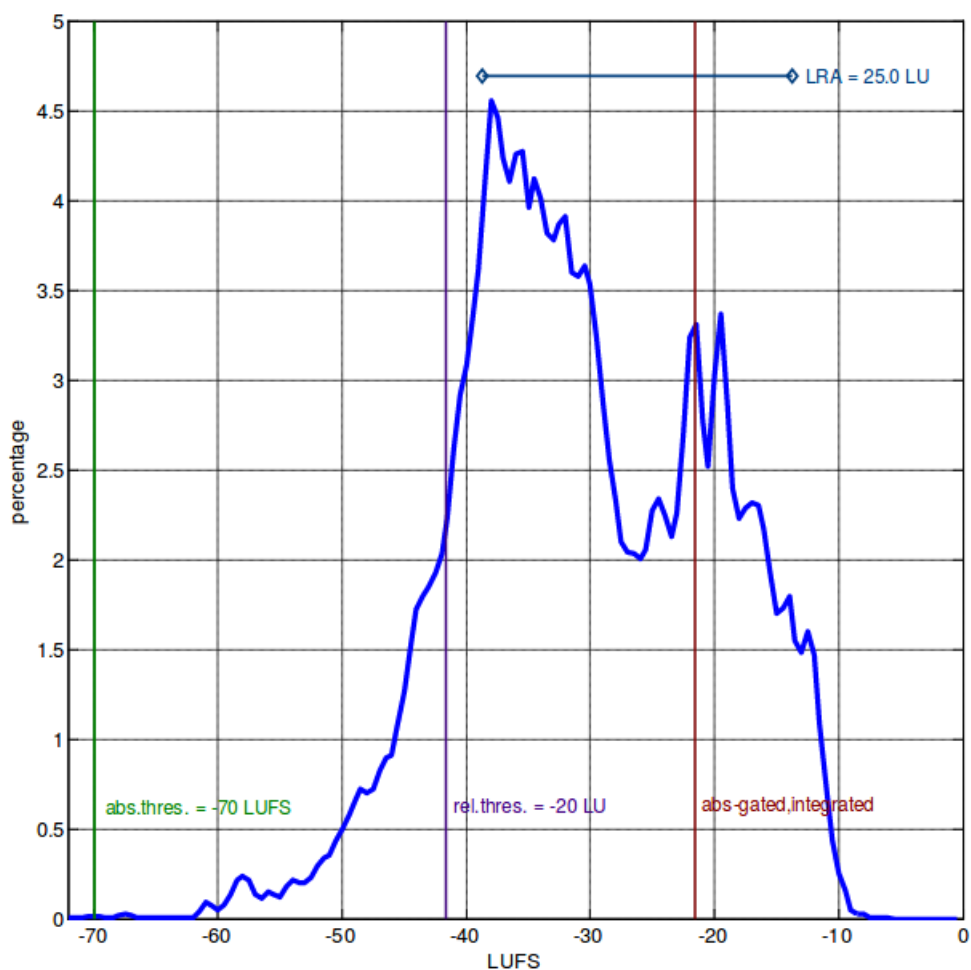
Considerando uma janela deslizante de 3 segundos, cumpre notar que medidas em peças de áudio curtas (a exemplo de peças publicitárias de 30 segundos), ou com trechos de silêncio, podem levar a LRA a ter valores não significativos estatisticamente. Eliminados silêncio, ruído de fundo e passagens com nível de *loudness* muito baixo, a distribuição dos níveis de *loudness* restantes é quantificada não na forma de um histograma, mas na forma de faixa de percentis. A faixa de *loudness* é calculada pela diferença entre o 10^o e o 95^o percentis da distribuição, como justificado no caderno suplementar.

O percentil mais baixo de 10% pode, por exemplo, prevenir que o desvanecimento (*fade-out*) de uma trilha musical domine a Faixa de *Loudness*. O percentil superior de 95% garante que um som muito intenso de caráter único e incomum, como um tiro de arma de fogo num filme, possa ser inteiramente responsável por uma faixa larga de *loudness* (EBU, 2016c, p. 6).

Um exemplo de distribuição de *loudness* é ilustrado na [Figura 3.15](#). O limiar absoluto (*abs.thres.*) está fixado em -70 LKFS. O *loudness* entrecortado medido com este limiar (*abs-gated,integrated*) resultou em $-21,6$ LKFS e o limiar relativo (*rel.thres.*) é situado a 20 LU abaixo em $-41,6$ LKFS. A faixa de *loudness* resultante (LRA = 25,0 LU) é delimitada entre o décimo e o nonagésimo quinto percentis da distribuição de níveis de *loudness* superiores ao limiar relativo.

Os valores alvo recomendados para programação contínua são um *loudness* médio de -23 LKFS, com uma tolerância de $\pm 0,5$ LU para programação de estúdio e de ± 1 LU para programas ao vivo, e um nível máximo de pico verdadeiro de -1 dBTP em todo o segmento. Valores de faixa de *loudness* não são recomendados propriamente, mas as práticas de produção trabalham com LRA = 20 LU para programação *surround* e LRA = 15 LU para programação em estéreo (EBU, 2016d).

Figura 3.15 – Distribuição de *Loudness* com limiares de fechamento e de Faixa de *Loudness* para o filme “Matrix” em masterização para DVD

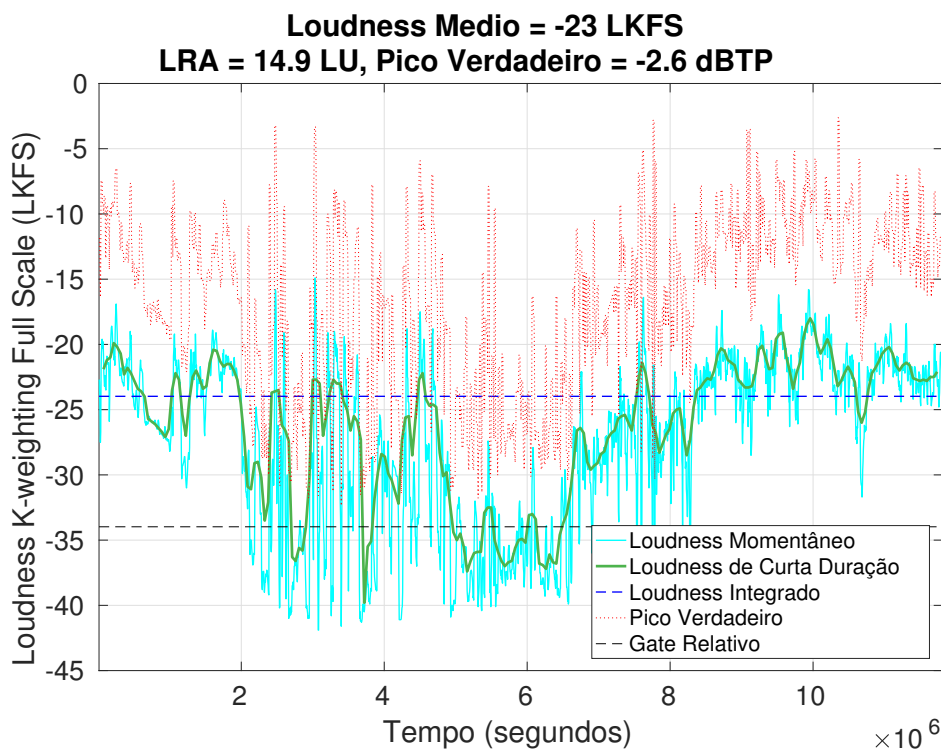


Fonte: EBU (2016c, Fig.1, p. 7).

Na Figura 3.16 há um exemplo de medidas consolidadas de *loudness* em um segmento de teledramaturgia com faixa dinâmica larga. O segmento possui um *loudness* médio de -23 LKFS e uma faixa de *loudness* de $14,9$ LU, caracterizando-o como o caso desejável para programas de televisão em estéreo. Note que o pico verdadeiro do sinal teve bastante margem de excursão, atingindo um valor máximo de $-2,6$ dBTP, seguramente distante da saturação e consequente ceifamento (*clipping*).

Já na Figura 3.17 têm-se as mesmas medidas consolidadas de *loudness* agora executadas numa trilha musical hiper-comprimida dinamicamente. Um *loudness* médio de -10 LKFS sugere que esta canção é percebida como tendo mais que o dobro da intensidade do segmento medido na Figura 3.16, se fizermos uma aproximação da Lei de Potência de Stevens com $n = 0,3$ ($10^{0,3} \approx 2$).

Figura 3.16 – Medidas consolidadas de *loudness* conforme Rec. ITU-R BS.1770-4 para um segmento de teledramaturgia com faixa dinâmica larga.



Fonte: Elaborada pelo autor.

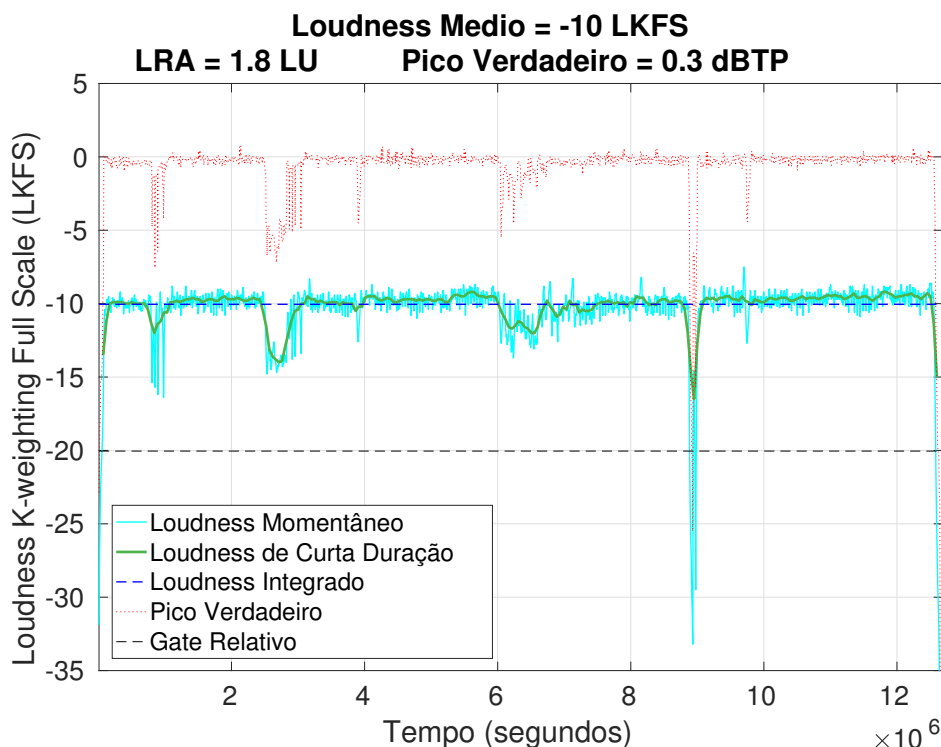
Nota – Este arquivo é identificado como “Caso de Testes número 7” na *suite* de testes da EBU para conformidade de medidores de *loudness*.

Uma faixa de *loudness* de 1,8 LU sugere que a variação de intensidade percebida deu-se pouco acima da diferença no limite do observável (JND) típica de 1 dB. Consequentemente, o pico verdadeiro do sinal quase não excursiona, beirando o fundo de escala quase que em toda a duração da música e até mesmo ultrapassando-o em alguns pontos, pois a leitura máxima foi de 0,3 dBTP. Muito provavelmente um examinador atento da forma de onda não encontraria pontos de ceifamento em taxas de amostragem de 44.100 ou 48.000 amostras por segundo. Mas com uma sobreamostragem de fator $L = 4$, picos situados entre as amostras originais podem atingir o fundo de escala e, na reconstrução do sinal analógico, resultarem em cliques durante a reprodução.

Outros Métodos

As Recomendações ITU-R BS.1770 e EBU R.128 tornaram-se padrões *de facto* – e *de jure* em alguns países – para a radiodifusão digital, pouco a

Figura 3.17 – Diagrama em blocos do medidor de pico verdadeiro conforme especificações no Anexo 2 da Rec. ITU-R BS.1770-3



Fonte: Elaborada pelo autor.

Nota – Este arquivo é de uma trilha musical não licenciada comercialmente e composta em *home studio*.

pouco influenciando a produção e normalização de áudio para outros veículos de distribuição (cinema, *streaming*, etc.), e tornando obsoletas algumas estratégias de normalização que se destacaram em suas épocas. Destacam-se as ferramentas *ReplayGain* e *Dialnorm*.

ReplayGain (ROBINSON, 2001), proposto originalmente em 2001 e ainda embarcado em muitos produtos, calcula um ganho que ajuste o *loudness* de uma música para casar com o *loudness* do ruído rosa em -20 dBFS quando reproduzido a 83 dB SPL ponderado pela curva *C*. A intenção original seria embarcar o valor desejado do *ReplayGain* como metadado de um arquivo de música. Seu procedimento consiste na ponderação em frequência, no cálculo dos valores RMS em blocos de 50 ms e a partir de um histograma, calcular o valor de *loudness* como sendo o excedido por 5% dos blocos em uma trilha. O algoritmo *Sound Check* da *Apple*, disponível em iPods® e iPads®, é uma adaptação do *ReplayGain*.

Dialogue Normalization, ou simplesmente *Dialnorm* (KROPUENSKE,

2007), é o parâmetro de metadados que controla o ganho de reprodução de áudio no sistema de compressão de áudio AC-3 dos laboratórios Dolby. Estes recomendavam que o valor de *dialnorm* fosse determinado medindo-se a trilha de diálogo ao longo do programa como uma soma de potências dos canais, ponderadas pela curva A. Todavia, verificou-se que a curva A não era adequada para ponderar medidas de *loudness* em conteúdos musicais (NORCROSS; SOULODRE; LAVOIE, 2003), e a recomendação caiu em desuso. Um estudo recente (SKOVENBORG; LUND, 2013) comparou a normalização de diálogos com a normalização de *loudness* no padrão ITU de filmes em *home video* e observou que esta última normalizava a níveis menores e com uma maior margem de excursão. Não só as diferenças de normalização eram de 5,5 dB em média, como também os valores alvo de *dialnorm* embarcados nos metadados das mídias de reprodução eram descasados dos valores efetivamente medidos.

Considerações

Neste terceiro capítulo, foi apresentada a arborescência de classificação dos modelos de *loudness*, bem como o estado da arte em suas implementações, com especial destaque para o método de faixa única ITU-R BS.1770 e seus descritores associados, plenamente absorvidos pelo mercado de radiodifusão e pela produção de conteúdo de áudio digital para veiculação neste meio, com perspectiva de incorporação progressiva a melhores práticas de distribuição de conteúdo para outras plataformas em operação, a exemplo do SeAC e de serviços de *streaming* via Internet. A estabilidade do método ITU-R na radiodifusão sugere que sua adoção em outras mídias dê-se de forma natural por meio de recomendações setoriais, a exemplo da Recomendação de *loudness* TD1004.1.15-10 da AES (2015) para *streaming* de áudio e reprodução de arquivos em rede, cabendo somente acompanhamento dos órgãos reguladores quanto a possíveis alterações em regulamentos técnicos e coibições de excessos quando necessário.

Em se tratando de contribuições científicas para a avaliação de *loudness* em áudio digital, deve-se portanto mover o foco das plataformas de distribuição e olhar os caminhos da pesquisa recente em *loudness*, assim como a evolução dos próprios modelos de definição de áudio espacial. Estes são os objetos do [Capítulo 4](#).

PROPOSTAS E EXPERIMENTOS PRELIMINARES

Os três capítulos anteriores delimitaram o arcabouço teórico necessário para que se pudesse perseguir com o propósito estabelecido na [seção 1.4](#) de se aprimorar o modelo BS.1770 de modo a habilitá-lo para medir *loudness* em sistemas de distribuição de áudio imersivo.

Este capítulo, por sua vez, está organizado de modo que sua primeira seção elenque os desenvolvimentos mais recentes em *loudness* em conteúdo de radiodifusão conforme a consolidação do padrão ITU-R na indústria, de modo a vislumbrar-se o horizonte para o qual a pesquisa de *loudness* está caminhando. As seções seguintes contemplarão os dois primeiros experimentos conduzidos para investigação dos problemas elencados na [seção 1.4](#). O primeiro experimento propõe um controlador de *loudness* em conteúdo de formato curto para radiodifusão baseado em critérios complementares aos adotados na regulamentação brasileira. A proposta seguinte é a de um modelo de *loudness* para áudio imersivo baseado em canais empregando técnicas de auralização.

4.1 Loudness na Radiodifusão Digital

4.1.1 Consolidação do padrão ITU-R

Nos primeiros anos que se sucederam à publicação da primeira versão da Recomendação ITU-R BS.1770, em 2006, muito se trabalhou em prol de melhorias e aprimoramento do algoritmo. No mesmo ano, Lyman e Seefeldt (2006), dos laboratórios Dolby, propuseram um novo modelo multifaixa baseado nos modelos de Glasberg e Moore (2002) e de Zwicker *et al.* (1991), e uma extensão do algoritmo ITU-R conjugando o pré-filtro com o uso de uma função de transferência relativa à cabeça por canal. O estudo observou que modelos multifaixa combinados com modelos binauriculares obtiveram melhores resultados, mas o algoritmo ITU-R ainda detinha a melhor razão desempenho/simplicidade. No mesmo ano, Lavoie e Soulodre (2006) do Centro de Pesquisas em Comunicações do Canadá (CRC) investigaram a adaptação do $L_{eq}(RLB)$ a integrações de curta duração e, em comparação com testes subjetivos, uma janela deslizante de 3 segundos obteve o melhor desempenho na suavização do vetor de *loudness* momentâneo, como visto na subseção 3.3.4. Contudo, os resultados desses mesmos testes subjetivos indicaram que ouvintes tendem a responder mais rapidamente a mudanças maiores de intensidade, e esta constatação daria suporte ao argumento de que um medidor de *loudness* de curta duração deveria ter uma resposta não-linear.

Em 2007, uma parceria da TC Electronic com a Universidade McGill propôs janelas instantâneas (500 ms) e de curta duração (2,5 s), além de um descritor estatístico denominado “Medida de Consistência”, projetado para indicar variações de *loudness* intrínsecas a um único programa e relacioná-las a uma tabela de adequação de faixa dinâmica ao veículo de conteúdo (LUND, 2007). Um protótipo com estas características foi apresentado por Skovenborg e Nielsen (2007). Na convenção da AES daquele ano, Alessandro Travaglini (2007) apresentou um *case* de controle de *loudness* da Fox Internacional / Sky Italia com reconfigurações de processadores de áudio em razão do controle de *loudness* com o algoritmo ITU, resultando em ganhos de qualidade e redução das reclamações de usuários.

Dois estudos importantes destacaram-se em 2008: o primeiro foi de autoria

da Empresa de Radiodifusão da Austrália (ABC) em conjunto com a Universidade de Sydney. [Cabrera, Miranda e Dash \(2008\)](#) conduziram um teste subjetivo para confirmar os ganhos de espacialidade e testar a ponderação em frequência do algoritmo ITU, mas desta vez com ruídos com larguras de banda de uma oitava. Os autores concluíram que tanto a ponderação da curva K quanto os pesos G_{Ls} e G_{Rs} destoavam dos resultados dos testes subjetivos nas duas oitavas abaixo de 250 Hz e propuseram filtros de correção da ponderação para baixas frequências. Já o segundo foi conduzido pela rede de radiodifusão norte-americana CBS, que preparou um extenso relatório sobre medida, análise e controle de *loudness* na sua cadeia de distribuição apresentado ao grupo de trabalho de produção de conteúdo da [CBS \(2008\)](#). A sonora conclusão em relação aos testes com *Dialnorm* de que a ferramenta “não é de muita ajuda quando o problema está em equalizar *loudness* nas transições entre programas e comerciais” decretou um ponto final para o parâmetro nas discussões de *loudness* que se seguiriam no âmbito do ITU-R. Por fim, [Skovenborg e Lund \(2008\)](#) apresentaram o produto final da TC Electronic baseado na Rec. ITU-R BS.1770, que promoveu o descritor de longa duração denominado “Medida de Consistência”, e [Sivonen e Ellermeier \(2008\)](#) apresentaram um novo modelo multifaixa de *loudness* como uma alternativa ao pré-filtro do algoritmo multicanal ITU-R na forma de uma HRTF resultante de uma soma de energia das pressões sonoras em ambos os ouvidos.

No ano seguinte, as discussões quanto ao desempenho do algoritmo de [\(ITU-R, 2015b\)](#) em baixas frequências ganham corpo. Ian [Dash \(2009\)](#), da emissora australiana ABC, dá continuidade ao estudo de [Cabrera, Miranda e Dash \(2008\)](#) e executa uma análise em faixas de oitavas na *suite* de testes de conformidade de medidores pela Rec. BS.1771 do [ITU-R \(2012a\)](#) para testar a audibilidade do conteúdo em baixas frequências e obter dados para uma análise de regressão, com o objetivo de encontrar coeficientes ótimos de ponderação para cada oitava. Os modelos multifaixa resultantes obtiveram uma correlação melhor com os dados subjetivos se comparados à curva K do método BS.1770. Por outro lado, [Norcross e Lavoie \(2009\)](#) do CRC analisaram os prós e contras da inclusão do canal reforçador de baixas frequências (LFE) no cálculo do *loudness*, e testes subjetivos indicaram uma pequena contribuição do canal, porém mensurável. Contudo, para a inclusão do LFE no algoritmo ITU, seria preciso definir um ganho para o canal, bem como filtrá-lo numa frequência de corte de 120 Hz

para normalizar o limite de banda em todos os dispositivos de reprodução. Já [Skovenborg e Lund \(2009\)](#), da TC Electronic, apresentaram uma evolução de seus descritores. A “Medida de Consistência” se torna “Faixa de *Loudness*” e passa a quantificar as variações de *loudness* nas unidades absoluta LKFS e relativa LU.

O ano de 2010 foi marcado pela forte atuação da EBU nas discussões de *loudness*. No primeiro semestre, [Grimm, Everdingen e Schöpping \(2010\)](#) propuseram um valor alvo de *loudness* e apresentaram as motivações para a criação de um grupo de estudo dentro da organização à imprensa especializada. Surge então o primeiro trabalho de impacto na exploração do algoritmo padrão: [Grimm, Skovenborg e Spikofski \(2010\)](#) sugerem a aplicação de funções retangulares – ou funções portão – na saída do medidor com o objetivo de controlar melhor áudios com faixa larga de *loudness* e/ou com longos períodos de silêncio. As funções poderiam ser absolutas, relativas e recursivas. O projeto de experimento foi diferente dos demais utilizados até então (casamento de *loudness*, ITU-R BS.1116 e ITU-R BS.1534). Conteúdos longos de faixas de *loudness* largas e estreitas eram executados na sequência e a pergunta foi: “Qual destas sequências fornece a melhor experiência quanto a não ser preciso ajustar o volume?”. Testes subsequentes foram realizados com sequências normalizadas por diferentes funções portão. As funções com melhor desempenho foram as relativas com limiares -6 LU e -10 LU, com blocos de 400 ms. As conclusões deste estudo, aliadas aos resultados de [Skovenborg e Lund \(2009\)](#), formaram a base da Rec. EBU R.128 e do algoritmo ITU revisado tal como publicado em sua segunda versão.

Contudo, no que diz respeito ao LRA, os resultados de [Skovenborg e Lund \(2009\)](#) não foram apreciados na Rec. ITU-R BS.1770-2 em razão do estudo de [Boley, Danner e Lester \(2010\)](#) que testou a faixa de *loudness* da EBU contra outras estatísticas de faixa dinâmica e, apesar do parecer sobre 400 ms ser um bom tamanho de bloco, nenhuma das estatísticas apresentou diferença significativa em relação às demais métricas existentes na estimação da faixa dinâmica percebida. Numa publicação conjunta da ABC australiana e da Universidade de Sidney, [Dash, Bassett e Cabrera \(2010\)](#) discutiram a ancoragem dos níveis das demais faixas de um programa à faixa de voz. Os autores sugeriram

que o *loudness* médio deva ser observado apenas para controle de saltos em intervalos comerciais, e para programas longos, o controle deva se dar apenas na faixa de voz de modo a não comprometer a dinâmica da produção.

Foi 2011 o ano no qual se discutiram questões de ordem prática afeitas ao operador de áudio. A EBU apresentou sua Rec. R 128 à comunidade com seus novos descritores de “faixa de *loudness*”, voltado à produção de programas, “*loudness* momentâneo” (400 ms), voltado ao *display* de medidores, e “*loudness* de curta duração” (3 s), voltado à percepção da audiência (LUND, 2011). Norcross, Poulin e Lavoie (2011), no CRC, fizeram uma avaliação de balísticas para medidores e sugeriram a introdução de um filtro suavizador de ganho no caminho entre a saída do algoritmo ITU e o *display*, tal que as marcações no dispositivo fossem de ataque rápido e decaimento lento, como nos medidores analógicos tradicionais (VU e PPM). Já a *Free TV Australia* testou o algoritmo ITU com integrações assimétricas (diferentes tempos de ataque e decaimento) sem encontrar diferenças perceptíveis (ITU-R, 2011b).

Paralelamente às questões de balística, foram publicados outros estudos relevantes à canalização. Norcross e Lavoie (2011) observaram que no somatório dos canais durante o processo de *downmix* de 5.1 para estéreo podem ocorrer diferenças de percepção de intensidade e, no experimento, aproximadamente 10% das diferenças observadas foram maiores que 2 LU. Um artigo da Empresa Britânica de Radiodifusão (BBC) sobre o uso do canal LFE na radiodifusão propôs que o canal silencioso fosse a norma, devido a este ser descartado no *downmixing* de 5.1 para estéreo e a comprometer a qualidade do áudio se os níveis forem mal configurados na etapa de produção (MASON, 2011). Frente aos recentes resultados, o *Chairman* do grupo de trabalho ITU-R elencou, dentre as questões não resolvidas, avaliar o indicador “faixa de *loudness*”, descobrir como computar a pequena contribuição do LFE no cálculo do *loudness* e, à luz dos resultados de Nielsen e Lund (2003) e Lund (2006), definir sinais de testes para avaliar um medidor de pico verdadeiro (ITU-R, 2011a).

Em 2012, Travaglini, Alemanno e Uncini (2012), da Universidade de Roma, propuseram melhoramentos no método ITU incluindo uma função portão recursiva e filtros individualizados por canal cujas curvas são inversões diretas da curva de 65 *phon*, a menos de modificações causadas por espacialidade e

filtros passa-altas e passa-baixas projetados para representar os tempos de subida e descida da audição. Por consequência, o bloco do pré-filtro foi retirado e o ganho de calibração da soma de potências foi modificado. O desempenho foi avaliado como sendo melhor que o do algoritmo padrão em um teste subjetivo, contudo o teste foi somente de Pontuação Média Opinitiva (MOS), cabendo reprodução. Em outra publicação, [Travaglini, Alemanno e Lantini \(2012\)](#) também posicionaram a “Faixa de *Loudness*” em desfavor a outros descritores de faixa dinâmica. A TC Electronic saiu em defesa do descritor publicando as decisões de projeto que a levou a propô-lo à EBU na forma que foi publicado ([SKOVENBORG, 2012b](#)). [Norcross e Lavoie \(2012\)](#) deram sequência ao trabalho de [Dash, Bassett e Cabrera \(2010\)](#) e estudaram diferenças entre medidas de *loudness* na programação geral e medidas ancoradas na faixa de voz. No experimento, as diferenças no *loudness* medido aumentavam proporcionalmente ao aumento nas variações de faixa dinâmica. A interessante conclusão foi de que a abordagem de medidas ancoradas, apesar de ser mais apropriada considerando o balanceamento dos elementos chave da mixagem, só faz sentido na etapa de produção ou de pós-produção, e não na cadeia de distribuição da radiodifusão. Por fim, a terceira versão da Rec. ITU BS.1770 é publicada contemplando um medidor de pico verdadeiro desenvolvido pela ABC australiana, cujas especificações detalhadas só foram publicadas dois anos depois ([DASH, 2014](#)), no mesmo ano em que publiquei minha própria implementação do medidor ([PIRES, 2014](#)), cujo melhoramento foi utilizado em experimento de controle de *loudness*, relatado mais adiante na [seção 4.2](#).

4.1.2 Desenvolvimentos recentes

O período posterior à publicação da terceira versão de ([ITU-R, 2015b](#)) foi de absorção do padrão pelo mercado de radiodifusão. A comunidade estava ciente do problema e os radiodifusores compravam seus medidores *EBU mode* enquanto os marcos regulatórios regionais eram escritos. Nos EUA, o *Commercial Advertisement Loudness Mitigation Act* (CALM Act) ([USC, 2010](#)) foi implementado pelo regulador em 2012 com validade a partir de 2013 ([FCC, 2014](#)). O CRTC canadense publicou a política regulatória 2011-584, vigorando a partir de setembro de 2012 ([CRTC, 2011](#)). O parlamento francês aprovou o

Art. 177 da Lei nº 2010-788 sobre padronização de volume, disciplinada pelo regulador CSA em 2011 (CSA, 2011). No Brasil, a normatização técnica da Lei nº 12.810/2013 deu-se pelo Ministério das Comunicações em 2012 (MC, 2012). Seus desdobramentos resultaram, em 2013, na restrição do escopo de (PR, 2001) para somente radiodifusão digital e, em 2014, no procedimento de fiscalização de *loudness* pela Anatel (ANATEL, 2014).

Dada a situação estável na radiodifusão tradicional, o caminho lógico subsequente deu-se em direção às transmissões via internet (*streaming*) e aos aparelhos móveis. Thomas Lund (2013) conduziu um estudo de *loudness* para dispositivos móveis procurando estabelecer um denominador comum para dispositivos portáteis e propôs um *loudness* alvo de -16 LKFS, em sintonia com a necessidade de normalização de conteúdo especificamente veiculado nestes dispositivos (CAMERER *et al.*, 2012). A Convenção da AES de dezembro de 2014 hospedou um *workshop* no qual especialistas discutiram novos rumos para controle de *loudness* e de volume excessivo no cinema e em dispositivos de usuários (RUMSEY, 2015). Esta resultou, em 2015, no documento técnico AES TD1004.1.15-10 para *streaming* de áudio e reprodução de arquivos (AES, 2015), que recomenda um *loudness* alvo não superior a -16 LKFS e nem inferior a -20 LKFS. Especificamente para o denominado conteúdo de formato curto (anúncios, vinhetas e inserções rápidas), o *loudness* de curta duração máximo não poderia ser maior que 5 LU acima do *loudness* integrado. Um pico verdadeiro máximo de $-1,0$ dBTP foi recomendado para todos os casos.

Guardadas as devidas proporções, a preocupação com o conteúdo de formato curto foi herdada do problema original de comerciais de volume elevado na radiodifusão. A recomendação R 128 da EBU (2014) incorporou um suplemento específico sobre o assunto no final de 2014, objetivando tratá-lo com descritores de curta duração (EBU, 2016a). Nesta, o *loudness* alvo continuou sendo -23 LKFS e, além do limite relativo de 5 LU para o *loudness* de curta duração máximo, o *loudness* momentâneo máximo não poderia ser maior que 8 LU acima do *loudness* médio. Este suplemento foi a base do controlador de *loudness* proposto em 2016 por Pires, Vieira e Yehia (2016), discutido mais adiante na seção 4.2.

No que tange ao conteúdo cinematográfico, o descritor “Faixa de *Loudness*”

volta a ganhar força. Em 2014, a TC Electronic testou seis medidas objetivas de microdinâmica, que seriam medidas de faixa dinâmica em menor escala (SKOVENBORG, 2012b). Algumas das medidas testadas eram baseadas em nível de *loudness*, outras em relações pico/média, mas a medida de macrodinâmica consolidada foi a de faixa de *loudness*. Em 2015, Kean, Johnson e Sheffield (2015), nos laboratórios da Rádio Pública Nacional norte-americana (NPR), fizeram um amplo levantamento de faixas de *loudness* adequadas para consumidores em vários modos de escuta e em diferentes ambientes com o objetivo de recomendar níveis de compensação ao se projetar ganhos de dispositivos de consumo.

Outra direção para as pesquisas de *loudness* deu-se no sentido dos sistemas avançados de áudio, capitaneada principalmente pela Empresa de Radiodifusão do Japão (NHK) e pelo instituto de pesquisas alemão *Fraunhofer IIS*, este último impulsionado principalmente pelas especificações de *3D Audio* (3DA) do padrão MPEG-H (2015) (HERRE *et al.*, 2015), cujo decodificador conta com um módulo controlador de faixa dinâmica – e normalizador de *loudness* – a partir de metadados de entrada. Todos os descritores e padrões vigentes descritos nesse levantamento são aceitos pelo MPEG-H 3DA (KUECH *et al.*, 2015), mas ainda era preciso saber se o padrão ITU carecia de aprimoramentos para sistemas com mais de cinco canais. Nesse sentido, em 2014 a administração japonesa propôs flexibilizar o número de canais recalculando os ganhos direcionais para quaisquer posicionamentos de caixas acústicas em passos de 30 graus tanto em azimute quanto em elevação (ITU-R, 2014d). Os resultados são apresentados Tabela 4.1.

Tabela 4.1 – Pesos direcionais inicialmente propostos em (ITU-R, 2014d)

Direção do canal		Pesos G_i	
Azimute (graus)	Elevação (graus)		
$ \theta < 45$	$ \phi < 30$	1,00	(+0,0 dB)
$45 \leq \theta < 120$	$ \phi < 30$	1,41	(+1,5 dB)
$120 \leq \theta < 150$	$ \phi < 30$	1,00	(+0,0 dB)
$150 \leq \theta \leq 180$	$ \phi < 30$	0,71	(-1,5 dB)
$ \theta < 120$	$30 \leq \phi < 70$	1,00	(+0,0 dB)
$120 \leq \theta \leq 180$	$30 \leq \phi < 70$	0,71	(-1,5 dB)
$ \theta \leq 180$	$70 \leq \phi$	1,00	(+0,0 dB)
$ \theta \leq 45$	$\phi \leq -30$	1,00	(+0,0 dB)

Em 2015, testes subjetivos feitos por Komori *et al.* (2015) em 5.1, 7.1 e 22.1 apresentaram boa correlação com a medida objetiva e a recomendação ITU-R (2015b) foi republicada em sua quarta versão contemplando uma extensão para sistemas avançados, com uma distribuição de pesos tal como na Tabela 4.2. Em 2016, uma nova versão deste estudo foi feita em conjunto com a NHK, a Fraunhofer IIS e os laboratórios Dolby, corroborando os resultados anteriores (NORCROSS; NANDA; COHEN, 2016).

Tabela 4.2 – Pesos direcionais incluídos na quarta versão de (ITU-R, 2015b)

Elevação (ϕ)	Azimute (θ)		
	$ \theta < 60^\circ$	$60^\circ \leq \theta \leq 120^\circ$	$120^\circ < \theta \leq 180^\circ$
$ \phi < 30$	1,00(± 0 dB)	1,41(+1,5 dB)	1,00(± 0 dB)
demais casos	1,00(± 0 dB)		

Paralelamente, Francombe *et al.* (2015a), na Universidade de Surrey em conjunto com a emissora BBC, conduziram um experimento de casamento de loudness em vários *set-ups* espaciais: 5, 9, 22 canais e cuboide padrão Ambisonics (FRANK; ZOTTER; SONTACCHI, 2015), testando modelos de faixa única ($L_{eq}(Lin)$, $L_{eq}(A)$ e ITU-R BS.1770-3) e multifaixa (Moore/Glasberg e Zwicker/Fastl). O estudo concluiu que o método de Glasberg e Moore (2002) resultou no menor erro médio quadrático, contudo o tempo de execução ainda o pretere em relação ao método ITU (ITU-R, 2015b).

4.1.3 Linhas de investigação

À luz da fortuna crítica sobre o tema, entendo que melhorias no método de medida objetiva de loudness passam pela incorporação de elementos psicoacústicos ao modelo e, possivelmente, pela construção de curvas de ponderação usando modelos de regressão e/ou pela modificação da função portão de um limiar relativo para um limiar recursivo. Porém, não basta identificar os pontos de ajuste sem ter uma ideia clara do que se pretende perseguir com os ajustes do modelo.

Na reunião do grupo de trabalho do ITU-R em fevereiro de 2016, deliberou-se pela continuidade de um grupo relator discutindo o algoritmo de medida de

loudness para sistemas avançados de áudio (ITU-R, 2016b), cujos termos de referência seriam (grifos meus):

- Continuar o estudo das particularidades de um algoritmo de medida de *loudness* para **sistemas avançados de áudio**;
- Continuar conduzindo testes subjetivos para discutir o algoritmo de medida de *loudness* para sistemas avançados de **áudio baseado em objetos**;
- Estudar o efeito das características de frequência (canal LFE ou filtro passa baixas com frequência de corte de 18 kHz) na medida de *loudness*;
- Investigar o algoritmo para sistemas avançados de **áudio baseado em cenas**;
- Verificar os sinais de teste propostos, como também construir mais sinais de teste para incluir no Relatório ITU-R BS.2217 para verificar o algoritmo atualizado na última revisão da Recomendação ITU-R BS.1770.

Isto posto, aponto duas grandes linhas de investigação e suas arborescências:

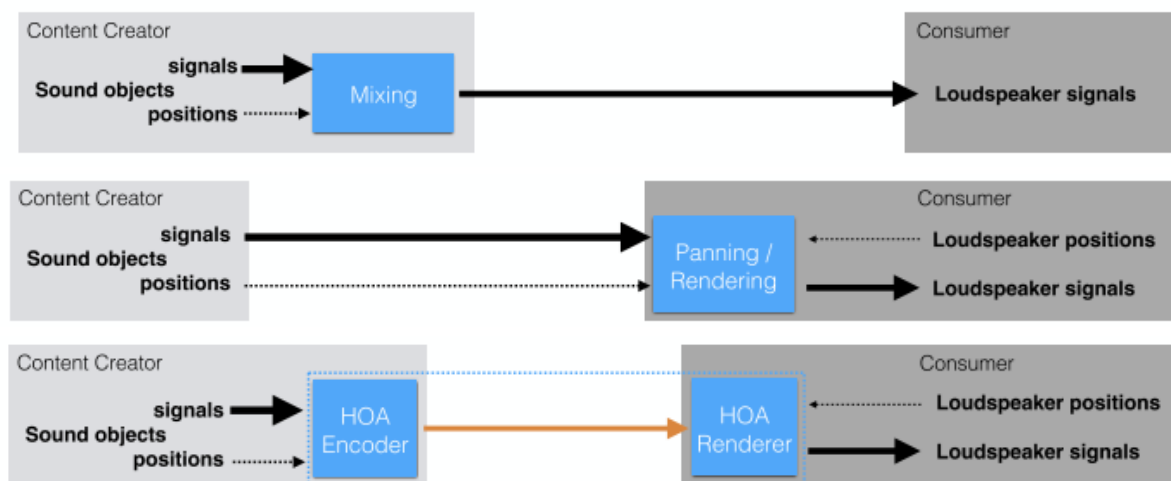
1. Características de frequência
 - a) Contribuição do canal LFE (*Low Frequency Effects*);
 - b) Filtro passa-baixas em 18 kHz.
2. Modelos de definição de áudio
 - a) Áudio baseado em canais;
 - b) Áudio baseado em objetos;
 - c) Áudio baseado em cenas.

À época de definição do algoritmo padrão, muito se debateu sobre a inclusão (ou não) do canal LFE no modelo multicanal (NORCROSS; LAVOIE, 2009; MASON, 2011; DASH, 2009), e o principal argumento pró-exclusão foi de que o canal não era aproveitado no *downmixing* de 5.1 para estéreo (NORCROSS; LAVOIE, 2011). Todavia, ao se considerar sistemas avançados de áudio superiores a cinco canais com mais de um canal LFE (FRANCOMBE *et al.*, 2015a; KOMORI *et al.*, 2015; SUGIMOTO; OODE; NAKAYAMA, 2015) e renderização de áudio 3D independente do posicionamento dos alto-falantes

([HERRE et al., 2015](#); [KUECH et al., 2015](#)), entendo como importante a retomada da discussão, e enxergo uma oportunidade de melhoramento do algoritmo ITU para acomodá-lo e comparar o *loudness* resultante com testes subjetivos.

Não me sinto competente para afirmar no que tange à produção de áudio para cinema, mas em produções para música e televisão é comum cortar toda a *mix* em 18 kHz, ao se sacrificar parte do áudio que a maioria dos consumidores não ouve, numa tentativa de se ganhar margem (*headroom*) de energia espectral seja para aumentar a excursão dinâmica do sinal ou para aumentar o *loudness*. Considerando a perspectiva de produção de áudio imersivo para radiodifusão, saber se a contribuição da faixa final do espectro audível é relevante para o *loudness* percebido é uma questão válida, de tal forma que este filtro possa ser introduzido na entrada do medidor e assim coibir abusos decorrentes desta prática¹. Sugere-se realização de testes subjetivos com conteúdo filtrado/não-filtrado e comparar as medidas objetivas para ambos os casos.

Figura 4.1 – Modelos de definição de áudio imersivo: a) Áudio baseado em canais, b) Áudio baseado em objetos e c) Áudio baseado em cenas



Fonte: [Peters et al. \(2015\)](#).

O primeiro dos três conceitos de áudio imersivo ilustrados pela [Figura 4.1](#) e definidos pela Recomendação BS.2266 do [ITU-R \(2014b\)](#), é o baseado em canais. É o formato *de facto* da indústria, empregado no desenvolvimento dos

¹ Em fevereiro de 2017, na lista de discussão do Grupo P-LOUD da EBU, foram reportadas práticas de produtoras embarcando um tom senoidal (inaudível) de 19 kHz em suas peças de áudio, com o objetivo de deixar a função portão do algoritmo ITU-R permanentemente aberta para que não haja silêncio a ser descartado e, conseqüentemente, alterar a leitura de medidores de *loudness* em alguns decibels para menos.

sistemas estéreo (2.0) e *surround* (5.1). Neste, os sinais enviados aos alto-falantes são mixados na etapa de produção para arranjos de caixas acústicas pré-definidos e são entregues numa ordem também pré-definida. A ilusão de posicionamento dos objetos sonoros é construída a partir das diferenças de níveis entre os canais (*panning*). O algoritmo BS.1770 foi desenvolvido especificamente para esta modalidade de áudio, limitada a cinco canais. Como dito num trabalho mais recente sobre o assunto, “*com os novos conteúdos imersivos e personalizados de áudio, a medida e o controle de loudness ainda está na sua infância*” (NORCROSS; NANDA; COHEN, 2016). Aprimorar o algoritmo passa inexoravelmente por abarcar estes novos conteúdos.

No conceito de áudio baseado em objetos, o áudio é entregue em componentes não mixados na forma de objetos, em conjunto com metadados que os definam. Com base nestes metadados, os sinais são adaptados para a canalização ou “renderizados” (*rendering*) e balanceados em diferentes níveis por canal no próprio equipamento do usuário, de acordo com o posicionamento dos alto-falantes. Por mais que o conceito seja promissor no sentido de o usuário poder controlar o *loudness* dos objetos individualmente (PAULUS, 2015), os custos de largura de banda e de processamento associados à transmissão e renderização de muitos objetos de áudio simultaneamente são desafiadores (PETERS *et al.*, 2015). Os estudos aqui desenvolvidos poderão comparar conteúdos idênticos codificados neste conceito e compará-los com sinais mixados para saber se há diferença perceptível no *loudness* após a renderização do conteúdo. Caso positivo, deve-se ajustar o modelo e compará-lo com outras medidas objetivas usando resultados de testes subjetivos como referência.

Já o áudio baseado em cenas é um híbrido dos dois conceitos anteriores. Tal como no modelo de canais, o conteúdo é preparado na etapa de produção num número fixo de sinais de áudio. Por outro lado, tal como no modelo de objetos, o número e o posicionamento dos alto-falantes é transparente para a transmissão. Os sinais transmitidos são coeficientes de harmônicos esféricos do *Ambisonics* de Alta Ordem (*Higher Order Ambisonics* – HOA) que descrevem os sons e suas propriedades espaciais variantes no tempo dentro de uma determinada cena (FRANK; ZOTTER; SONTACCHI, 2015). Estes sinais são então renderizados no receptor de acordo com o posicionamento dos alto-falantes. A renderização

do HOA passa pelo uso de HRTFs para a construção da espacialidade a partir dos coeficientes de harmônicos esféricos. A possibilidade a ser perseguida aqui seria de usá-las enquanto substituições do pré-filtro da curva K (função de sombreamento de cabeça) para incorporá-las ao método de medida. Posteriormente, compará-las com métodos tradicionais usando resultados de testes subjetivos como referência.

4.1.4 *Cursos de ação*

Posto o histórico recente e apresentadas as perspectivas de trabalho, foram conduzidos dois experimentos, cada qual na tentativa de se responder uma das perguntas abaixo:

1. Considerando as regulamentações regionais publicadas e as atualizações recentes das Recomendações ITU-R BS.1770 e EBU R. 128, sob que aspectos a norma brasileira de *loudness* para a radiodifusão pode ser revisada?
2. Considerando as perspectivas de desenvolvimento do tema no ITU-R, o advento dos sistemas avançados de reprodução e a finalização do padrão MPEG-H 3D *Audio*, como o modelo de *loudness* da ITU pode ser aprimorado nesse contexto?

As propostas e os experimentos que as acompanham serão descritos na [seção 4.2](#) e na [seção 4.3](#) a seguir.

4.2 Controle Automático de *Loudness* em Conteúdo de Formato Curto

A Lei nº 10.222, de 9 de maio de 2001, que “padroniza o volume de áudio das transmissões de rádio e televisão nos espaços dedicados à propaganda e dá outras providências”, foi originalmente promulgada com a seguinte redação

(...)

Art. 1º Os serviços de **radiodifusão sonora e de sons e imagens** padronizarão seus sinais de áudio, de modo a que não haja, no momento

da recepção, **elevação injustificável de volume nos intervalos comerciais**.

Art. 2º O Poder Executivo criará, no período de **cento e vinte dias**, a contar da publicação desta Lei, os mecanismos necessários à normalização técnica da matéria, bem como à fiscalização de seu cumprimento.

Art. 3º O descumprimento do disposto nesta Lei sujeitará o infrator à pena de **suspensão da atividade** pelo prazo de trinta dias, triplicada em caso de reincidência.

(...)(PR, 2001, grifos meus)

Em 2001, as obrigações desta lei contemplavam todos os serviços de entrega de conteúdo audiovisual disponíveis no país – rádios e TVs analógicas – no sentido de coibir apenas os saltos de intensidade entre programas e comerciais, não exigindo uma normalização do áudio em toda a transmissão. Não foram previstos casos particulares como os de programas ao vivo e de inserções de chamada, que eram comumente transmitidos com níveis de *loudness* distintos dos da programação de estúdio.

À época não havia como se medir objetivamente a diferença perceptiva de intensidade entre a programação e os intervalos comerciais, e a normalização técnica da matéria, de criação prevista para cento e vinte dias após a publicação da Lei, deu-se onze anos depois. Consequentemente, nenhuma emissora de radiodifusão foi punida com suspensão das atividades no período de vigência da redação original.

Em razão dos trabalhos em curso durante os anos de 2012 e 2013 pela normatização do tema e pela fiscalização das obrigações previstas, a Lei nº 12.810, de 15 de maio de 2013, alterou a redação da Lei nº 10.222/2001 da seguinte forma:

(...)

Art. 18. Os arts. 1º e 3º da Lei nº 10.222, de 9 de maio de 2001, passam a vigorar com a seguinte redação:

“Art. 1º Os serviços de radiodifusão sonora e de som e imagens **transmitidos com tecnologia digital** controlarão seus sinais de áudio de modo que não haja elevação injustificável de volume nos intervalos comerciais.” (NR)

“Art. 3º O descumprimento do disposto nesta Lei sujeitará o infrator às **penalidades prescritas no Código Brasileiro de Comunicações**.” (NR)

(...)(PR, 2013, grifos meus)

A Comissão Mista encarregada de examinar a Medida Provisória convertida na Lei nº 12.810/2013 justificou as alterações na Lei nº 10.222/2001 em relatório.

Propomos modificação da Lei nº 10.222, de 9 de maio de 2001, que dispõe sobre o volume de áudio das transmissões de rádio e televisão nos espaços dedicados à propaganda, chamado de aumento injustificado do volume de áudio nos intervalos comerciais, para que tal exigência se aplique **somente à transmissão digital**. Nesses doze anos de vigência da Lei, não foi possível implementar tal dispositivo por **razões tecnológicas**. Atualmente, considerando o alto investimento das empresas difusoras para implantar o sistema digital em todo o País, criar sistemas paralelos para gerenciamento do volume de áudio das transmissões analógicas encontra dificuldades de custos e tecnologia. Sendo assim, como os sistemas digitais encontram-se em **fase avançada de implantação**, é preciso dispensar as emissoras analógicas dessa obrigação, **impossível** de ser cumprida por falta de soluções técnicas viáveis. (CN, 2013, grifos meus)

Embora a exposição de motivos não apresente justificativa quanto à conversão das apenações de “suspensão da atividade” para “penalidades prescritas no Código Brasileiro de Comunicações”, há fundamentos para desobrigar as emissoras analógicas do controle de *loudness* nas suas programações. Em primeiro lugar porque o problema da percepção de intensidade sonora na radiodifusão analógica não é resolvido inteiramente pelo controle de níveis do áudio, mas também pela limitação do desvio de modulação, de ± 75 kHz na radiodifusão em Frequência Modulada (FM) (ANATEL, 2010, Item 7.2.2 b, p. 31) e de ± 50 kHz na radiodifusão de sons e imagens (ANATEL, 2001, Item 3.2.3.1.3, p. 19). A ausência ou o mau funcionamento de um limitador de modulação pode fazer com que os desvios sejam superiores a esses limites, caracterizando uma situação de sobremodulação na qual o áudio da programação é percebido de modo mais intenso como um todo, se comparado às emissoras concorrentes na mesma área geográfica. E em segundo lugar porque distorções na etapa de demodulação são mais ou menos prejudiciais dependendo do sistema receptor, ou seja, as condições de intensidade de sinal e de recepção modificam a razão sinal/ruído e a percepção de intensidade.

Como há mais fatores de impacto no âmbito da engenharia de Radiofrequência (RF) do que no escopo da engenharia de áudio, o controle de *loudness* na radiodifusão analógica é de fato mais custoso. Por outro lado, embora o crono-

grama de desligamento do sinal analógico de televisão já esteja em curso, ainda não há uma definição para o padrão de rádio digital no país e, por consequência, o sinal analógico na radiodifusão sonora permanecerá no ar ainda por muitos anos, assim como os problemas de *loudness* inerentes à modalidade de transmissão.

A normatização veio com a publicação da Portaria nº 354, de 11 de julho de 2012 do antigo Ministério das Comunicações, com seus principais trechos destacados abaixo (MC, 2012).

Art. 2º Para efeitos desta Portaria, aplicam-se as definições a seguir:

(...)

II - Faixa de *Loudness* - **faixa na qual varia a intensidade subjetiva de áudio** ao longo de um período de medição;

III - Intensidade subjetiva de áudio (*Loudness*) - percepção da intensidade do som ou dos sinais de áudio quando estes são reproduzidos acusticamente, tratando-se de uma **função complexa**, que pode ser medida objetivamente por meio de algoritmos definidos na Recomendação ITU-R BS.1770-2 e na Recomendação EBU R-128- 2011;

IV - Intensidade média subjetiva de áudio (*Loudness* médio) - **média da intensidade subjetiva de áudio** medida em um intervalo de tempo;

Art. 3º Para efeito do controle dos sinais de áudio de que trata esta Portaria, de modo que não haja elevação injustificável de volume entre um bloco de programa e o intervalo comercial imediatamente posterior, serão considerados:

I - os **limites de modulação** e os critérios de fiscalização constantes nos regulamentos específicos de cada serviço; e

II - o **padrão internacional** e os algoritmos recomendados pela União Internacional de Telecomunicações.

§ 1º Na programação transmitida, serão observados os seguintes parâmetros:

I - a intensidade subjetiva de áudio (*Loudness*) dos blocos de programas deverá ser centrada em -23 LKFS, com tolerância, para mais ou para menos, de **2 LKFS**;

II - a intensidade subjetiva de áudio (*Loudness*) dos intervalos comerciais deverá ser centrada em -23 LKFS, com tolerância, para mais ou para menos, de **2 LKFS**; e

III - a Faixa de *Loudness* do canal de áudio principal dos programas e dos intervalos comerciais não deve ultrapassar o valor de 15 LU.

Art. 1º (...)

§5º Quando (...) a intensidade média subjetiva do áudio do intervalo comercial for **superior** à do bloco de programa a ele anterior **em mais de 2 LKFS**, será caracterizada infração ao disposto na Lei nº 10.222, de 9 de maio de 2001, e nesta Portaria. (MC, 2012, grifos meus)

Com a massa de conhecimento adquirido sobre o tema nos capítulos anteriores, é possível fazer uma leitura crítica da norma de *loudness* para a radiodifusão brasileira:

- Na [subseção 3.3.4](#), foi visto que a faixa de *loudness* não é exatamente a faixa na qual a intensidade percebida varia no tempo, e sim a diferença entre o 10^o e o 95^o percentis da distribuição dos níveis de *loudness* fechado medidos a intervalos de 3 segundos com sobreposição de 66%, distintos dos intervalos de entrecorte do *loudness* médio ao qual a norma se refere.
- *Intensidade subjetiva de áudio* é uma tradução boa para o *loudness*. Contudo, a tradição bibliográfica trata como subjetivo somente o teste auditivo, sendo o próprio *loudness* descrito como uma *sensação* a se perceber (ver [Capítulo 2](#)). Nesse sentido, optou-se por traduzir *loudness* como intensidade *percebida* de áudio ao longo deste texto.
- Com base na [subseção 3.3.3](#), sabe-se que o *loudness* médio não é a média da intensidade subjetiva de áudio, e sim a medida de *loudness* fechado em intervalos de 400 ms com 75% de sobreposição a um limiar absoluto de -70 LKFS e a um limiar relativo de -10 LU.
- O Art. 3^o foi correto ao considerar os limites de modulação como um elemento de controle de *loudness* na radiodifusão, porém esta redação é anterior à Lei n^o 12.810/2013 e portanto, sua vigência enseja retificação do inciso I.
- O Art. 3^o também foi correto ao considerar o padrão internacional para além da União Internacional de Telecomunicações de modo a assegurar a introdução de recomendações relevantes, como a EBU R 128 e seus cadernos suplementares.
- Onde se lê “2 LKFS” dever-se-ia ler “2 LU”, dado que é um desvio e não um valor absoluto. Ainda assim, considerando uma normalização recomendada em -23 LKFS $\pm 0,5$ LU para programas de estúdio e em -23 LKFS ± 1 LU para programas ao vivo (ver [subseção 3.3.4](#)), uma tolerância de ± 2 LU para todos os casos é considerada permissiva se comparada à experiência internacional.

- Note que a infração é caracterizada quando o nível de *loudness* do bloco de propaganda for superior ao nível de *loudness* do bloco de programação em 2 LU, dando margem a situações de irregularidade enquadrada no Art. 4^o, §5^o, com regularidade enquadrada no Art. 4^o, §1^o, incisos I e II. Exemplo: *loudness* médio de um bloco de programa medido a $-24,5$ LKFS e de um bloco de propaganda subsequente medido a -22 LKFS.

Considerando o objetivo da norma de coibir a “elevação injustificada de volume” nos intervalos comerciais, duas coisas chamam a atenção: a ausência de descritores curtos de *loudness* e o uso do descritor “faixa de *loudness*” na caracterização de uma peça áudio como sendo regular ou ofensora. No que tange ao conteúdo de formato curto (comerciais, vinhetas e inserções), o Suplemento n^o 1 da Recomendação R 128 da EBU (2016a) dispõe que:

- o parâmetro “Faixa de *Loudness*” não é aplicável por ser baseado numa análise estatística de valores de “*Loudness* de Curta Duração” (3 segundos) já que, para comerciais e vinhetas, resulta num conjunto de valores pequeno demais para se obter um resultado significativo;
- os descritores “*Loudness* Médio” e “Nível Máximo de Pico Verdadeiro” por si só são insuficientes para caracterização de comerciais, vinhetas e inserções. E por isso é recomendado o uso dos parâmetros “*Loudness* Momentâneo Máximo” (400 milissegundos) e “*Loudness* de Curta Duração” (SKOVENBORG; NIELSEN, 2007) para controle de peças altamente comprimidas.

Um bloco de 30 segundos de propaganda resultaria num vetor de *loudness* de curta duração não fechado com trinta elementos no melhor caso. Considerando vetores de programas típicos entre uma e duas horas de duração contendo de 3600 a 7200 medidas, e supondo uma verificação de regularidade como um teste estatístico, é razoável presumir que um vetor de 30 amostras resultaria num teste estatístico de baixa potência, com alta sensibilidade a erros do tipo II, ou seja, de não se rejeitar a hipótese nula ($LRA \leq 15$ LU) quando esta for falsa (MONTGOMERY, 2003).

Poucos segundos muito intensos numa peça de propaganda comercial são suficientes para incomodar a audiência, mesmo que o restante do tempo esteja com intensidade inferior ao nível de *loudness* médio de referência e que a peça seja caracterizada como regular pelos critérios da Portaria nº 354/2012. Daí a necessidade de descritores curtos de *loudness* capazes de avaliar a regularidade de blocos de propaganda pelos seus vetores de *loudness* momentâneo ou de curta duração.

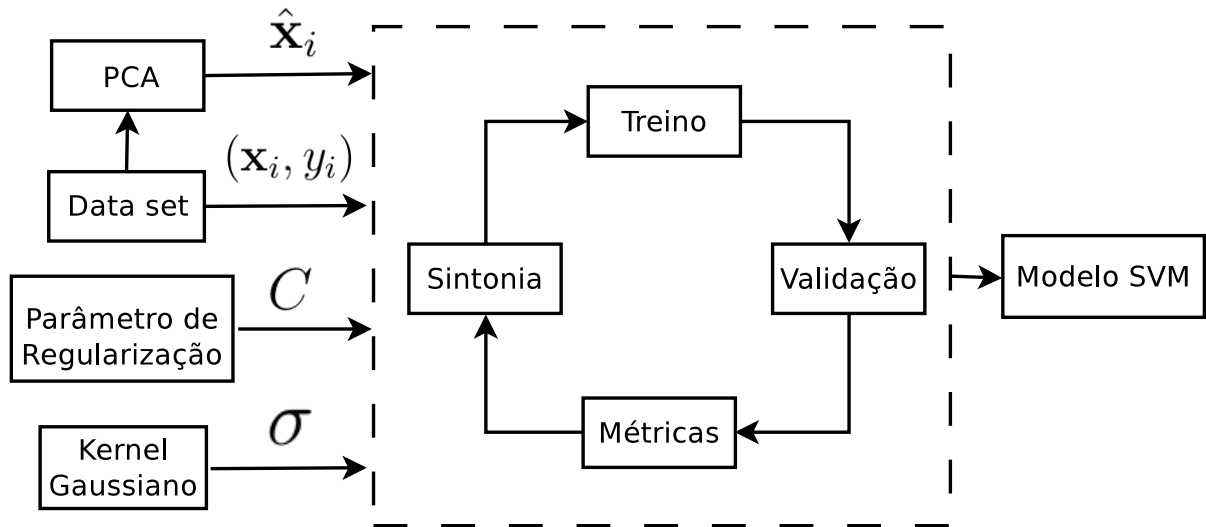
Isto posto, o objetivo deste experimento foi propor um esquema de controle de conteúdo de formato curto diretamente na cadeia de distribuição do sinal de áudio na radiodifusão com base nestas novas recomendações que, embora ainda não abarcadas pela regulamentação brasileira, têm impacto na Qualidade da Experiência (QoE) do usuário. Quando da sua concepção, o trabalho vislumbrou potencial implementação num Processador Digital de Sinais (DSP) de baixo custo, destinado a emissoras de menor porte que por ventura não disponham de *hardware* dedicado de processamento de áudio, ou que tipicamente executem peças publicitárias no estado durante suas programações.

A proposta é composta de um detetor de conteúdo de formato curto seguida de um controlador de *loudness*. Seus funcionamentos serão descritos nas subseções a seguir.

4.2.1 Detetor de conteúdo de formato curto

Paralelamente à regressão logística, aos sistemas de inferência nebulosa e às redes neuronais artificiais, a máquina de aprendizado denominada Máquina de Vetor de Suporte (SVM) é também um algoritmo poderoso e cada vez mais popular tanto na indústria quanto no meio acadêmico (DUDA; HART; STORK, 2012; ABE, 2006). Por aprendizado supervisionado, o detetor baseado em SVM aprende funções de decisão que classificam dados de entrada em uma das duas classes: *conteúdo de formato curto* ou *programa*. A Figura 4.2 ilustra os processos de treinamento e sintonia propostos para o detetor.

Figura 4.2 – Diagrama em blocos do detetor de conteúdo de formato curto



Fonte: Elaborada pelo autor.

Conjunto de dados e extração de características

Para os experimentos usou-se a base de dados “*TV News Channel Commercial Detection Dataset*” construída por [Vyas et al. \(2014\)](#). O conjunto de dados contém cinco características de áudio com duas dimensões cada (média e variância), 5 de vídeo e 129685 exemplos correspondentes a 150 horas de gravação de cinco canais de notícias, três locais e dois internacionais (CNN e BBC), dos quais $\frac{2}{3}$ compuseram o conjunto de treinamento e $\frac{1}{3}$ o de validação. Estão entre as características de áudio, médias e variâncias das seguintes medidas: Energia de Curta Duração (STE), Taxa de Cruzamento de Zeros (ZCR), centroides espectrais, decaimento espectral e fluxo espectral ([RABINER; SCHAFER, 2007](#)).

Para um sinal $x[m]$, sua energia de curta duração (STE) , ou num único quadro, foi calculada como sendo o quadrado das amplitudes das amostras contidas na janela limitante do quadro centrado no instante de análise \hat{n} , $w[\hat{n} - m]$:

$$E_{\hat{n}} = \sum_{m=0}^{N-1} (x[m]w[\hat{n} - m])^2 = \sum_{m=0}^{N-1} x^2[m]w^2[\hat{n} - m]. \quad (4.1)$$

Ao longo de quadros sucessivos, a envoltória STE do sinal alterna rapidamente entre estados de alta e baixa energia em função da atividade sonora.

A taxa de cruzamento de zeros (ZCR) num quadro de áudio corresponde

à taxa de mudanças entre valores positivos e negativos que o sinal pode assumir nos limites do quadro:

$$Z_{\hat{n}} = \sum_{m=0}^{N-1} \frac{1}{2} |\operatorname{sgn}\{x[m]\} - \operatorname{sgn}\{x[m-1]\}| w[\hat{n} - m]. \quad (4.2)$$

A ZCR pode ser interpretada como uma medida de quão ruidoso é um sinal, alternando rapidamente entre estados de alta e baixa taxa em função das pausas na atividade sonora. A ZCR é usada em conjunto com a STE em aplicações de detecção de atividade de voz ou identificação de eventos sonoros.

O ponto de decaimento (*rolloff*) espectral é definido como sendo a frequência abaixo da qual está concentrado um percentil P da distribuição de magnitude do sinal. Portanto, se o m -ésimo coeficiente da DFT corresponde ao decaimento espectral do i -ésimo quadro:

$$\sum_{k=0}^{m-1} X_i[k] \approx P \sum_{k=0}^{N-1} X_i[k], \quad (4.3)$$

onde os valores de k representam as raias espectrais do sinal. A medida é usada para distinguir fontes com maior/menor concentração de magnitudes nas altas frequências. No experimento, foi adotado um valor de $P = 80\%$.

O centroide espectral é calculado como sendo a média das frequências presentes no sinal, ponderada pelas suas magnitudes:

$$\text{Centroide} = \frac{\sum_{k=0}^{N-1} k |X[k]|}{\sum_{k=0}^{N-1} |X[k]|}. \quad (4.4)$$

Esta medida indica onde se encontra o “centro de massa” do espectro. Perceptivamente, está relacionado com a percepção de “brilho” do som.

O fluxo espectral é calculado como o quadrado da diferença entre as magnitudes normalizadas dos espectros de dois blocos consecutivos:

$$Fl_{(i,i-1)} = \sum_{k=0}^{N-1} \left(\frac{|X_i[k]|}{\sum_{l=0}^{N-1} |X_i[l]|} - \frac{|X_{i-1}[k]|}{\sum_{l=0}^{N-1} |X_{i-1}[l]|} \right)^2. \quad (4.5)$$

A fala apresenta fluxos maiores do que a música em razão de um maior número de transitórios e alterações entre fala vozeada e surda, por exemplo (GIANNAKOPOULOS; PIKRAKIS, 2014, p. 85). A medida também é usada em detecção de timbre e identificação de *onsets*.

Redução de dimensionalidade

A Análise de Componentes Principais, ou simplesmente PCA (*Principal Component Analysis*), é uma técnica que consiste na redução do número de variáveis aleatórias do problema, com o intuito de diminuir sua complexidade, eliminar redundâncias e facilitar a visualização dos dados, quando possível. Quando a dimensionalidade dos dados é alta, a PCA pode ser utilizada para projetá-los num espaço de dimensionalidade inferior retendo o máximo de informação possível (GUYON *et al.*, 2008). As coordenadas dos pontos projetados podem ser reaproveitadas como novas características. Quando não nos damos ao luxo de ter dados graficamente representáveis, PCA se torna uma ferramenta interessante de análise. Expressando o conjunto de características de áudio como uma variável aleatória multidimensional \mathbf{X} , de média $\mathbf{m}_x = E[\mathbf{X}]$, sua matriz de covariância é dada por:

$$\Sigma_x = E \left[(\mathbf{X} - \mathbf{m}_x)^T (\mathbf{X} - \mathbf{m}_x) \right]. \quad (4.6)$$

Σ_x é uma matriz positiva definida que pode ser decomposta via Decomposição em Valores Singulares (SVD) ou “decomposição em autovalores”:

$$\Sigma_x = \mathbf{U} \mathbf{S} \mathbf{U}^T, \quad (4.7)$$

onde \mathbf{S} é uma matriz quadrada diagonal cujos elementos diferentes de zero são os autovalores de Σ_x em ordem decrescente, e \mathbf{U} é uma matriz unitária, ou de rotação, cujas colunas são os autovetores normalizados de Σ_x :

$$\mathbf{U} \mathbf{U}^T = \mathbf{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}. \quad (4.8)$$

Agora, se definirmos:

$$\mathbf{Y} = \mathbf{U}^T (\mathbf{X} - \mathbf{m}_x), \quad (4.9)$$

pela Equação 4.6, teremos:

$$\Sigma_y = E \left[\mathbf{U}^T (\mathbf{X} - \mathbf{m}_x) (\mathbf{X} - \mathbf{m}_x)^T \mathbf{U} \right] \quad (4.10)$$

$$= \mathbf{U}^T E \left[(\mathbf{X} - \mathbf{m}_x) (\mathbf{X} - \mathbf{m}_x)^T \right] \mathbf{U} \quad (4.11)$$

$$= \mathbf{U}^T \Sigma_x \mathbf{U} \quad (4.12)$$

$$= \mathbf{U}^T \mathbf{U} \mathbf{S} \mathbf{U}^T \mathbf{U} \quad (4.13)$$

$$= \mathbf{I} \mathbf{S} \mathbf{I} \quad (4.14)$$

$$= \mathbf{S}. \quad (4.15)$$

A covariância de \mathbf{Y} é uma matriz diagonal, ou seja, as componentes de \mathbf{Y} são descorrelacionadas. Logo:

$$\mathbf{X} = \mathbf{U} \mathbf{Y} + \mathbf{m}_x. \quad (4.16)$$

Assim sendo, a Equação 4.9 e a Equação 4.16 definem a representação de \mathbf{X} em componentes principais e sua inversa (DUDA; HART; STORK, 2012).

O emprego de PCA neste trabalho objetivou reduzir o espaço de m características para k componentes principais e aumentar a velocidade do detetor. A matriz de covariância do conjunto de dados das características de áudio é decomposta em valores singulares para se determinar o máximo de dimensões redutíveis do problema. Daí, escolhe-se o menor número de dimensões tal que o erro quadrático de projeção dividido pela variância total seja inferior a 1%, ou seja, tal que o número de componentes principais retenha 99% da variância dos dados, testado pela soma cumulativa dos autovalores de Σ_x na matriz diagonal \mathbf{S} :

$$\frac{\sum_{i=1}^k S_{ii}}{\sum_{i=1}^m S_{ii}} \geq 0,99. \quad (4.17)$$

Parâmetros sintonizáveis

Para o aprendizado de complexas regiões de decisão, o SVM precisa fazer um mapeamento não-linear do espaço original de características antes da minimização da função objetivo. O algoritmo faz esse mapeamento por meio funções denominadas núcleos (*kernels*) que medem a similaridade entre um par de treinamento (\mathbf{x}_j, y_j) e um marcador \mathbf{z}_i posicionado no espaço de características.

Dados m pares e m marcadores, cada par (\mathbf{x}_i, y_i) será então descrito por um novo vetor de características \mathbf{f} com elementos $\mathbf{f}_i = H(\mathbf{x}_i, \mathbf{z}_i)$, onde $H(\cdot, \cdot)$ é a função *kernel* e cada amostra do conjunto de treinamento \mathbf{x}_i é descrita por um novo vetor composto de graus de similaridade entre \mathbf{x}_i e cada marcador \mathbf{z}_i .

O mapeamento do conjunto de características por núcleos consiste na execução das etapas abaixo relacionadas (ABE, 2006):

- Dados $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_m, y_m)$, escolha $\mathbf{z}_1 = \mathbf{x}_1, \mathbf{z}_2 = \mathbf{x}_2, \dots, \mathbf{z}_m = \mathbf{x}_m$
- Para cada exemplo \mathbf{x}_i :

$$\mathbf{f} = \begin{bmatrix} f_1 = \text{similaridade}(\mathbf{x}_i, \mathbf{z}_1) \\ f_2 = \text{similaridade}(\mathbf{x}_i, \mathbf{z}_2) \\ \vdots \\ f_m = \text{similaridade}(\mathbf{x}_i, \mathbf{z}_m) \end{bmatrix} \quad (4.18)$$

no qual a função de similaridade pode ser, uma função de base radial da distância euclidiana entre a amostra \mathbf{x}_i e o marcador correspondente. Neste experimento, foi utilizado um núcleo gaussiano escrito da forma:

$$f_j = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{z}_i\|^2}{2\sigma^2}\right) \quad (4.19)$$

Neste trabalho, além do desvio padrão do núcleo gaussiano σ , há outro parâmetro de ajuste identificável: o termo de regularização C da máquina de aprendizado, como será detalhado a seguir.

Modelo SVM

Considere o conjunto de dados com m amostras de treinamento, cada amostra com n características e k componentes principais. Sejam m entradas de treinamento com k dimensões, $\hat{\mathbf{x}}_i (i = 1, \dots, m)$, pertencentes à classe *conteúdo de formato curto*, rotulada por $y_i = 1$, ou à classe *programa*, rotulada por $y_i = 0$. Adicionalmente, sejam \mathbf{f}_i as novas l entradas de treinamento obtidas após o mapeamento por *kernels* de $\hat{\mathbf{x}}_i$, e \mathbf{w} o vetor de coeficientes do hiperplano de separação. A função objetivo usada para treinamento do SVM como classificador

binário é dada por (ABE, 2006):

$$\min_{\mathbf{w}} \sum_{i=1}^m \left[y_i \text{custo}_1(\mathbf{w}^T \mathbf{f}_i) + (1 - y_i) \text{custo}_0(\mathbf{w}^T \mathbf{f}_i) \right] + \frac{1}{2C} \sum_{j=1}^l \mathbf{w}_j^2, \quad (4.20)$$

onde, para cada i -ésimo exemplo do conjunto de treinamento, $\mathbf{w}^T \mathbf{f}_i$ é uma função discriminante linear no espaço de características l -dimensional e as funções $\text{custo}_1(\cdot)$ e $\text{custo}_0(\cdot)$ são, respectivamente, as funções de custo SVM para as classes $y_i = 1$ e $y_i = 0$, e são aproximações lineares por partes das funções de custo log-sigmoide do algoritmo de regressão logística.

A expressão fora do somatório dos exemplos é o chamado *termo de regularização* (BRAGA, 2007), técnica anti-sobreajuste (*overfitting*) que consiste em penalizar valores elevados dos parâmetros do modelo, tal que ao término da minimização da função objetivo tem-se um vetor de parâmetros da função discriminante linear \mathbf{w} de magnitude reduzida que, no espaço original de características de um problema não linearmente separável, traduz-se numa região de separação menos sinuosa e, por consequência, menos aderente ao conjunto de dados de treinamento e com maior potencial de generalização para novos dados. C é o parâmetro de ajuste da regularização.

Como parâmetro de sintonia do modelo SVM, se C é muito grande há pouca regularização da função objetivo e propensão ao sobreajuste do modelo aos dados de treinamento. Já se for demasiadamente pequeno, há muita regularização, pouca aderência aos dados de treinamento e, por consequência, uma margem mais larga de classificação. O mesmo efeito pode ser observado para desvios padrão do *kernel* gaussiano estreitos e largos por estarem associados a variações bruscas e suaves das características \mathbf{f} , respectivamente.

Com os custos na Equação 4.20 sendo zero, ficamos com o problema reduzido de otimização sujeito a restrições:

$$\min_{\mathbf{w}} C \cdot 0 + \frac{1}{2} \sum_{j=1}^l \mathbf{w}_j^2 \text{ sujeito a:} \quad (4.21)$$

$$\mathbf{w}^T \mathbf{f}_i \geq 1 \text{ se } y_i = 1 \text{ e} \quad (4.22)$$

$$\mathbf{w}^T \mathbf{f}_i \leq -1 \text{ se } y_i = 0 \quad (4.23)$$

Note que o produto interno $\mathbf{w}^T \mathbf{f}_i$ pode ser escrito como a magnitude da

projeção de \mathbf{f}_i que multiplica a norma do vetor \mathbf{w} . Assim sendo, a [Equação 4.21](#) pode ser reescrita da forma:

$$\min_{\mathbf{w}} C \cdot 0 + \frac{1}{2} \sum_{j=1}^n \mathbf{w}_j^2 \text{ sujeito a :} \quad (4.24)$$

$$p_i \cdot \|\mathbf{w}\| \geq 1 \text{ se } y^{(i)} = 1 \text{ e} \quad (4.25)$$

$$p_i \cdot \|\mathbf{w}\| \leq -1 \text{ se } y^{(i)} = 0 \quad (4.26)$$

Dado que a função objetivo reduzida consiste na minimização da norma do vetor \mathbf{w} , sujeita a restrições do produto $p_i \cdot \|\mathbf{w}\|$, o SVM treinado terá suas projeções p_i , ou margens, como sendo as maiores possíveis.

Fluxo de treinamento

O detetor de conteúdo foi treinado com diferentes pares de C e σ . Para cada ajuste, o classificador recém treinado foi testado no conjunto de validação, e o modelo SVM selecionado é o de melhor pontuação num dado conjunto de métricas baseadas no número de Verdadeiros Positivos (VP), Verdadeiros Negativos (VN), Falsos Positivos (FP) e Falsos Negativos (FN). A primeira delas é o percentual de acertos, ou acurácia, definido da forma:

$$Acc(\%) = 100 \cdot \left(\frac{VP + VN}{VP + FP + VN + FN} \right). \quad (4.27)$$

A acurácia é a razão das instâncias corretamente classificadas para todas as instâncias. Já as métricas de sensibilidade e especificidade são, respectivamente, as medidas estatísticas das instâncias corretamente classificadas positiva e negativamente:

$$\text{sensibilidade} = \frac{VP}{VP + FN}, \text{ e} \quad (4.28)$$

$$\text{especificidade} = \frac{VN}{VN + FP}. \quad (4.29)$$

A última métrica utilizada foi o Coeficiente de Correlação de Matthews (MCC), uma medida de qualidade de classificação binária estável mesmo quando as densidades das classes são consideravelmente diferentes ([ALPAYDIN, 2014](#)). O MCC é um coeficiente de correlação entre as classificações binárias rotuladas

e as preditas pelo modelo SVM, e assume valores entre -1 e $+1$. Sua formulação é dada abaixo:

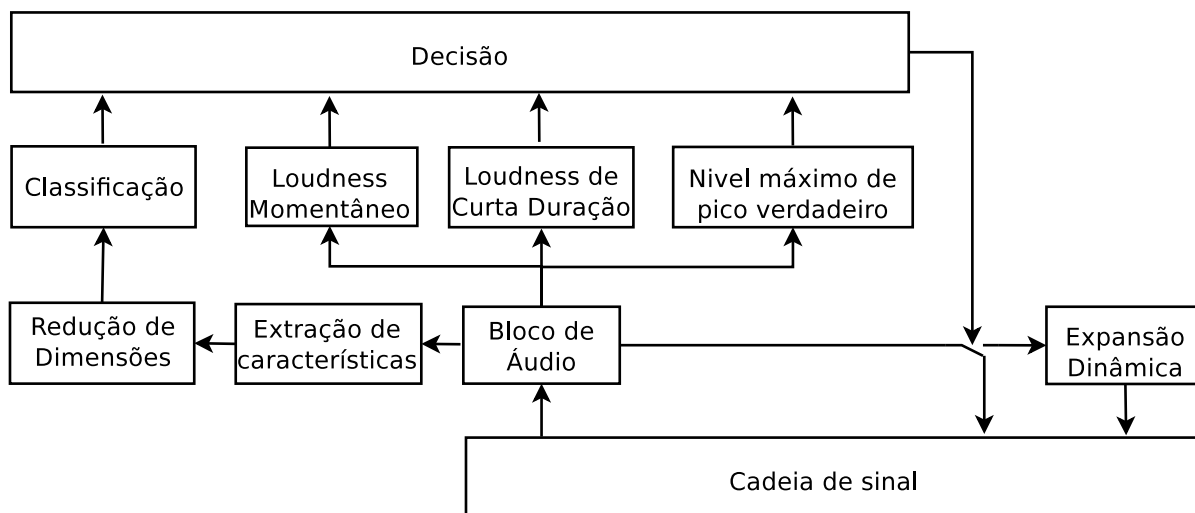
$$MCC = \frac{VP \cdot VN - FP \cdot FN}{\sqrt{(VP + FP)(VP + FN)(VN + FP)(VN + FN)}}. \quad (4.30)$$

O coeficiente assume o valor $+1$ quando o classificador faz predições perfeitas, -1 quando as predições e os rótulos de treinamento são diferentes por completo, e zero quando a classificação não é melhor que uma predição aleatória.

4.2.2 Controlador de loudness

Os passos de processamento de sinais de áudio propostos para o controlador de *loudness* são ilustrados na [Figura 4.3](#).

Figura 4.3 – Diagrama em blocos do controlador de *loudness*



Fonte: Elaborada pelo autor.

Os segmentos de áudio classificados como *conteúdo de formato curto* são processados em dinâmica caso violem algum dos critérios de níveis máximos recomendados pelo Suplemento nº 1 da Recomendação R 128 da EBU (2016a):

- nível máximo de pico verdadeiro = -1 dBTP;
- *loudness* de curta duração = -18 LKFS;
- *loudness* momentâneo = -15 LKFS.

Os testes de medida e processamento foram feitos sobre um conjunto de blocos de programação e de intervalo comercial provenientes de uma emissora

de radiodifusão digital de sons e imagens cujas trilhas de áudio foram extraídas do *transport stream* capturado, e sobre suas contrapartes distorcidas pela demodulação do áudio por placas de captura.

Integrações de loudness

O experimento usou os descritores curtos anteriormente mencionados *loudness* de curta duração e *loudness* momentâneo. Os vetores são calculados na forma da [Equação 3.53](#) e da [Equação 3.54](#) usando uma janela deslizante retangular de duração T correspondendo a 3 s e a 400 ms, respectivamente. Este último conta com introdução de um filtro IIR de primeira ordem na saída, descrito pela equação de diferenças de mesma formulação da [Equação 3.34](#) e da [Equação 3.36](#) como no modelo de [Chalupper e Fastl \(2002\)](#). As medidas de pico verdadeiro deram-se com a implementação do diagrama em blocos da [Figura 3.14](#).

Considerando $f_s = 48000$ amostras por segundo, a trilha de áudio é processada em *streaming* em blocos de 144000 amostras. São feitas quinze medidas de *loudness* momentâneo em cada sub-bloco de 19200 amostras com 50% de sobreposição, uma medida de *loudness* de curta duração e outra de pico verdadeiro para todo o bloco. A violação de qualquer um dos critérios anteriores dispara um expensor dinâmico para controlar os níveis de *loudness* de saída. Para os casos de pico verdadeiro superiores a -1 dBTP, os valores acima do limiar são atenuados em 1 dB para evitar ceifamento. O bloco “descomprimido” volta à cadeia de sinal e o bloco subsequente é capturado, classificado, medido e expandido se necessário. O processo é realizado continuamente durante a transmissão.

Expansão dinâmica

Abordagens tradicionais de expansão dinâmica derivam suas curvas de ganho de forma direta, usando os valores absolutos de amostras multiplicados por uma razão escalar ([ZÖLZER, 2011](#)). Por outro lado, compressores e expansores modernos empregam filtros de suavização de um único polo nos estágios de detecção de nível e picos de modo a se minimizar artefatos indesejáveis devido às variações abruptas das magnitudes das amostras no estágio de cálculo de ganho ([GIANNOULIS; MASSBERG; REISS, 2012](#)). Neste trabalho, a preservação das

partes mais intensas e a atenuação das partes mais suaves é decidida com base num limiar calculado por sobre a envoltória do sinal, detectada por transformada de Hilbert (FEILAT, 2006). Um sinal de áudio (de valores reais) pode ser escrito da forma:

$$x[n] = x_r[n] + jx_i[n], \quad (4.31)$$

onde $x_r[n]$ e $x_i[n]$ também são sequências reais. O sinal $x[n]$ é denominado *sinal analítico* e sua transformada discreta de Fourier de tamanho N , iguala zero na metade inferior do círculo unitário ($-\pi \leq \frac{2k\pi}{N} \leq 0$) (OPPENHEIM; SCHAFER, 2013). Sejam $X_r[k]$ e $X_i[k]$ as transformadas discretas de Fourier de $x_r[n]$ e $x_i[n]$, então:

$$X[k] = X_r[k] + jX_i[k], \quad (4.32)$$

sendo

$$X_r[k] = \frac{1}{2} [X[k] + X^*[k]] \quad (4.33)$$

e

$$jX_i[k] = \frac{1}{2} [X[k] - X^*[k]]. \quad (4.34)$$

Sendo $X[k] = 0$ para $-\pi \leq \frac{2k\pi}{N} \leq 0$, não há sobreposição entre as partes diferentes de zero de $X[k]$ e $X^*[k]$. Logo, o sinal analítico $X[k]$ pode ser recuperado a partir unicamente de $X_r[k]$ ou de $X_i[k]$. A relação entre $X_r[k]$ e $X_i[k]$ é da forma:

$$X_i[k] = \begin{cases} -jX_r[k], & 0 \leq \frac{2k\pi}{N} \leq \pi, \\ jX_r[k], & -\pi \leq \frac{2k\pi}{N} \leq 0. \end{cases}, \quad (4.35)$$

ou

$$X_i[k] = H[k]X_r[k], \quad (4.36)$$

onde

$$H[k] = \begin{cases} -j, & 0 \leq \frac{2k\pi}{N} \leq \pi, \\ j, & -\pi \leq \frac{2k\pi}{N} \leq 0. \end{cases}. \quad (4.37)$$

Este sistema linear é denominado *deslocador de fase em 90 graus* ou simplesmente *transformador de Hilbert*. Sua resposta ao impulso, a partir da Equação 4.37, é dada por:

$$h[n] = \frac{1}{N} \sum_{k=0}^{\frac{N}{2}-1} j e^{j(\frac{2\pi}{N})kn} - \frac{1}{N} \sum_{k=\frac{N}{2}}^{N-1} j e^{j(\frac{2\pi}{N})kn} \quad (4.38)$$

ou

$$h[n] = \begin{cases} \frac{2}{\pi} \frac{\text{sen}^2(\frac{\pi n}{2})}{n}, & n \neq 0, \\ 0, & n = 0 \end{cases}. \quad (4.39)$$

através da qual $x_i[n]$ e $x_r[n]$ podem ser obtidos um a partir do outro por convolução com $h[n]$:

$$x_i[n] = \sum_{m=-\infty}^{\infty} h[n-m]x_r[m] \quad (4.40)$$

e

$$x_r[n] = - \sum_{m=-\infty}^{\infty} h[n-m]x_i[m]. \quad (4.41)$$

Portanto, um sinal real $x_r[n]$ na entrada de um transformador de Hilbert produz uma saída real $x_i[n]$ e ambos compõem o sinal $x[n]$ representado na [Equação 4.31](#). A magnitude do sinal analítico $|x[n]|$ é denominada *amplitude instantânea* e corresponde à envoltória do sinal de entrada $x_r[n]$ utilizada pelo expansor dinâmico.

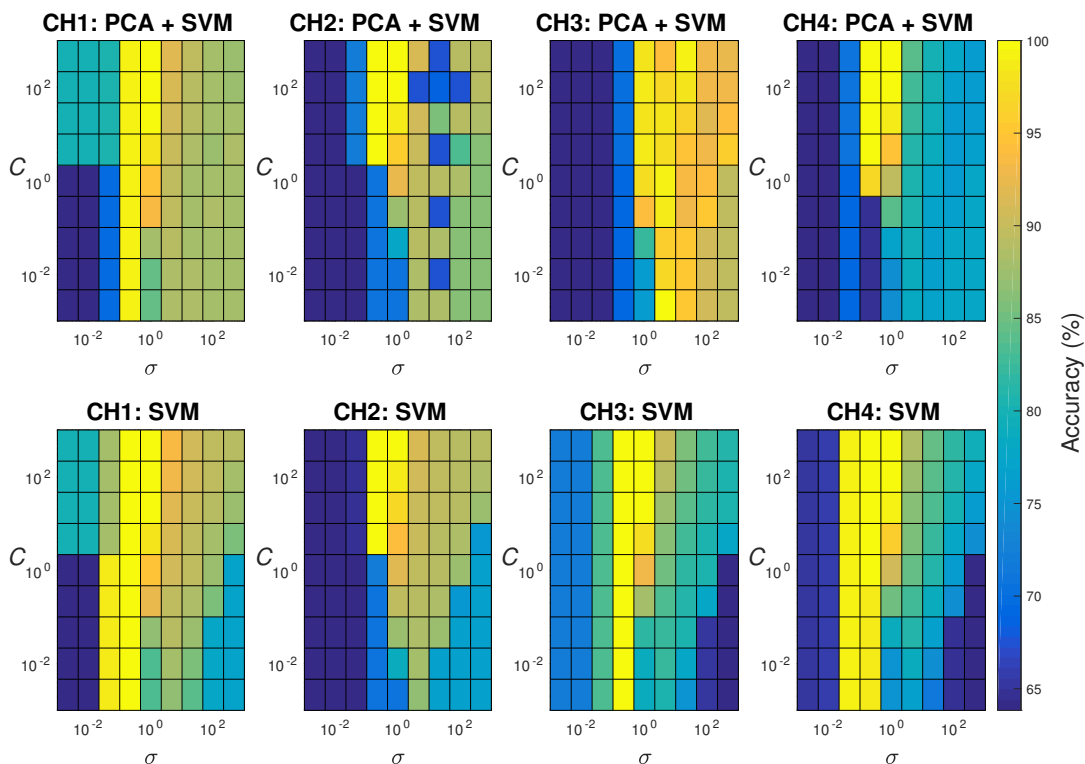
A atenuação é empregada nas amostras da envoltória cujas magnitudes se situam abaixo de um limiar relativo calculado como sendo a razão entre o nível da amostra e um limiar absoluto ajustado 10 dB abaixo do pico verdadeiro medido no segmento de áudio. O valor foi escolhido por duas razões:

1. Uma variação de nível em 10 dB corresponde aproximadamente a um fator de dois na sensação de *loudness* pela Lei de Potência de Stevens com $n = 0,3 \rightarrow 10^{0,3} \approx 2$ (ver [Capítulo 2](#)).
2. Um fator de crista – diferença entre níveis de pico e valores RMS – de 10 dB é o fator de crista aproximado que seria obtido por uma gravação analógica em fita, prática anterior à hiper-compressão dinâmica de peças de áudio ([VICKERS, 2010](#)).

4.2.3 Resultados e discussões

A efetividade do detetor de conteúdo proposto foi avaliada pelos percentuais de acerto no conjunto de validação para os modelos SVM treinados com diferentes valores de termo de regularização C e de desvio padrão do *kernel* gaussiano σ , variando entre 10^{-3} e 10^3 numa escala logarítmica. Os resultados obtidos para quatro canais de TV são ilustrados na [Figura 4.4](#).

Figura 4.4 – Treinamento e sintonia do detetor de conteúdo de formato curto.



Fonte: Elaborada pelo autor.

Nota – As cores indicam os percentuais de acerto em conjuntos de validação com características de áudio provenientes de quatro canais de televisão. Resultados para dados de dimensões reduzidas são ilustrados na primeira linha e desempenhos de classificação no espaço de características original são ilustrados na segunda linha. Diferentes conjuntos de termos de regularização C e de desvios padrão do *kernel* gaussiano resultam em valores distintos de acurácia da classificação.

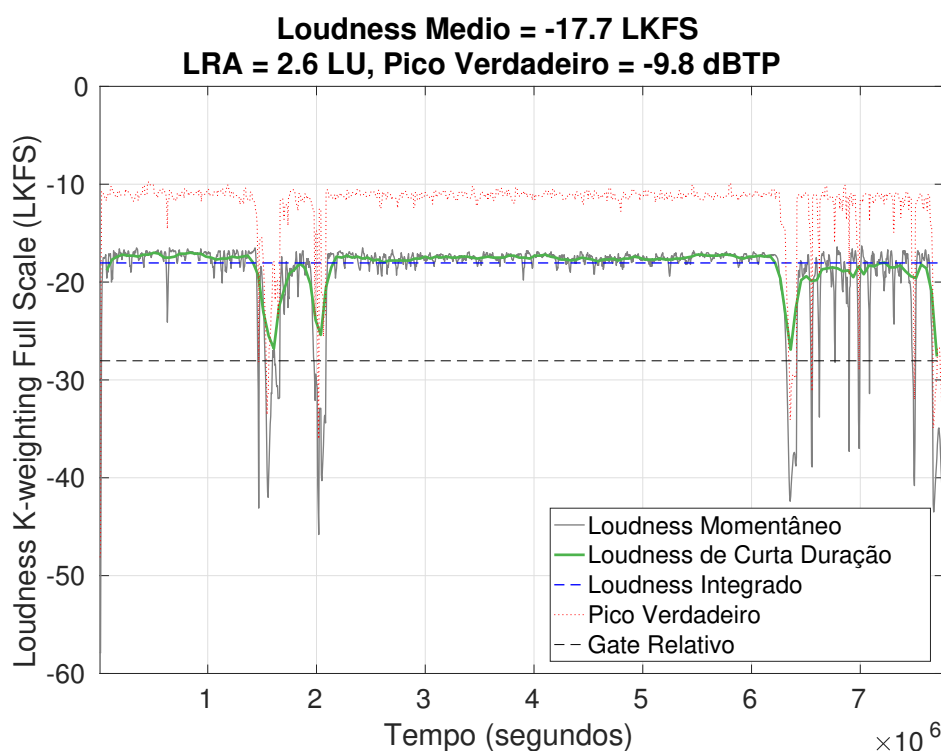
Note que, para curtos desvios padrão do *kernel* (pequenos valores de σ), uma regularização forte do modelo SVM (pequenos valores de C) suavizou a função objetivo do SVM e resultou em subajuste (*underfitting*). Por outro lado, a Figura 4.4 mostra que quando σ cresce, a acurácia também cresce quando a regularização do modelo é fraca, sugerindo que o modelo não tende a se sobreajustar (*overfit*) aos dados. Adicionalmente, classificadores com e sem dados reduzidos em dimensionalidade atingiram seus melhores desempenhos de classificação para uma mesma região de pares (C, σ) , ilustrando a importância de se usar um classificador bem sintonizado, como também sugerindo que as características selecionadas foram suficientemente relevantes e não-redundantes.

A significância estatística das diferenças de acurácia entre os classificadores com e sem emprego de PCA foram avaliadas usando dois testes estatísticos

unilaterais de McNemar (DIETTERICH, 1998), com a hipótese nula rejeitada no intervalo de confiança de 95% quando do uso de um número de componentes principais inferior a sete. Logo, o espaço de características de áudio, originalmente representado por dez características explícitas, pode ser representado por suas sete primeiras componentes principais retendo 99% da variância dos dados e sem perdas significativas de acurácia.

O controlador de *loudness* proposto processou trilhas de áudio cujos níveis violaram os critérios adotados, enquanto manteve as diferenças de integrações de *loudness* de curta duração entre blocos adjacentes inferiores a 1 LU para evitar saltos perceptíveis de intensidade nos programas veiculados. Sua efetividade pode ser ilustrada por um exemplo.

Figura 4.5 – Integrações de *loudness* ao longo de um trecho de um programa de televisão que intercala passagens de diálogo com números musicais altamente comprimidos em dinâmica

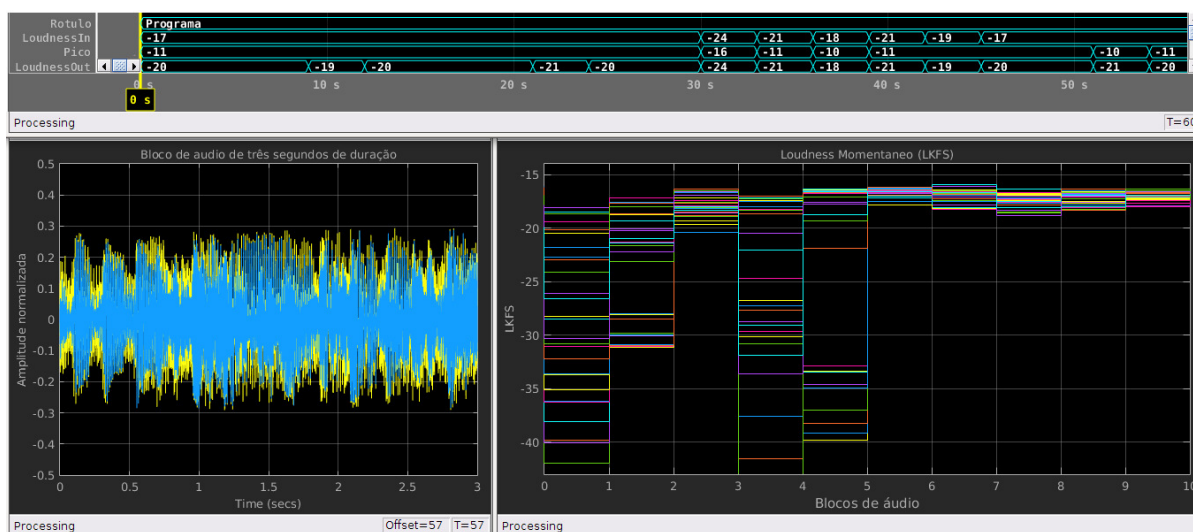


Nota – O vetor de *loudness* momentâneo é ilustrado em linha fechada cinza e o de *loudness* de curta duração em linha fechada verde. Nas partes hipercomprimidas, os valores de *loudness* de curta duração são ocasionalmente superiores a -18 LKFS, disparando o controlador de *loudness* proposto. As linhas tracejadas indicam o *loudness* médio em azul e o limiar relativo da função portão em preto. A linha pontilhada em vermelho indicam os valores máximos de pico verdadeiro atingidos.

Considere as integrações de *loudness* realizadas em tempo diferido ilus-

tradas na [Figura 4.5](#). Este extrato é oriundo de um programa de televisão que intercala passagens de diálogo com números musicais altamente comprimidos em dinâmica, no qual pode-se observar que as integrações de *loudness* de curta duração ultrapassaram o critério de nível máximo em alguns momentos. As medidas de *loudness* executadas ao longo do programa e as expansões dinâmicas pontuais na forma de onda de saída são exemplificadas na [Figura 4.6](#).

Figura 4.6 – Captura de tela do controlador automático de *loudness* atuando no mesmo conteúdo de áudio da [Figura 4.5](#) com duração de 160 segundos.



Fonte: Elaborada pelo autor.

Nota – A captura deu-se numa transição entre blocos de um diálogo com boa excursão dinâmica e blocos de um número musical comprimido em dinâmica. Integrações de *loudness* de curta duração de entrada e de saída, assim como medidas de pico verdadeiro são atualizadas no *display* lógico superior. Integrações de *loudness* momentâneo bloco a bloco são mostradas no monitor gráfico em escada à direita, e as expansões dinâmicas são exibidas em azul por cima da forma de onda original em amarelo no osciloscópio à esquerda. Um vídeo de demonstração está disponível na versão *online* da publicação de Pires, Vieira e Yehia (2017)

A captura de tela do controlador de *loudness* ilustra a transição de blocos de um diálogo com boa excursão dinâmica e blocos de um número musical comprimido em dinâmica, identificados pela dispersão das medidas de *loudness* momentâneo nos primeiros, e pela concentração das mesmas medidas nestes últimos. Quando as medidas de *loudness* de curta duração no bloco de áudio na entrada (*LoudnessIn*) ultrapassaram o valor de -18 LKFS, o bloco de expansão dinâmica foi acionado, gerando uma forma de onda de saída (em azul) com uma faixa dinâmica maior que a forma de onda de entrada (em amarelo). Um vídeo de demonstração está disponível na versão *online* da publicação de Pires, Vieira

e Yehia (2017).

Muito embora detectores de blocos comerciais que fazem uso de parâmetros audiovisuais produzam regiões de decisão mais complexas (VYAS *et al.*, 2014), um dos desafios deste trabalho residiu em descobrir se seria possível alcançar resultados utilizando um número limitado de parâmetros de áudio, dada a preocupação em produzir um classificador tão simples quanto possível tal que suas operações não introduzam retardos significativos. Trabalhos futuros deverão avaliar o impacto da introdução de outros descritores de baixa redundância no desempenho do classificador.

Cabe ressaltar que o detector de conteúdo de formato curto aqui proposto foi testado somente no conjunto de dados descrito na [subseção 4.2.1](#), pois o material bruto utilizado nos testes de processamento era insuficiente para o treinamento do algoritmo de aprendizado. Para fins práticos, o treinamento do SVM deve ser feito a partir da própria programação da emissora ao qual se destina.

Já a preocupação com o processador residiu principalmente na escolha do tamanho do bloco de processamento. O número de amostras escolhido corresponde a três segundos de reprodução, não somente pelo intervalo de integração para cálculo do *loudness* de curta duração, mas por ser um limitante confortável para um processamento ao vivo. Os retardos de grupo máximos introduzidos pelos filtros da curva *K* são inferiores a quinze amostras. Trabalhos futuros poderão avaliar a implementação desta solução em *hardware*, atentando para alterações do esquema para operação em aritmética de ponto fixo, a exemplo da modificação alertada no diagrama em blocos da [Figura 3.14](#).

A atuação do expensor dinâmico pode ser revista – e até mesmo substituída por outra – conforme o caso. Para emissoras que disponham de bons processadores de áudio, a solução proposta poderia gerar um sinal de controle para o processador e habilitar um *preset* de configuração de faixa dinâmica adequado.

Para fins de observação dos gatilhos oriundos do medidor de pico verdadeiro, o controlador de *loudness* também foi testado em conteúdo artificialmente comprimido e amplificado tal que resultasse em formas de onda com mínima

excursão de faixa dinâmica e com picos próximos do fundo de escala digital. Quanto mais próximo desta condição o sinal estiver, maior será sua transformação e maiores são os riscos de tornar a trilha sem presença. Embora o método aqui proposto destine-se aos casos de peças de programação ou propaganda não normalizadas previamente, não há melhor cenário do que peças originalmente mixadas com uma dinâmica saudável.

Discussões de implementação à parte, o objetivo principal desta proposta foi alcançado. Foi possível fazer uma leitura crítica da norma brasileira de *loudness* para a radiodifusão evidenciando potenciais correções e sugerir aprimoramentos dos seus dispositivos de controle baseados na experiência internacional e demonstrados de forma prática e eficiente.

4.3 Medição de *Loudness* para Áudio Espacial

Considerando o advento dos formatos de áudio imersivo elencados na [subseção 4.1.3](#) e a recente finalização do padrão MPEG-H 3D Audio da [ISO \(2015\)](#) de transmissão de áudio para Televisão em Ultra Alta Definição (UHDTV), cabe repensar a medição de intensidade sonora nesse contexto. Muito embora a revisão mais recente da Rec. ITU-R BS.1770 contemple o suporte a um número arbitrário de canais e alto-falantes, tanto o algoritmo padrão quanto os demais modelos de *loudness* existentes não foram suficientemente testados em tais condições. E, ao contrário do áudio tradicional baseado em canais, os conceitos de áudio baseado em objetos/cenas são agnósticos em relação ao número de alto-falantes. Nestes, o conteúdo de áudio é transmitido acompanhado de metadados ou em forma de coeficientes de harmônicos esféricos que descrevem todos os sons e suas propriedades espaciais variantes no tempo dentro de uma cena sonora, independentemente do número de fontes e da disposição dos alto-falantes na reprodução. Como consequência, para que algoritmos de controle de faixa dinâmica e *loudness* sejam aplicados no padrão MPEG-H 3D Audio, o conteúdo codificado deve ser renderizado num arranjo virtual de alto-falantes. Os sinais oriundos dos alto-falantes virtuais são controlados em intensidade e dinâmica, e daí recodificados até a etapa final de renderização ([PETERS et al., 2015](#)).

Este trabalho apresenta uma proposta de aprimoramento do algoritmo

ITU por meio de modificações em alguns de seus elementos psicoacústicos, mais especificamente por meio da substituição do pré-filtro de sombreado de cabeça pela simulação de som em salas de estar de referência seguida de síntese binauricular. As salas são simuladas por um modelo de fonte imagem a partir de suas respostas ao impulso e a síntese binauricular é feita com pares de Funções de Transferência Relativas à Cabeça (HRTFs) para cada alto-falante virtualmente posicionado na sala de estar simulada. A ponderação em frequências pela curva RLB também é modificada com a introdução de um pré-filtro projetado para modelar os efeitos do meato acústico. O método proposto foi testado em conteúdos para reprodução imersiva com oito canais (cubóide), nove canais (9.1) e vinte e dois canais (22.2), além de ter seu desempenho comparado com o algoritmo ITU e outros métodos identificados na literatura.

4.3.1 Métodos

Modelo de fonte imagem

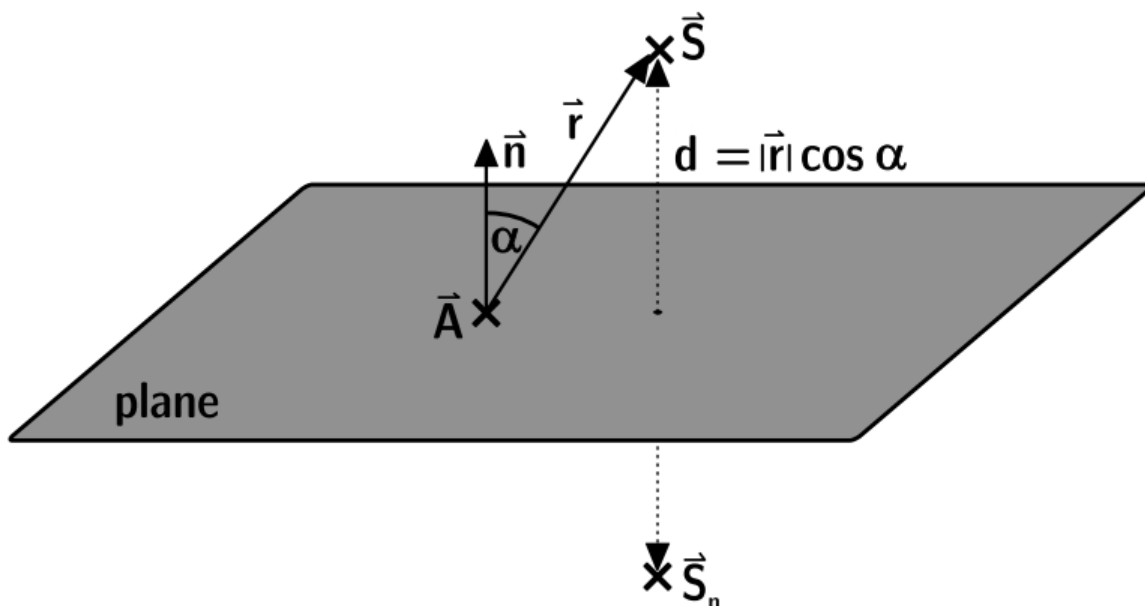
Pelo princípio da fonte imagem, a pressão sonora total do som direto e de suas múltiplas reflexões pode ser modelada pela soma das amplitudes das ondas esféricas correspondentes (VORLÄNDER, 2007). Na Figura 4.7(a), sejam \vec{S} a posição da fonte, \vec{S}_n a posição da fonte imagem, \vec{n} o vetor unitário normal ao plano \vec{A} , \vec{r} o vetor posição da fonte S , e d a distância entre a fonte e o plano obtida pelo produto escalar $\vec{r} \cdot \vec{n}$, a posição da fonte imagem é dada por:

$$\vec{S}_n = \vec{S} - 2d\vec{n}. \quad (4.42)$$

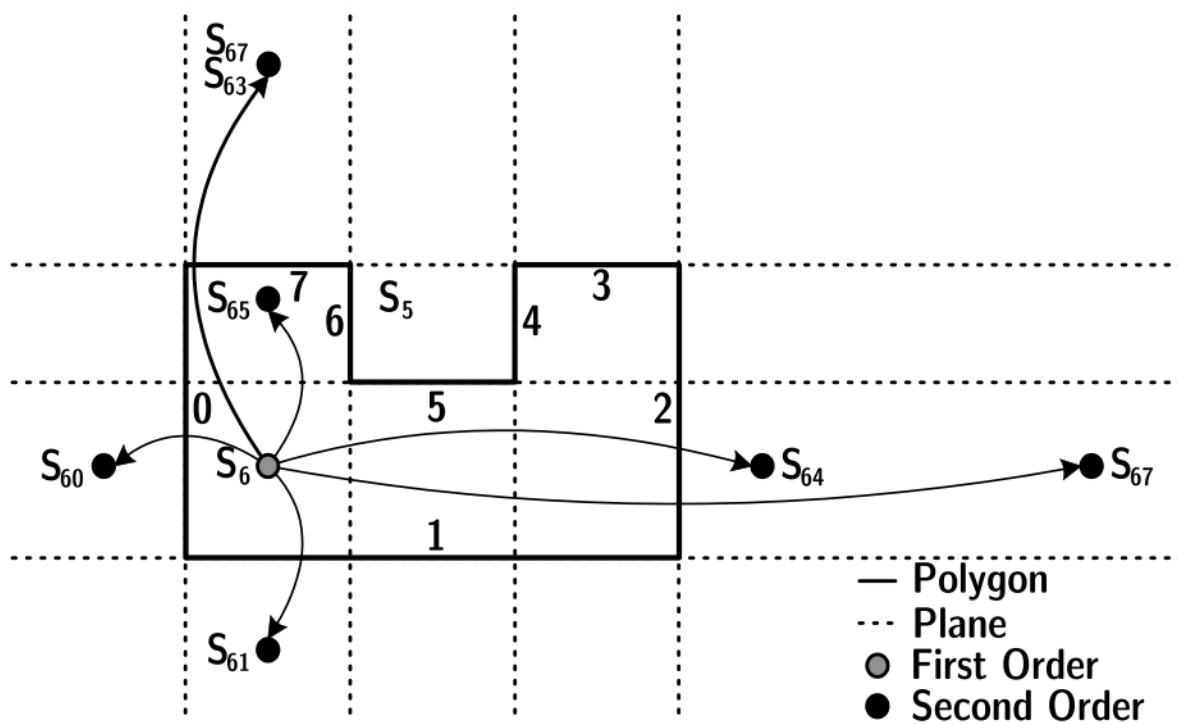
A Figura 4.7(b) ilustra este procedimento aplicado a todas as paredes de uma sala para a construção de fontes imagem de primeira ordem. Fontes imagem de ordens superiores são criadas da mesma maneira a partir das anteriores até um tempo limite de truncamento.

Para a decomposição das respostas ao impulso de salas num conjunto de fontes imagem \mathbf{f}_r , foi utilizado o método proposto por Tervo *et al.* (2013), no qual as pressões sonoras $x_i(n)$ de uma resposta ao impulso de uma sala, medidas num arranjo de microfones $i = 1, 2, \dots, I$, são apresentadas como a soma de todos os eventos acústicos $r = 1, 2, \dots, R$. O retardo associado à incidência de uma

Figura 4.7 – (a) Construção de uma fonte imagem. (b) Fontes imagem construídas para uma sala em duas dimensões.



Fonte: Vorländer (2007, p. 211).



Fonte: Vorländer (2007, p. 211).

frente de onda plana r num dado microfone i pode ser escrito como:

$$\tau_i = (d_r + \mathbf{m}_i \mathbf{n}_r^T) / c, \quad (4.43)$$

onde d_r é a distância da origem da frente de onda até o centro do arranjo, \mathbf{m}_i é a posição do microfone em relação ao centro do arranjo, \mathbf{n}_r é o vetor normal à frente de onda, c é a velocidade do som, e $(\cdot)^T$ é a notação de matriz transposta. Já a diferença de tempos de chegada da frente de onda em dois microfones distintos i e j pode ser escrita da forma:

$$\tau_{i,j} = \tau_i - \tau_j = (\mathbf{m}_i - \mathbf{m}_j) \mathbf{n}_r^T / c. \quad (4.44)$$

Os autores modelaram a medida das diferenças de tempo de chegada como um processo estocástico afetado por um ruído branco gaussiano w ($\hat{\tau}_{i,j} = \tau_{i,j} + w$), e a estimação de \mathbf{n}_r como um problema de minimização do erro médio quadrático destas medidas:

$$\hat{\mathbf{n}}_r = \arg \min_{\mathbf{n}_r} \sum_{\{i,j\}=1}^M (\tau_{i,j} - \hat{\tau}_{i,j})^2, \quad (4.45)$$

onde os pares de microfones $\{i, j\} = 1, \dots, M$ são $\{1, 2\}, \{1, 3\}, \dots, \{i, j\}, \dots, \{I, I-1\}$. A solução da direção de chegada da onda plana é dada em coordenadas cartesianas por:

$$\hat{\mathbf{n}}_r = \mathbf{V}^\dagger \hat{\boldsymbol{\tau}}, \quad (4.46)$$

onde $\mathbf{V} = [\mathbf{m}_1 - \mathbf{m}_2, \mathbf{m}_1 - \mathbf{m}_3, \dots, \mathbf{m}_I - \mathbf{m}_{I-1}]^T$ são as diferenças de posições entre pares de microfones, $(\cdot)^\dagger$ é a notação da matriz pseudo-inversa de Moore-Penrose e $\hat{\boldsymbol{\tau}} = [\hat{\tau}_{1,2}, \hat{\tau}_{1,3}, \dots, \hat{\tau}_{I,I-1}]^T$ são as diferenças dos tempos de chegada da onda plana a cada microfone. Estas são estimadas como sendo os argumentos máximos do vetor de correlação cruzada entre os sinais \mathbf{x}_i e \mathbf{x}_j em cada par de microfones, calculado pela Transformada Discreta Inversa de Fourier (IDFT):

$$\hat{\tau}_{i,j} = \arg \max_{\tau} (R_{i,j}(\tau)) / f_s = \arg \max_{\tau} (\text{IDFT}(\mathbf{X}_i \mathbf{X}_j^*)(\tau)) / f_s, \quad (4.47)$$

sendo o espectro \mathbf{X}_i obtido pela Transformada Discreta de Fourier (DFT):

$$\mathbf{X}_i = \text{DFT}([x_i(n - L/2), \dots, x_i(n + L/2 - 1)])(k), \quad (4.48)$$

onde n é o índice do quadro, L é o tamanho do quadro, f_s é a frequência de amostragem e $k = 1, \dots, K$ são as raias espectrais da DFT. Ao fim do cômputo,

têm-se as \mathbf{f}_r posições das fontes imagem em relação ao centro do arranjo de microfones, obtidas em função dos vetores unitários das direções de chegada na forma $\hat{\mathbf{n}}_r / \|\hat{\mathbf{n}}_r\|$, e da distância d_r calculada pela [Equação 4.43](#).

Funções de Transferência Relativas à Cabeça – HRTFs

Como visto na [subseção 3.3.1](#), o filtro K , cuja resposta em frequência é exibida na [Figura 3.12\(b\)](#), é dado pelo filtro RLB (ver [Figura 3.11](#)) precedido de um pré-filtro de sombreamento de cabeça, modelada como uma esfera rígida. Porém, as funções de transferência de um alto-falante à superfície da esfera são dependentes dos ângulos de incidência ([LARA; PASQUAL, 2014](#)). Este pré-filtro é um passa-altas com ganho de 4 dB para frequências superiores a 1 kHz e é tido como uma média das respostas ao longo dos ângulos de incidência de um ambiente de até cinco canais. Cumpre notar as diferenças anteriormente apresentadas entre as curvas de resposta em frequência do pré-filtro do algoritmo ITU-R na [Figura 3.12\(a\)](#), de HRTFs paramétricas na [Figura 2.16\(a\)](#) e de HRTFs medidas em laboratório na [Figura 2.17\(b\)](#). Logo, considerando áudio imersivo com um número de canais superior a seis, a média aproximada das HRTFs paramétricas de cinco canais utilizada como pré-filtro no algoritmo ITU não mais contempla todas as possibilidades. Por isso, neste trabalho usou-se uma base de HRTFs medidas em laboratório ([ALGAZI *et al.*, 2001](#)).

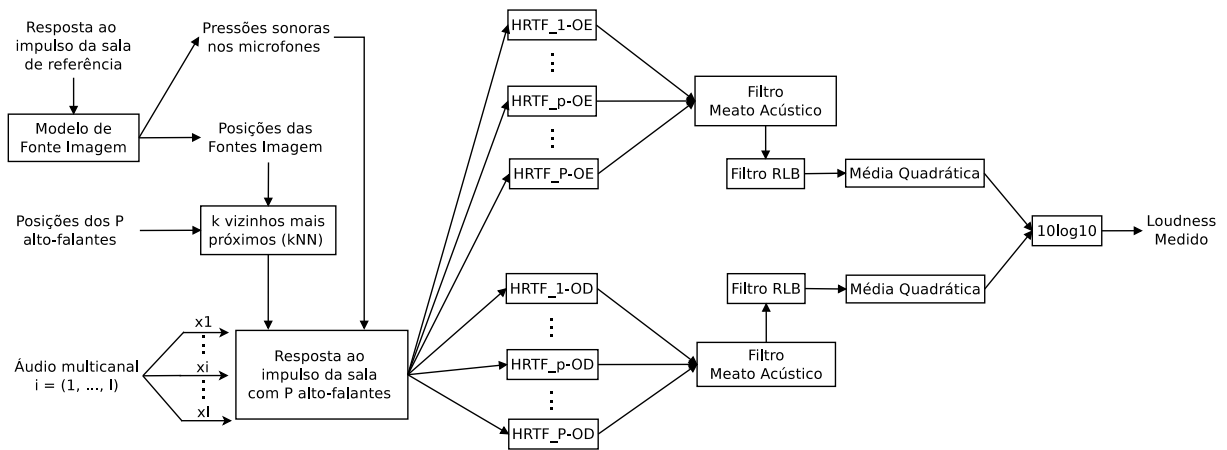
Modelo do meato acústico

De modo geral, as curvas de ponderação têm resposta plana em frequências superiores a 1 kHz, pois como visto na [subseção 3.2.1](#), resultados clássicos da psicoacústica sugerem que o ouvido interno seja igualmente sensível nessa faixa, e que as transformações na onda acústica entre 1 e 20 kHz ocorrem no ouvido médio ([ZWICKER; FASTL, 2013](#)). [Moore, Glasberg e Baer \(1997\)](#) levantaram esta função de transferência do meato acústico. Um filtro FIR de mesma resposta em frequência foi utilizado neste trabalho como filtro de ponderação juntamente com o filtro RLB do modelo de loudness cuja resposta em frequência é exibida na [Figura 3.5\(a\)](#).

4.3.2 Proposta

O diagrama em blocos do método proposto é ilustrado na [Figura 4.8](#). Este modelo de *loudness* modifica o algoritmo ITU-R numa tentativa de torná-lo mais adequado ao processamento de áudio imersivo com mais de 6 canais e de áudio baseado em cenas. Os objetivos aqui são contabilizar os efeitos acústicos da sala por meio de sua resposta ao impulso e preservar o agnosticismo do modelo de áudio baseado em cenas no que se refere ao número e à disposição de alto-falantes no sistema de reprodução do consumidor final.

Figura 4.8 – Diagrama em blocos do modelo de *loudness* proposto com o objetivo de se contabilizar os efeitos acústicos da sala por meio de sua resposta ao impulso e preservar o agnosticismo do modelo de áudio baseado em cenas no que se refere ao número e à disposição de alto-falantes no sistema de reprodução do consumidor final.



Fonte: Elaborada pelo autor.

Dada uma sala de estar de referência, esta é simulada para hospedar um arranjo virtual de P alto-falantes posicionados conforme dispostos no ambiente de reprodução do usuário. O mapeamento das fontes-imagem para os alto-falantes é dado da forma:

$$p = \arg \min_s (\| \mathbf{f}_r - \mathbf{a}_s \|), \quad s = 1, 2, \dots, P, \quad (4.49)$$

onde \mathbf{f}_r são as coordenadas das fontes imagem calculadas pelo método descrito na [subseção 4.3.1](#), \mathbf{a}_s é a posição do alto-falante em relação ao ponto de escuta e p é o índice do alto-falante mais próximo da fonte imagem em distância euclidiana. A [Equação 4.49](#) é computada considerando as fontes imagem como dados de entrada num problema de classificação, no qual os alto-falantes são as classes. Para cada amostra do sinal, a pressão sonora é associada à classe a qual a

fonte imagem pertence, computada pelo algoritmo dos k vizinhos mais próximos (k -NN) (DUDA; HART; STORK, 2012, Sub. 4.5.4.). A convolução do áudio de entrada com a nova resposta ao impulso obtida auraliza o sinal considerando os efeitos da sala simulada com os alto-falantes virtuais.

4.3.3 Testes

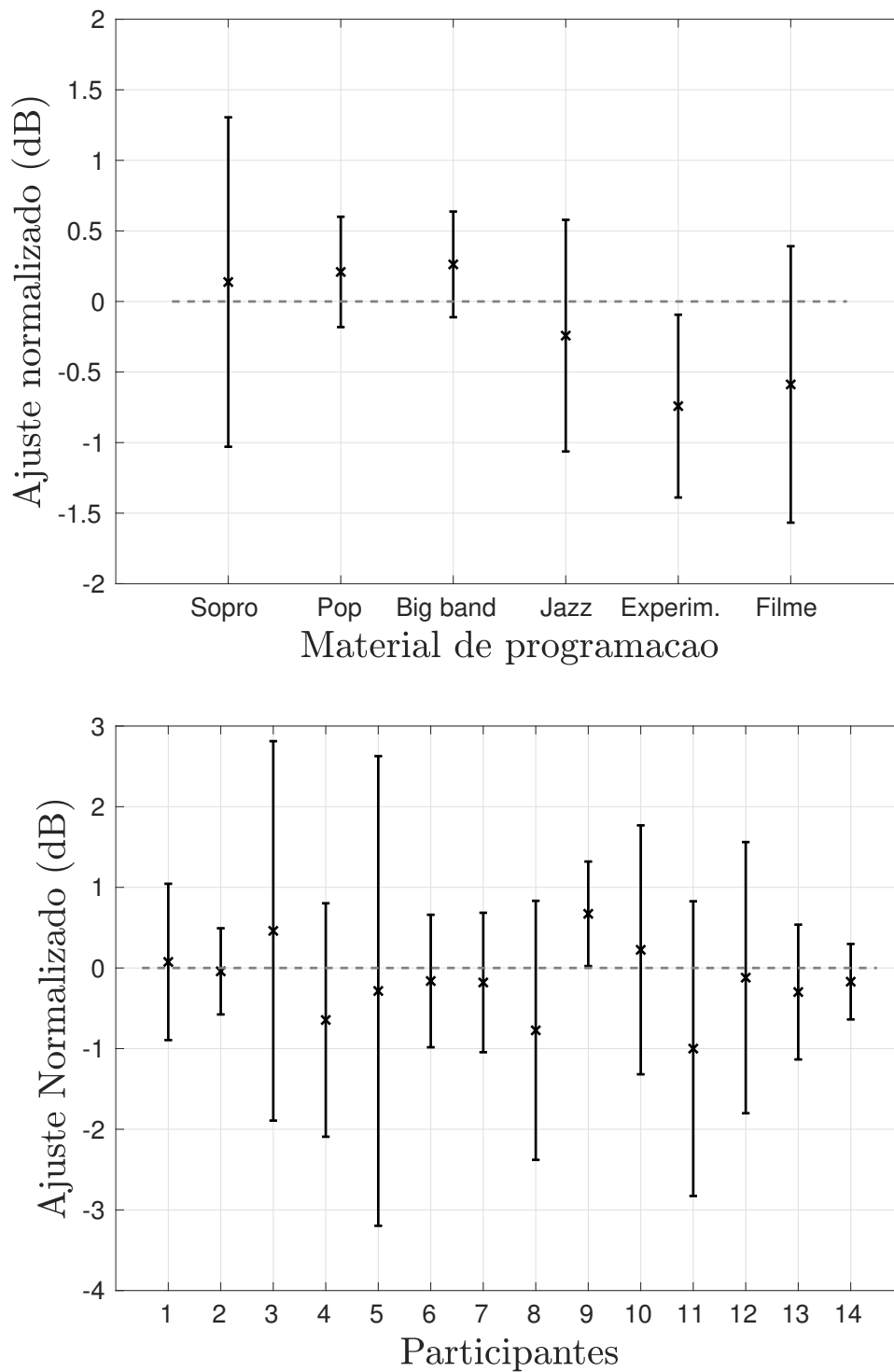
A disposição dos alto-falantes para áudio em formatos tradicionais de reprodução (mono, estéreo e 5.1) e em formatos de áudio imersivo (9.1, 22.2 e cuboide), seguiu a Recomendação ITU-R BS.2051, que especifica azimutes e elevações de cada um dos *layouts* de reprodução (ITU-R, 2014a).

O conteúdo de áudio imersivo utilizado neste trabalho foi elaborado para os testes auditivos de casamento de *loudness* conduzidos por Francombe *et al.* (2015a) e gentilmente cedido pelos autores. Nestes testes, 14 participantes ajustaram níveis de *loudness* com níveis de referência em cinco canais para trechos de 20 segundos de conteúdo musical e cinematográfico apresentados nos *layouts* de reprodução do parágrafo anterior. Os resultados categorizados por conteúdo e por participante estão ilustrados na Figura 4.9. Nos gráficos, as barras de erro mostram intervalos de confiança de 95% calculados utilizando a distribuição t de *student*. Médias superiores a zero indicam ajustes em níveis superiores aos níveis de referência e vice-versa.

Com relação às salas simuladas, foram utilizadas respostas ao impulso com 192 mil amostras por segundo de três salas de estar distintas (TERVO *et al.*, 2013). A primeira, com dois alto-falantes posicionados a um metro de distância do arranjo de microfones, a segunda com uma distância de dois metros entre microfone e alto-falantes e a terceira com a mesma distância da anterior, mas com alto-falantes ativos. As respostas ao impulso foram capturadas com uma sonda de intensidade vetorial G.R.A.S. 50VI, cujas dimensões foram parâmetros de entrada do arranjo de microfones no modelo de fonte imagem. A caracterização das salas é dependente do número e das posições dos alto-falantes, assim como da posição do arranjo de microfones. Suas respostas espaciais e em frequência são ilustradas na Figura 4.10, na Figura 4.11 e na Figura 4.12.

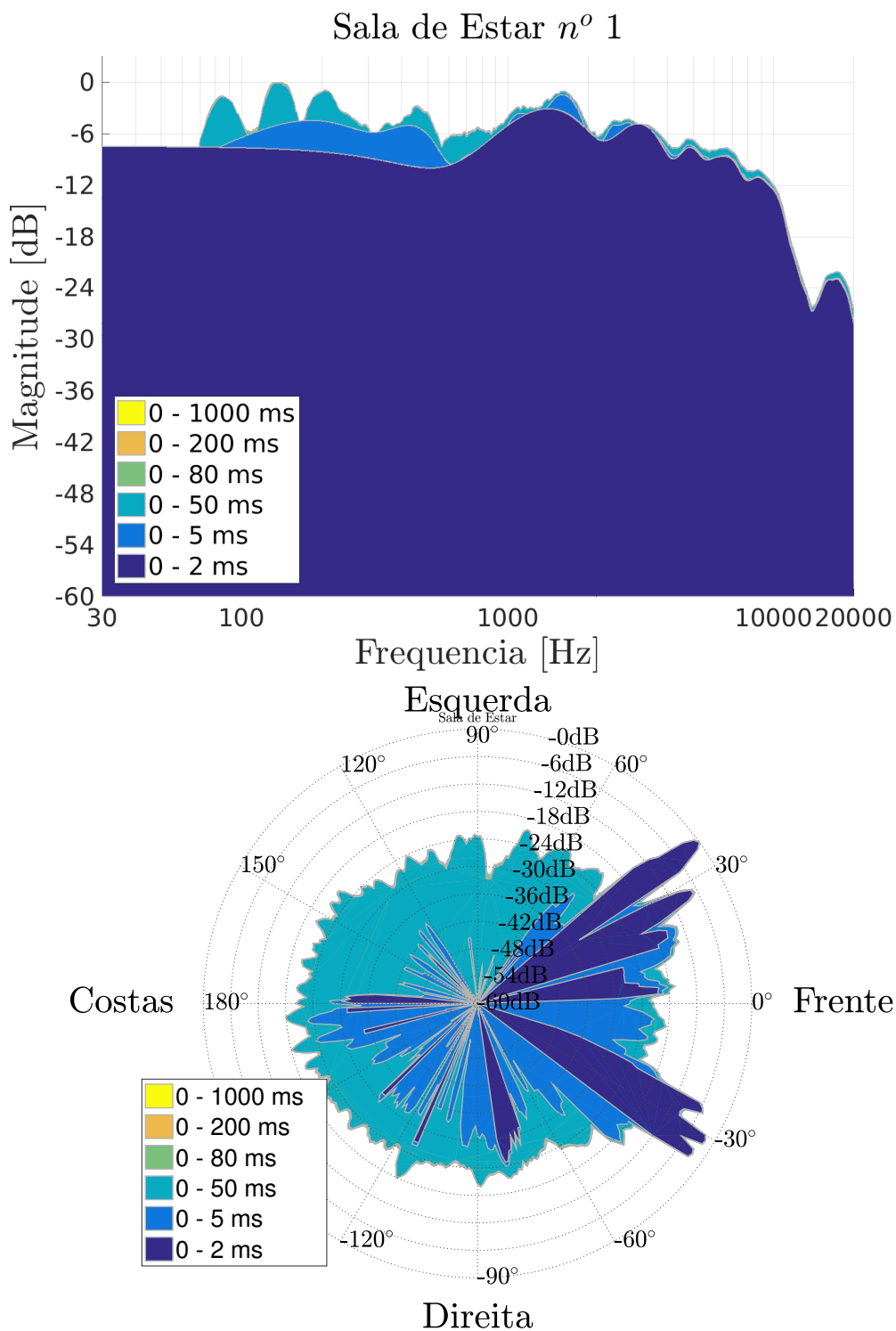
Cumprе notar que, em razão da menor distância entre alto-falantes e

Figura 4.9 – Ajustes de nível em relação ao conteúdo de referência (a) por tipo de conteúdo (b) por participante.



Fonte: Adaptada de [Francombe et al. \(2015a\)](#), dados da pesquisa).

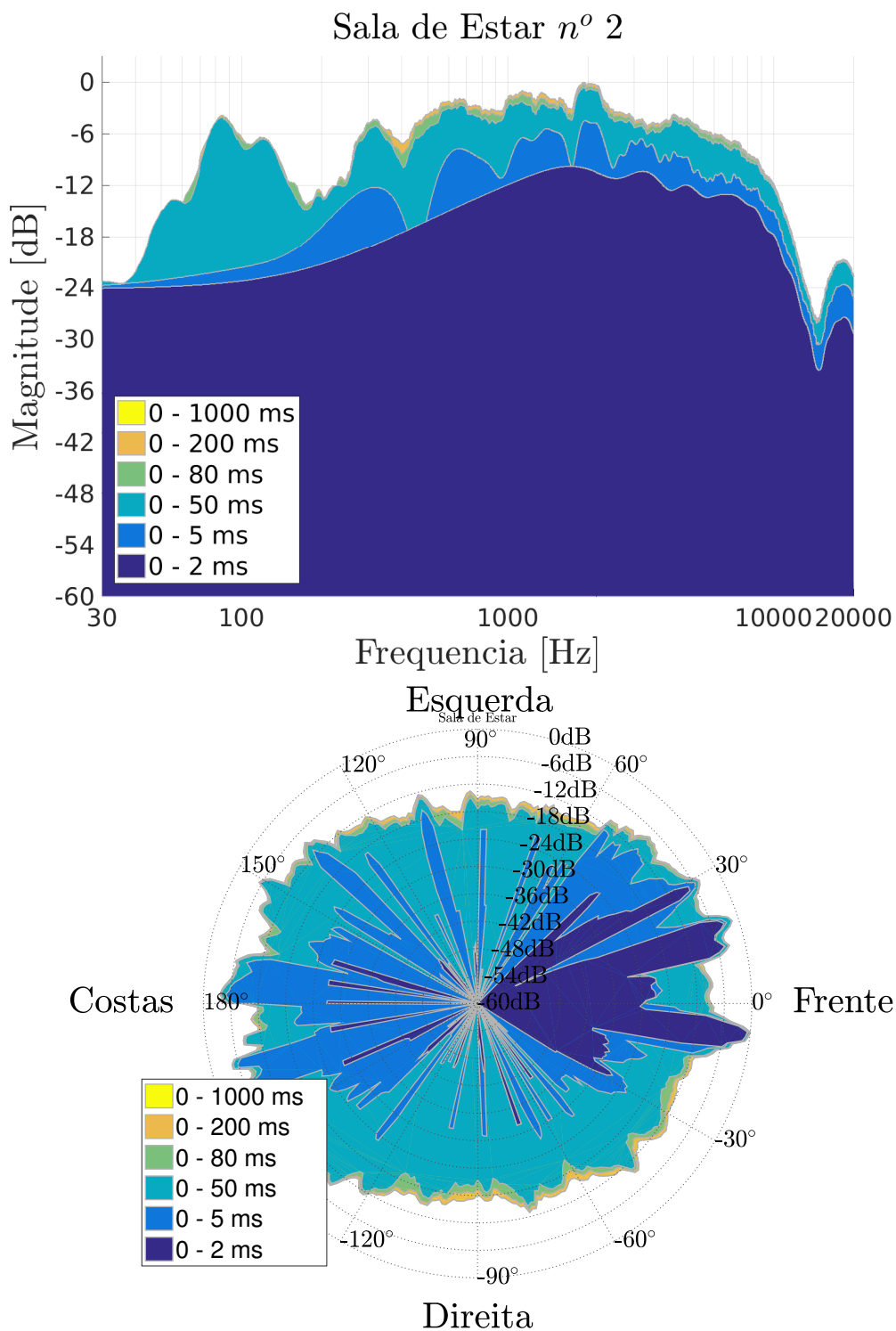
Figura 4.10 – Resposta em frequência e diagrama de captação da sala virtual nº 1



Fonte: Elaborada pelo autor.

Nota – Resposta ao impulso de sala estar com de dois alto-falantes posicionados a um metro de distância da sonda de intensidade vetorial G.R.A.S. 50VI capturada por [Tervo et al. \(2013\)](#).

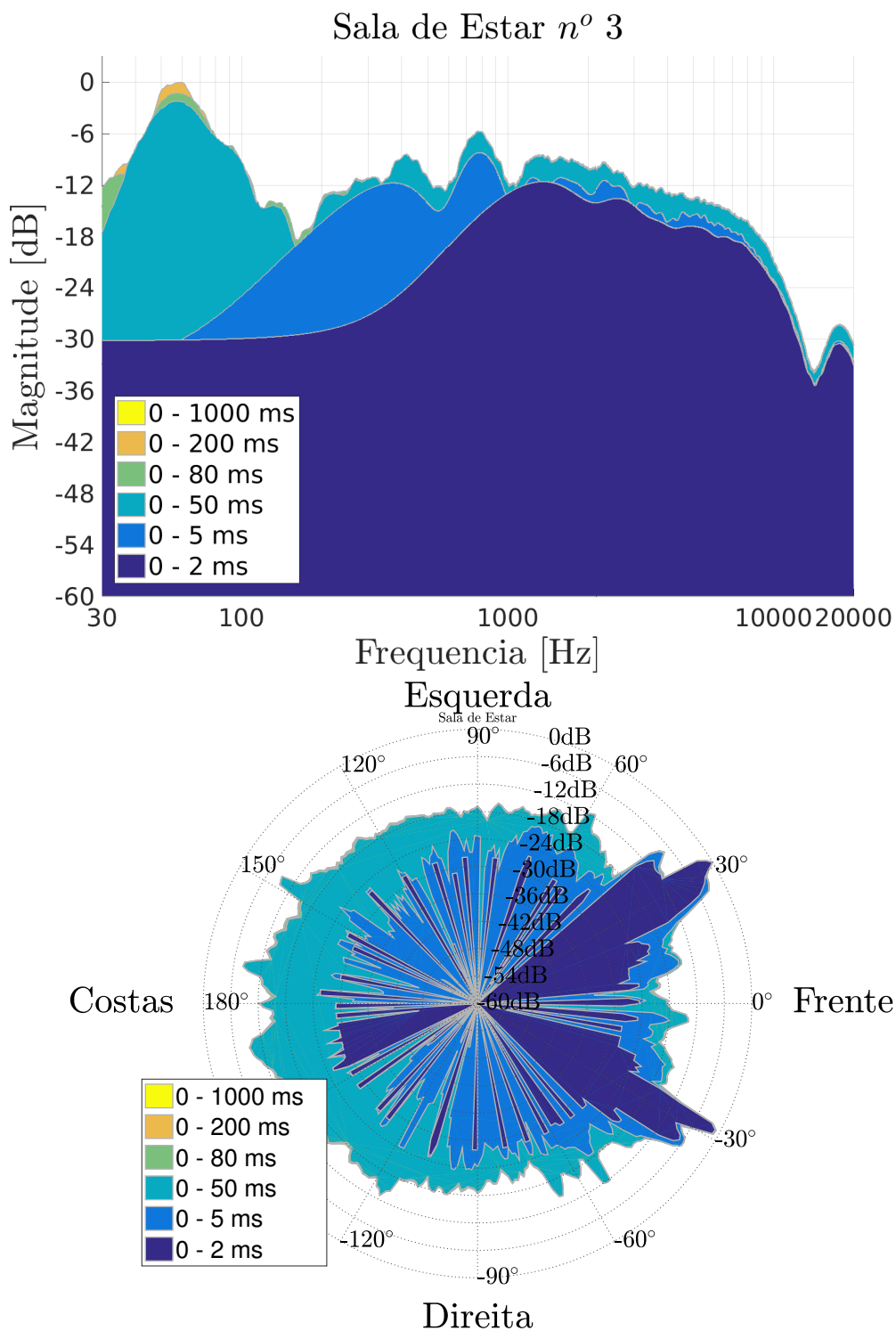
Figura 4.11 – Resposta em frequência e diagrama de captação da sala virtual nº 2



Fonte: Elaborada pelo autor.

Nota – Resposta ao impulso de sala estar com de dois alto-falantes posicionados a dois metros do arranjo de microfones capturada por Tervo *et al.* (2013).

Figura 4.12 – Resposta em frequência e diagrama de captação da sala virtual nº 3



Fonte: Elaborada pelo autor.

Nota – Resposta ao impulso de sala de estar de dois alto-falantes ativos posicionados a dois metros de distância do arranjo de microfones capturada por [Tervo et al. \(2013\)](#).

microfone, a contribuição da onda direta é maior na primeira sala. Por outro lado, a energia das respostas ao impulso das três salas está maciçamente concentrada na onda direta e nas reflexões com retardos inferiores a 50-80 ms, caracterizando-as como pouco reverberantes e com clarezas para fala (C_{50}) e música (C_{80}) satisfatórias para a condução dos testes (VORLÄNDER, 2007).

A fortuna crítica de modelos de *loudness* utilizada para fins comparativos é listada abaixo:

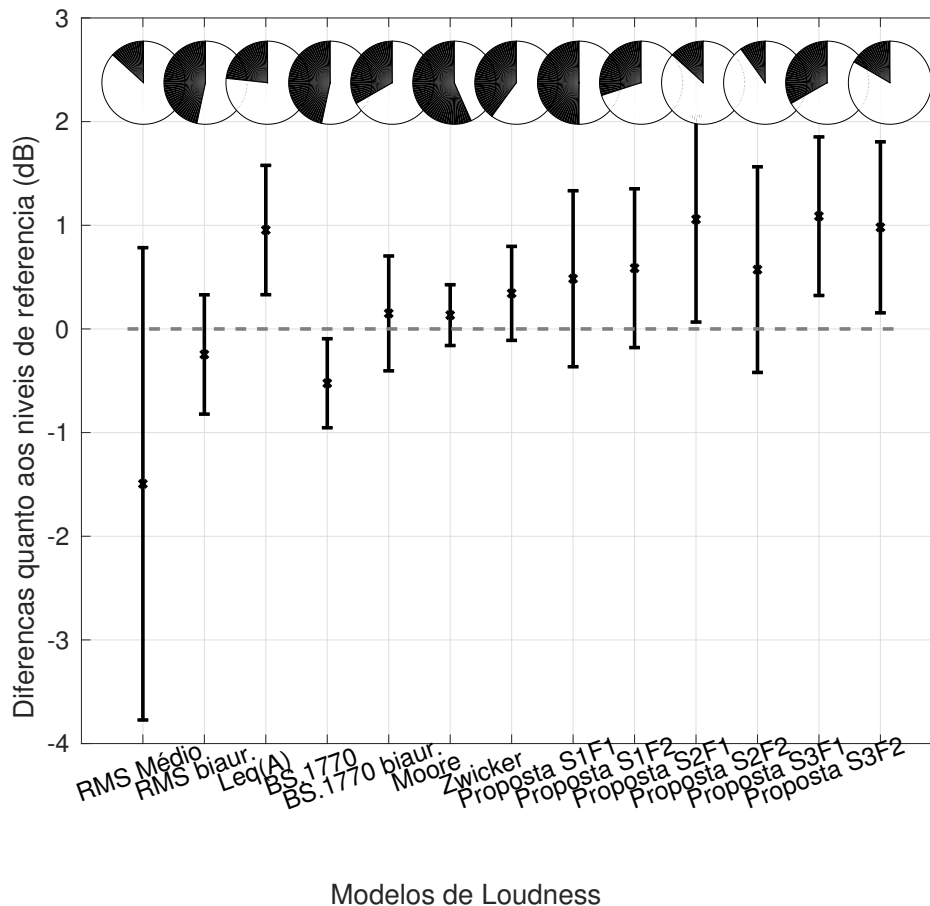
- Modelo clássico de detecção de energia da seção 3.1, ou RMS médio;
- RMS médio precedido de síntese biauricular;
- $Leq(A)_{(20s)}$, ou simplesmente decibelímetro;
- ITU-R BS.1770-4, para um número arbitrário de canais (ITU-R, 2015b);
- ITU-R BS.1770 com entrada biauricular por Pike e Melchior (2013);
- Modelo de Zwicker e Fastl (1999) para sons variantes no tempo;
- Modelo de Glasberg e Moore (2002) para sons variantes no tempo.

4.3.4 Resultados e discussões

As diferenças médias das predições de cada modelo de *loudness* em relação aos conteúdos de referência de cinco canais estão ilustradas na Figura 4.13. As predições do modelo proposto com técnicas de auralização são ilustradas na metade posterior do gráfico em seis combinações entre as três salas virtuais (S1, S2 e S3) e duas filtragens de ponderação, a primeira (F1) somente com a curva *RLB* e a segunda (F2) contando com a inclusão do filtro de transmissão do meato acústico. Os segmentos escuros nos gráficos do tipo pizza representam as predições que caíram dentro dos intervalos de confiança dos testes subjetivos.

A contabilização dos efeitos da sala fez com que as médias das predições fossem todas superiores aos níveis de referência. Embora as médias da sala n^o 1, que conta com mais energia na resposta ao impulso da onda direta, sejam mais próximas da referência que as das salas n^o 2 e 3, cujas respostas ao impulso são mais influenciadas pelas primeiras reflexões, a introdução do filtro de transmissão do meato acústico teve um impacto positivo nestas últimas, diminuindo as diferenças entre suas médias e os valores de referência. Com exceção da configuração “sala n^o 2 sem filtro de transmissão”, as médias das predições

Figura 4.13 – Diferenças médias entre predições para os demais sistemas de áudio em relação à referência de cinco canais. Os segmentos escuros nos gráficos do tipo pizza representam as predições que caíram dentro dos intervalos de confiança dos testes subjetivos.



Modelos de Loudness

Fonte: Elaborada pelo autor.

Nota – Os segmentos escuros nos gráficos do tipo pizza representam as predições que caíram dentro dos intervalos de confiança dos testes subjetivos.

do modelo proposto diferiram da referência em menos de 1 dB, apresentando melhor desempenho em relação ao decibelímetro tradicional. A configuração “sala nº 1 sem filtro de transmissão” se destaca por uma diferença próxima à do algoritmo ITU-R, em módulo.

Os intervalos de confiança mais largos em relação aos principais métodos apontam para a necessidade de calibração do modelo proposto. Tanto o algoritmo padrão quanto os demais modelos que representam o estado da arte em medição de *loudness* (Glasberg e Moore (2002), Zwicker e Fastl (1999) e BS.1770 biauricular) foram calibrados com o auxílio de testes auditivos de casamento de

intensidade sonora com uma das curvas de mesmo *loudness*, o que só reforça a necessidade de se fazer o mesmo com o algoritmo apresentado neste trabalho. Por outro lado, uma possível maneira de reduzir o intervalo de confiança seria com uma configuração que empregasse a resposta ao impulso da sala na qual os testes subjetivos foram realizados.

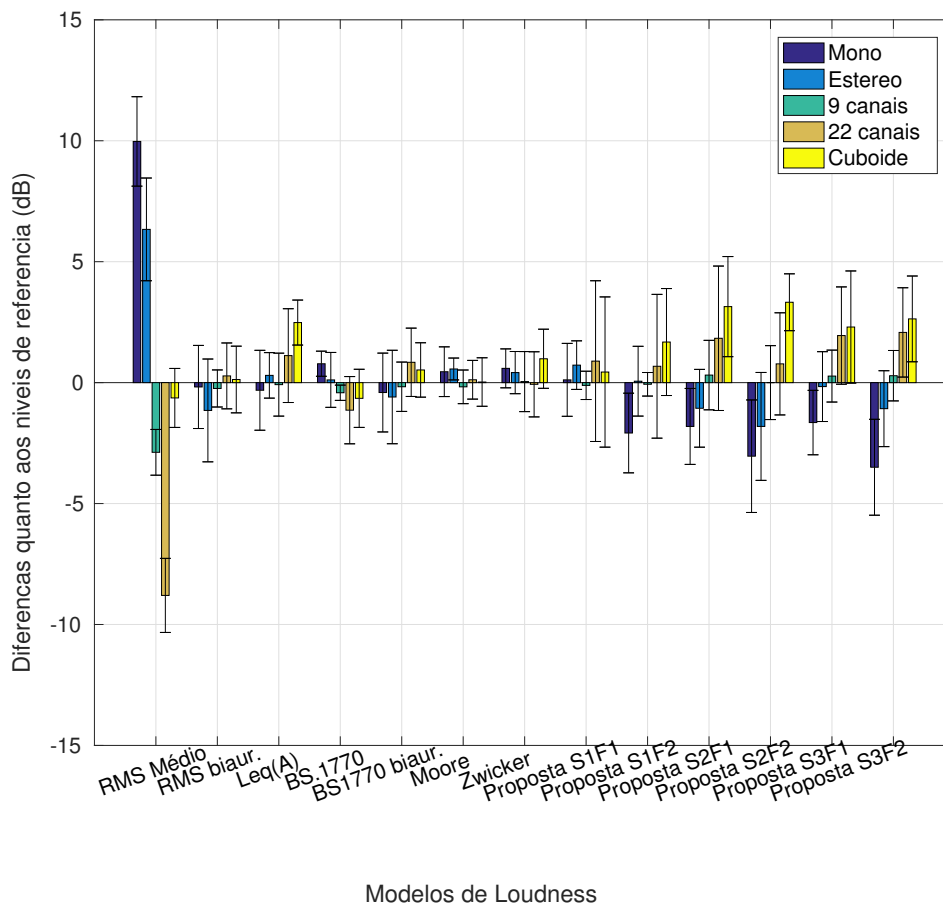
A configuração “sala nº 1 sem filtro de transmissão” também se destaca no que tange às predições que se situaram dentro do intervalo de confiança de 95% dos testes auditivos caracterizados na [Figura 4.9](#). Neste quesito, seu desempenho foi superior ao algoritmo ITU-R e aos demais métodos, com exceção do modelo de [Glasberg e Moore \(2002\)](#), o mais complexo e consistente dos modelos.

Os mesmos resultados discriminados por sistema de reprodução estão representados na [Figura 4.14](#). Para os sistemas de 9 canais, todas as configurações apresentaram erros tão baixos quanto os dos modelos consolidados. Apesar dos intervalos de confiança mais largos, a configuração “sala nº 1 sem filtro de transmissão” apresentou diferenças médias competitivas com os modelos na literatura em todos os sistemas de reprodução, com desempenhos marginalmente superiores aos do algoritmo ITU-R nos sistemas de áudio imersivo testados (9 canais, 22 canais e cuboide).

Em suma, a meta deste trabalho foi propor um método de medição objetiva que levasse em conta a influência de elementos não contemplados pelo padrão ITU-R BS.1770, como a função de transferência do meato acústico, funções de transferência relativas à cabeça (HRTFs) e efeitos acústicos de salas de estar de referência. O algoritmo proposto foi pensado para operação num decodificador MPEG-H 3D Audio, com atuação no controle de faixa dinâmica de conteúdo codificado espacialmente e com o uso de informações sobre o sistema de reprodução do consumidor final como parâmetros de entrada.

Os experimentos ilustraram a relevância do modelo proposto por meio da comparação com a fortuna crítica de modelos de *loudness*, envolvendo desde detetores de energia a modelos complexos do ponto de vista psicoacústico. A métrica utilizada foi a diferença de predições de um mesmo conteúdo em diferentes sistemas de reprodução, tomando os valores medidos no *layout 5.1* como referência. As estatísticas comuns sugerem erros médios inferiores a 1 dB. Em uma das configurações da proposta, foi observado um erro médio inferior

Figura 4.14 – Diferenças médias entre predições em relação à referência de cinco canais discriminadas por sistema de reprodução.



Fonte: Elaborada pelo autor.

e um maior número de predições dentro dos intervalos de confiança dos testes subjetivos, se comparada ao algoritmo ITU-R. Já nas estatísticas separadas por sistema de reprodução, a mesma configuração apresentou erros médios inferiores ao padrão BS.1770 para os três sistemas de áudio imersivo testados (9.1, 22.2 e cuboide).

Por outro lado, os intervalos de confiança das diferentes configurações do método proposto foram superiores aos dos modelos de *loudness* de referência. Como trabalho futuro, é sugerida a calibração em relação às curvas de mesmo *loudness* com o auxílio de testes subjetivos, de modo a se obter um escalamento adequado da cadeia de filtros utilizada.

PROPOSTAS E EXPERIMENTOS PRINCIPAIS

O [Capítulo 4](#) relatou a pesquisa recente em *loudness* para áudio digital, objetivando fundamentar uma compreensão sobre as necessidades correntes dos organismos de padronização e regulatórios, e assim prover contexto e relevância às perguntas formuladas na [subseção 4.1.4](#). Tanto a prova de bancada elaborada na [seção 4.2](#) quanto o medidor para áudio imersivo proposto na [seção 4.3](#) se prestaram a jogar alguma luz nessas questões, no que se refere ao aproveitamento das melhores práticas internacionais na atribuição de valores de referência e à importância de experimentos perceptivos para se modelar medidores mais próximos das sensações provocadas nos ouvintes.

Este capítulo se propõe a responder as perguntas da pesquisa apresentadas no capítulo anterior ao apresentar uma estratégia de revisão da norma brasileira e uma adaptação do modelo de ITU-R para objetos sonoros. A primeira proposta deriva do aprendizado obtido na construção do controlador da [seção 4.2](#) e da experiência de profissionais de radiodifusão. Já a segunda proposta origina-se de testes de hipóteses e de experimentos perceptivos referentes aos principais efeitos de localização de fontes sonoras na intensidade percebida, elaborados em sala de escuta crítica em conformidade com a Recomendação ITU-R BS.1116.

5.1 Estágio Doutoral

Após falar em melhores práticas internacionais, faz-se aqui um interlúdio para que eu mencione meu estágio doutoral de um ano na Universidade de Surrey, no Reino Unido. No período compreendido entre agosto de 2017 a julho de 2018, fui pesquisador visitante de seu Instituto de Gravação de Som (IoSR), referência mundial de pesquisa em engenharia psicoacústica. O instituto mantém relações com empresas importantes da indústria de áudio e possui instalações de primeira linha, incluindo um laboratório de áudio no formato de sala de audição em conformidade com a Recomendação ITU-R BS.1116, que dispõe sobre métodos para a avaliação subjetiva de pequenas deficiências em sistemas de áudio (ITU-R, 2015a), nos critérios de baixos ruídos, reflexões controladas, dimensões e tempos de reverberação (vide [Figura 5.1](#)).

Figura 5.1 – Laboratório de áudio no formato de sala de audição conforme com a Rec.ITU-R BS.1116 (ITU-R, 2015a)



Fonte: IoSR (2017, Instituto de Gravação de Som da Universidade de Surrey).

O laboratório é utilizado para experimentos de medidas em áudio e testes subjetivos de escuta e é equipado para diversos sistemas de reprodução, de estéreo até 22 canais. Sua conformidade com a Recomendação ITU-R BS.1116 foi fundamental para a observação de efeitos pequenos o suficiente para serem considerados como ruído experimental em ambientes comuns. Não há laboratórios estruturalmente similares em território brasileiro.

À época, havia um projeto em andamento no Instituto denominado Áudio Espacial Futuro para uma Experiência Imersiva do Ouvinte em Casa (S3A). Seu objetivo consistiu em habilitar a audiência para experimentar a sensação de se “estar lá” num evento ao vivo. O projeto teve início em 2013 e término em 2018

(IOSR, 2013). O período do Doutorado Sanduíche coincidiu com a fase final do projeto S3A, que focou no estabelecimento das diferenças-chave percebidas entre sistemas de reprodução e, no passo seguinte, modelar perceptivamente os mais importantes.

Na oportunidade, pude ter contato com especialistas na área e ganhar conhecimento para a condução dos meus próprios experimentos e testes de escuta no laboratório de áudio supracitado. O trabalho foi realizado sob supervisão do Diretor de Pesquisa do Instituto, Dr. Timothy Brookes, e do Coordenador da Graduação *Tonmeister*, Dr. Russell Mason, aos quais deixo aqui registrados meus mais sinceros agradecimentos.

5.2 *Loudness* na Radiodifusão Digital (revisitado)

Por mais que os experimentos de controle automático e de medição de áudio imersivo no [Capítulo 4](#) tenham resultado num maior discernimento sobre as perguntas desta pesquisa, a [subseção 4.1.3](#) deixou uma série de linhas de investigação em aberto. Foi preciso refletir sobre elas para só então fazer-se a difícil escolha dentre quais perseguir e quais deixar para trás. Para tanto, procurei conversar com profissionais de áudio para radiodifusão em busca de *insights* que pudessem me auxiliar na decisão. Durante meu estágio doutoral no Instituto de Gravação de Som na Universidade de Surrey, tive a oportunidade de conversar com funcionários da Rádio BBC Surrey e pesquisadores de áudio da BBC, com os objetivos de saber mais sobre medição de *loudness* no fluxo de trabalho de um estúdio de radiodifusão e sobre as discussões quanto ao aprimoramento do medidor no âmbito do ITU-R.

Como visto na [seção 1.3](#) e na [seção 4.2](#), a regulamentação brasileira é verticalmente estruturada, contando somente com uma Lei Federal de conteúdo geral que elenca a irregularidade e a pena (PR, 2001) e uma Portaria Ministerial que dispõe sobre os aspectos técnicos correspondentes (MC, 2012) em respeito à padronização internacional (ITU-R, 2015b; EBU, 2014). No caso do Reino Unido, a atuação do regulador é pautada num adendo (*Rule 47*) ao Código de Propaganda em Radiodifusão do Reino Unido (BCAP) que não só

elencar a irregularidade e a apenação, mas também estabelece a recomendação [ITU-R \(2015b\)](#) como referência. Já a recomendação da [EBU \(2014\)](#) e suas orientações técnicas quanto à especificação do medidor ([EBU, 2016b](#)), faixa de *loudness* ([EBU, 2016c](#)) e produção de peças ([EBU, 2016d](#)), são referências distribuídas setorialmente pela Parceria de Produção Digital (DPP), fundada pela BBC e pelas emissoras privadas *Independent Television* (Televisão Independente) (ITV) e *Channel 4* com o objetivo de padronizar a produção de conteúdo para radiodifusão no país.

“Setorialmente” – no caso em questão – refere-se à divisão entre padrões de conteúdo de programação de longa duração voltados aos radiodifusores ([DPP, 2018a](#)) e padrões de conteúdo de programação de curta duração voltados aos produtores de publicidade e propaganda ([DPP, 2018b](#)). Os níveis de referência correspondentes estão relacionados na [Tabela 5.1](#) e na [Tabela 5.2](#), respectivamente.

Ao comparar-se os valores na [Tabela 5.1](#) com os valores da Portaria [MC \(2012\)](#), foi possível obter algumas confirmações do que já se suspeitava, além de recomendações não antecipadas pela leitura crítica da Portaria Ministerial feita na [seção 4.2](#). Um desvio de ± 2 LU em relação ao nível de *loudness* de referência de -23 LKFS para o conteúdo de programação pode sim ser interpretado como permissivo, considerando desvios de $\pm 0,5$ LU para conteúdo de estúdio e de $\pm 1,0$ LU para conteúdo ao vivo, por exemplo. Um valor máximo de pico verdadeiro de -1 dBTP, citado apenas no procedimento de fiscalização da [Anatel \(2014\)](#) como prática a ser observada, é exigido para a radiodifusão britânica e acompanhado de um limite conservador recomendado de -3 dBTP. Já o parâmetro “Faixa de *Loudness*”, exigido como sendo inferior a 15 LU em programas e intervalos comerciais no Art 3º, inciso III, da Portaria 354/2012 do antigo [MC \(2012\)](#), é somente recomendado na [Tabela 5.1](#). A novidade aqui é a recomendação de faixa de *loudness* para diálogo de modo a garantir sua inteligibilidade independentemente do pano de fundo sonoro: sua dinâmica percebida não pode ultrapassar em 1/3 a dinâmica percebida máxima da programação, além de respeitar uma faixa dinâmica percebida “de guarda” entre diálogo e um fundo de música e efeitos. Esta medida não só garante QoE, como também prepara o ambiente de produção para os formatos imersivos de áudio na UHD TV, pois

Tabela 5.1 – Exigências de *loudness* para programas de formato longo no Reino Unido

Exigências DPP para entrega de conteúdo de formato longo			
<i>Loudness</i> Integrado (EBU, 2016d)	<i>Loudness</i> medido ao longo da duração do programa	LKFS	Conteúdo de estúdio: –23 LKFS \pm 0,5 LU Conteúdo ao vivo: –23 LKFS \pm 1,0 LU
Máximo Pico Verdadeiro (EBU, 2016d)	Valor máximo da forma de onda do sinal de áudio	dBTP	Fora de conformidade: > –1 dBTP Não recomendado: > –3 dBTP
Orientações de Faixa de <i>Loudness</i> somente para referência			
Faixa de <i>Loudness</i> (LRA) (EBU, 2016c)	Descreve a faixa dinâmica percebida medida ao longo da duração do programa	LU	Programas deverão atingir uma LRA não maior que 18 LU
Faixa de <i>Loudness</i> de Diálogo	Diálogo deve ser capturado e mixado tal que seja claro e de fácil compreensão	LU	Conteúdo de fala em programas deve atingir uma LRA de no máximo 6 LU É recomendada uma separação mínima de 4 LU entre diálogo e pano de fundo sonoro

Fonte: Adaptada de DPP (2018a, p. 15).

as codificações MPEG-H 3DA e Dolby AC-4 permitirão controlar o nível de diálogo separadamente do fundo sonoro (mais sobre isso na [subseção 5.2.2](#)).

Quanto às vinhetas e peças de propaganda, destaca-se que a [Tabela 5.2](#) se abstém de exigir ou recomendar Faixas de *Loudness* e passa a exigir um valor máximo de *Loudness* de Curta Duração de –18 LKFS, indo ao encontro do Suplemento nº 1 da Recomendação R 128 da EBU (2016a) quanto à não aplicabilidade do primeiro descritor e à recomendação do segundo para controle de conteúdo de formato curto, pelas razões dispostas na [seção 4.2](#). Estas exigências, datadas de 2018, corroboram o racional estabelecido na construção do controlador de *loudness* de Pires, Vieira e Yehia (2016) a menos de uma diferença regulatória fundamental: de que o conteúdo de formato curto não precisa

Tabela 5.2 – Exigências de *loudness* para programas de formato curto no Reino Unido

Exigências DPP para entrega de conteúdo de formato curto			
<i>Loudness</i> Integrado (EBU, 2016d)	<i>Loudness</i> medido ao longo da duração do comercial	LKFS	Conteúdo de estúdio: –23 LKFS $\pm 0,5$ LU Conteúdo ao vivo: –23 LKFS $\pm 1,0$ LU
Máximo Pico Verdadeiro (EBU, 2016d)	Valor máximo da forma de onda do sinal de áudio	dBTP	Não conforme: > –1 dBTP Não recomendado: > –3 dBTP
Máximo <i>Loudness</i> de Curta Duração (EBU, 2016a)	Nível máximo permitido de <i>loudness</i> de curta duração em conformidade com EBU (2016b)	LKFS	<i>Loudness</i> de Curta Duração Máximo: –18 LKFS

Fonte: Adaptada de DPP (2018b, p. 11).

ser controlado em tempo de programação, pois a responsabilidade sobre sua conformidade não recai sobre os radiodifusores, mas sim sobre os produtores das peças publicitárias, que poderão ser reprovadas para veiculação se não estiverem em conformidade com os valores recomendados.

5.2.1 *Sob que aspectos a norma brasileira de loudness pode ser revisada?*

Definidos os valores de referência e, principalmente, a atribuição de responsabilidades às partes interessadas, para prosseguir na formulação da resposta a esta primeira pergunta apresentada na [subseção 4.1.4](#), é preciso examinar as contribuições feitas ao ITU-R com relação às demandas do grupo de trabalho de *loudness*. Mais precisamente, quanto às questões relacionadas aos efeitos de frequência enumerados na [subseção 4.1.3](#): impacto da inclusão do canal de efeitos de baixa frequência (LFE) no cálculo do *loudness* e de um filtro passa-baixas de 18 kHz.

Em 2016, pesquisadores do Instituto Fraunhofer fizeram nova avaliação da contribuição do canal LFE para o *loudness*, independente da análise feita por

Mason (2011) que levou à exclusão do canal do cômputo no algoritmo original. Um teste de casamento de *loudness* foi feito com 15 ouvintes treinados, no qual o som sob teste deveria ser alinhado com uma referência calibrada para uma pressão sonora de 60 dBA na posição do ouvinte. O nível de calibração do *subwoofer* utilizado foi de 10 dB acima do nível dos canais principais (FRAUNHOFER, 2016). Os pesquisadores selecionaram 12 itens de teste, a partir de DVDs disponíveis comercialmente, pelo uso do canal LFE e pela quantidade de energia substancial em baixas frequências. Três variantes destes itens foram criadas: (i) reprodução do conteúdo 5.1 “no estado”, tal como nos DVDs; (ii) “Sem LFE”, ou seja, o conteúdo 5.1 foi reproduzido como sendo 5.0 e (iii) “Gerenciamento de graves”, no qual o sinal nos cinco canais principais é uma versão filtrada passa-altas do conteúdo do DVD e o sinal do canal LFE é a soma do sinal original do canal LFE do DVD com versões filtradas passa-baixas dos sinais dos 5 canais principais do DVD. A frequência de corte de todos os filtros foi de 100 Hz.

Os resultados do teste perceptivo foram comparados com o cálculo do *loudness* conforme o modelo ITU-R (2015b). As três variantes descritas no parágrafo anterior foram comparadas com três cômputos distintos: (i) sem LFE, (ii) com LFE e (iii) com LFE + 10 dB. Os coeficientes de correlação de Pearson (r) e os valores de Raiz Média Quadrática do Erro (RMSE) estão dispostos na Tabela 5.3. Cumpre notar que os cálculos de *loudness* incluindo o canal LFE possuem maior correlação com as respostas dos participantes para todas as variantes dos itens de teste, indicando que o canal LFE tem sua influência e deve ser considerado. Adicionalmente, o cálculo do *loudness* incluindo o canal LFE com um ganho de 10 dB possui o menor erro nas três variantes do item de teste, sugerindo que este ganho deva ser considerado no caso da inclusão do canal. Os autores recomendaram que este mesmo experimento fosse reproduzido em outro local, para confirmação das tendências observadas.

A questão do filtro passa-baixas com corte em 18 kHz chamou a atenção do grupo de trabalho do ITU-R ao se considerar que um sinal com raias espectrais de magnitude considerável em faixas não audíveis possa ser adicionado ao conteúdo resultando em medidas incorretas de *loudness*. Isso está relacionado à ponderação em frequência e à função portão implementada na Recomendação

Tabela 5.3 – Métricas de comparação entre cálculos de *loudness* e valores de testes perceptivos considerando inclusão do canal LFE.

Cômputo	“No estado”		“Sem LFE”		“Gerenciamento de graves”	
	<i>r</i>	<i>RMSE</i>	<i>r</i>	<i>RMSE</i>	<i>r</i>	<i>RMSE</i>
sem LFE	0,902	1,899	0,963	1,191	0,910	1,818
LFE	0,919	1,270	0,957	0,941	0,928	1,199
LFE + 10 dB	0,904	0,918	0,924	0,825	0,911	0,888

Fonte: Adaptada de Fraunhofer (2016, p. 20).

ITU-R (2015b). Como visto na Figura 3.12, a curva *K* não possui nenhum decaimento (*roll-off*) em alta frequência e, como visto na subseção 3.3.3, a função portão possui um limiar relativo para descarte de blocos de amostras que não contribuam para a sensação de *loudness* percebida. Logo, ao se inserir um tom inaudível no sinal de programação, se sua componente de alta-frequência contribuir com energia suficiente se comparada ao restante do conteúdo, ele manterá a função portão do algoritmo “aberta” de modo que segmentos de nível mais baixo contribuirão na integração de *loudness*, reduzindo o valor final da medida e podendo resultar em falsas condições de conformidade de uma dada peça de áudio.

O grupo relator testou modificações na curva *K* adicionando filtros passa-baixas do tipo Butterworth de segunda e oitava ordens, com frequências de corte em 16 kHz e 19 kHz (RG32, 2017). Em medidas feitas com 48 clipes de áudio monofônico e 8 clipes 5.1 (dentre estes, três filmes inteiros), as diferenças máximas entre o algoritmo original e suas modificações estão dispostas na Tabela 5.4. Cumpre notar que as diferenças encontradas foram marginais para conteúdos regulares, tornando estas modificações aplicáveis exclusivamente para mitigação de conteúdos ofensores, ou seja, contendo tons inaudíveis de alta energia.

Uma das razões para a adoção em massa da Recomendação BS.1770 deve-se à simplicidade do algoritmo: dois filtros IIR de segunda ordem seguidos de uma função portão. Quaisquer sugestões de aumento de complexidade requerem bastante discussão, e este parece ser o caso ao se considerar esta modificação da curva *K* para altas frequências. Verificações feitas pelo grupo MPEG e pela NHK mostraram que, para conteúdos com taxa de amostragem de 48.000 amostras/s,

Tabela 5.4 – Diferenças absolutas entre medidas de *loudness* feitas pela Recomendação BS.1770 do ITU-R (2015b) e suas variantes com filtros passa-baixas na saída da curva *K*

Curva <i>K</i> modificada	2ª ordem 19 kHz		8ª ordem 19 kHz		2ª ordem 16 kHz		8ª ordem 16 kHz	
	mono	5.1	mono	5.1	mono	5.1	mono	5.1
Diferença absoluta máxima (dB ou LU)	0,010	0,020	0,003	0,017	0,041	0,058	0,022	0,033

Fonte: Adaptada de RG32 (2017, p. 2-3).

frequências superiores a 16 kHz não são audíveis e um filtro passa-baixas de segunda ordem com corte em 16 kHz seria apropriado (RG32, 2017). Também é possível que a prática possa ser recomendada somente para conteúdos com taxas de amostragem mais altas, nos quais energias de raias espectrais até 24 kHz poderiam ser incluídas no cômputo do *loudness* integrado. É função do regulador acompanhar os desdobramentos desta discussão.

Respondendo – enfim – à pergunta desta subseção, ao se considerar o racional conduzido na seção 4.2 e divulgado em (PIRES; VIEIRA; YEHIA, 2016) e (PIRES; VIEIRA; YEHIA, 2017), aliada à experiência internacional reportada até aqui, uma proposta de reedição da Portaria n° 345 do antigo MC (2012) quanto aos seus valores de referência, passaria pelos itens abaixo relacionados:

1. Redução das tolerâncias do Art. 3º, parágrafo 1º, para $\pm 0,5$ LU, aplicáveis ao conteúdo de áudio digital produzido em estúdio (EBU, 2014; DPP, 2018a).
2. Inclusão de novo Inciso no Art. 3º, parágrafo 1º, definindo tolerâncias de $\pm 1,0$ LU, aplicáveis exclusivamente para conteúdo de áudio digital veiculado em programação ao vivo (EBU, 2014; DPP, 2018a).
3. No Art. 3º, parágrafo 1º, inciso III, restringir a exigência de um valor máximo de Faixa de *Loudness* somente para itens de programação (EBU, 2016a; PIRES; VIEIRA; YEHIA, 2016; DPP, 2018a).

4. No mesmo Art. 3^o, parágrafo 1^o, tanto para itens de programação quanto para intervalos comerciais, exigir-se um valor máximo de pico verdadeiro de -1 dBTP (EBU, 2014; PIRES; VIEIRA; YEHIA, 2016; PIRES; VIEIRA; YEHIA, 2017), cabendo recomendação de valor inferior de segurança (DPP, 2018a).
5. Com base na restrição do item 3, eliminar do Art. 4^o a exigência de durações mínimas de segmentos de áudio para efeito de fiscalização.
6. Discutir a adoção de descritores de formato curto para propagandas como *Loudness* Momentâneo e de Curta Duração (EBU, 2016a; PIRES; VIEIRA; YEHIA, 2016; DPP, 2018b).
7. Acompanhar o desenvolvimento das discussões sobre modificações do algoritmo e, dadas as diferenças máximas verificadas na Tabela 5.4 serem inferiores às novas tolerâncias propostas, considerar utilizar a versão modificada da curva de ponderação K com filtro passa-baixas Butterworth de 2^a ordem e $f_c = 16$ kHz testada em (RG32, 2017) para fins de fiscalização, objetivando mitigar vieses de medição causado por conteúdos de alta energia em frequências inaudíveis.

5.2.2 Como o modelo de loudness do ITU-R pode ser aprimorado para áudio imersivo?

No que tange aos novos formatos de áudio imersivo descritos na subseção 4.1.3 (baseado em canais, baseado em objetos e baseado em cenas), tanto o padrão de codificação MPEG-H 3D *Audio* quanto o padrão Dolby AC-4 foram desenvolvidos com a preocupação de que as características de um conteúdo de áudio devam se ajustar à condição individual do ouvinte, independentemente da origem ou do canal de distribuição deste conteúdo. Radiodifusores e *streamers* transmitindo um conteúdo não adaptável ao ambiente de reprodução – como é hoje – reduzem o problema normalização de *loudness* ao estabelecimento de um nível comum entre programas e intervalos comerciais e o alça a outros patamares: i) o consumidor precisa ajustar o volume de reprodução entre diferentes canais de distribuição porque o *loudness* não é consistente, ii) a inteligibilidade do diálogo

é afetada em suas partes mais suaves devido a um ambiente mais ruidoso, iii) a faixa dinâmica de uma peça pode ser larga demais para o nível de reprodução desejado (ex.: os níveis mais intensos de um filme o são ao ponto de irritação, enquanto as partes mais suaves simplesmente não são intensas o suficiente), iv) a faixa dinâmica de uma peça pode ser larga demais para o dispositivo de reprodução (ex.: alto-falantes de baixa qualidade em dispositivos portáteis) e v) um sinal de áudio imersivo muito intenso pode sofrer ceifamento quando remixado para um menor número de alto-falantes (*downmixed*).

Para tanto, os decodificadores de ambos os padrões contam com módulos de controle de faixa dinâmica e de normalização de *loudness* logo após a etapa de mapeamento dos sinais de áudio para o arranjo de alto-falantes no ambiente de reprodução. Para o áudio baseado em canais, este mapeamento é feito por um conversor de formato responsável pelo *downmix* dos sinais multicanal para um número inferior de canais específico do sistema de reprodução. Por outro lado, o mapeamento do áudio baseado em objetos para os canais dos alto-falantes é feito por um renderizador que, com base nos metadados transmitidos – ou nos metadados modificados pela interação do usuário – e na disposição dos alto-falantes na sala de reprodução, e o balanceamento dos alto-falantes é feito via Sistema Vetorial de Panorama por Amplitude (VBAP) (PULKKI, 1997) para criar a sensação de tridimensionalidade sonora. No caso dos sinais *Ambisonics* de áudio baseado em cenas, o mapeamento também é feito por um renderizador específico que opera baseado nos metadados HOA transmitidos. Só então o controle de faixa dinâmica e a normalização de *loudness* definitivos são executados para entrega dos sinais aos alto-falantes (RIEDMILLER *et al.*, 2017; BLEIDT *et al.*, 2017).

O algoritmo de medida de *loudness* em áudio imersivo proposto na seção 4.3, posteriormente divulgado por Pires *et al.* (2017), foi concebido ao encontro da ideia de flexibilidade do conteúdo de áudio para diferentes condições de reprodução, ao fazer uso de informações sobre disposição de alto-falantes – e da resposta ao impulso da sala de reprodução – de modo a estimar um nível de *loudness* mais próximo da experiência dos ouvintes. Todavia, a quantidade de processamento envolvido na renderização de objetos ou cenas sonoras, tanto no MPEG-H 3D Audio quanto no Dolby AC-4, não dispensa a necessidade de

normalização de *loudness* ao final do processo de decodificação. Tomo aqui a liberdade de citar trecho de uma conversa que tive em 2018 com o Sr. Andrew Mason, do setor de pesquisa e desenvolvimento da BBC e membro do grupo de trabalho de *loudness* do ITU-R:

No momento, a abordagem pragmática para se fazer a medida (ou estimação, como alguns preferem) do *loudness* de áudio baseado em objetos, é renderizá-lo para os canais dos alto-falantes e medi-los. Desta forma, não há necessidade de se levar em consideração, na medida de *loudness*, se o áudio é originalmente baseado em objetos ou em cenas, pois isso já foi levado em conta na etapa de renderização para os canais dos alto-falantes (**tradução minha**).

Salvo melhor juízo, independentemente da origem de áudio imersivo, o formato de entrada do modelo de *loudness* ITU-R continuará sendo de áudio multicanal ainda por muito tempo. Adicionalmente, do ponto de vista da produção, há benefícios em prover diferentes perfis de controle de faixa dinâmica para objetos individuais ou para grupos de canais. Por exemplo, o diálogo num filme pode ser de difícil compreensão num ambiente ruidoso. Nesse caso, seria vantajoso usar uma configuração específica de compressão para o objeto “diálogo”, ou para os canais que carreguem o diálogo de modo dominante, de modo a aumentar a inteligibilidade da fala. Analogamente, para um perfil “noturno” de controle de faixa dinâmica, os canais da “cama” sonora – ou pano de fundo sonoro – poderiam ser agressivamente comprimidos de modo a reduzir os picos de intensidade em cenas de ação, enquanto os trechos de diálogo teriam um processamento dinâmico mais natural (KUECH *et al.*, 2015). Este *layout* híbrido de alguns objetos sonoros ou de algumas cenas sonoras (acompanhado(a)s de metadados) por sobre uma cama de canais, é o que mais se aproxima de uma ideia de mixagem sonora para UHD TV. Em suma, não há perspectiva de uma “aposentadoria” do áudio multicanal para dar lugar aos novos formatos imersivos. Pelo contrário, ele proverá suporte às sensações de espacialidade e imersão provocadas por objetos sonoros destacados.

À luz de como os novos padrões de codificação de áudio para radiodifusão se anteciparam no que diz respeito ao controle de faixa dinâmica baseado em metadados e interações do usuário, fecha-se aqui a linha de investigação do algoritmo ITU-R quanto aos modelos de definição de áudio, aberta na [subseção 4.1.3](#)

e baseada nos termos de referência do grupo relator (ITU-R, 2016b). E pode ser repensada a partir dos termos de referência do ciclo de estudos anterior (ITU-R, 2014e), quando da elaboração da quarta edição da Recomendação BS.1770 em 2015, dentre os quais destaco três em particular:

- Checar se o filtro K deve ser dependente dos ângulos de azimute e elevação. Se sim, qualquer novo filtro K dependente da posição deve ser igual ao filtro K convencional quando o *loudness* de um sistema multicanal 3/2 (5 canais) é medido;
- Determinar os pesos G_i dependentes dos ângulos de elevação e azimute da posição de um elemento de áudio. Os novos G_i dependentes da posição deverão ser iguais aos G_i convencionais quando o *loudness* de um sistema multicanal 3/2 (5 canais) é medido;
- Determinar como o algoritmo de medida de *loudness* poderia incluir medidas de objetos sonoros que possuam localizações dinâmicas.

Mesmo com o objetivo de adaptar o modelo ITU-R para um número irrestrito de canais de áudio numa época em que ele era limitado a sistemas 5.1, estes termos de referência já traziam preocupações com relação à compatibilidade reversa e questões quanto à evolução do algoritmo rumo ao novo paradigma de áudio imersivo – mas principalmente sem focar nos formatos de áudio, e sim na posição dos objetos sonoros. Pode, então, ser possível aprimorar o modelo ITU-R ao se calcular o *loudness* de objetos sonoros a partir das informações posicionais contidas nos metadados que os acompanhem, mantendo o cômputo inalterado para a cama sonora baseada em canais.

Isto posto, os estudos referentes a esta pergunta da pesquisa que serão descritos nas próximas seções, dar-se-ão por experimentos perceptivos que investiguem as relações entre a sensação de *loudness* e atributos posicionais de fontes sonoras. A seção 5.3 abordará a relação do *loudness* com a distância e o quanto ela pode ser diferente da relação da intensidade com a distância da fonte sonora. Já a seção 5.4 trabalhará com a relação do *loudness* com a energia reverberante, capaz de torná-lo invariante com a distância. Na seção 5.5, investiga-se a influência direcional dada pelos ângulos de azimute e elevação, sendo este último não contemplado pela versão vigente da Recomendação BS.1770 à data desta reda-

ção. Por fim, a [seção 5.6](#) proporrá um modelo de *loudness* consolidado baseado nos experimentos anteriormente apresentados e comparará seus resultados com a fortuna crítica de modelos de *loudness* descrita no [Capítulo 3](#), como também com o método proposto na [seção 4.3](#), divulgado por [Pires et al. \(2017\)](#).

5.3 Relação entre *Loudness* e Distância

Experimentos progressos de *loudness* em câmaras anecoicas conduzidos por [Warren, Sersen e Pores \(1958\)](#) e [Stevens e Guirao \(1962\)](#) tiveram resultados consistentes com a relação de variação de intensidade proporcional ao inverso do quadrado da variação da distância, nos quais a ocorrência de meio *loudness* e dupla distância foram observadas próximas de um decremento de 6 dB. Estes experimentos foram conduzidos sob condições tais que a variação de intensidade era a dica principal – ou única – de variação de distância.

Por outro lado, sob condições nas quais mais dicas perceptivas estavam disponíveis para os ouvintes, relações inversas de proporcionalidade entre o *loudness* e a distância, não eram mais observadas. De fato, para sons em salas reverberantes, os experimentos de [Mohrmann \(1939\)](#) e [Zahorik e Wightman \(2001\)](#) reportaram, em épocas distintas, situações de invariância do *loudness* com a distância (*loudness constancy*). [Zahorik, Brungart e Bronkhorst \(2005\)](#) sugeriram que a informação sobre a energia reverberante pode ser a base das observações de invariância de *loudness* em salas de reprodução. Já na ausência de reverberação, os ouvintes usam somente a energia da onda sonora diretamente incidente para fazer seus próprios juízos de *loudness*, o que aproximaria as respostas destes experimentos com o comportamento conhecido da variação de intensidade sonora com a distância.

Isto pode ser interpretado como uma possível explicação para que as diferenças de níveis de *loudness* entre as medidas pelo método proposto na [seção 4.3](#) e as respostas dos participantes do experimento de [Francombe et al. \(2015a\)](#) ilustradas na [Figura 4.9](#) tenham se mostrado estatisticamente insignificantes, quando a resposta ao impulso da sala nº 1, menos reverberante, foi utilizada no cômputo do *loudness*. Na [Figura 4.13](#), também é possível observar que mais diferenças entre níveis de *loudness*, calculadas nas mesmas condições nas quais

a influência da energia reverberante foi menor, se situaram dentro dos limites dos intervalos de confiança das respostas dos participantes, se comparadas às versões que fizeram uso das respostas ao impulso das salas nº 2 e 3, com mais energia reverberante.

Stecker e Hafter (2000) reportaram descobertas similares num experimento que investigou a sensação de *loudness* com estímulos temporalmente assimétricos, quando estímulos de ataques rápidos e decaimentos lentos foram ouvidos de modo menos intenso com a distância se comparados a estímulos com ataques lentos e decaimentos rápidos. Os autores sugerem que a primeira categoria de estímulos foi ouvida como se composta de partes diretas e reverberantes, enquanto que a segunda categoria de estímulos sofreria menos os efeitos da reverberação. Num estudo mais recente, Wendt *et al.* (2016) testaram diferentes padrões de diretividade para observar a razão entre distância auditiva física e percebida. Seus resultados apresentaram correlação com a Razão entre Energia Direta e Energia Reverberante (DRR) para cada padrão de diretividade testado.

Inegavelmente, há uma relação de proporcionalidade inversa entre a sensação de *loudness* provocada e a distância da fonte sonora que pode ser degradada – ou mesmo terminada – em razão da quantidade de energia reverberante no ambiente de reprodução e/ou no material gravado, resultando no fenômeno de invariância de *loudness* com a distância, observado nos experimentos anteriormente descritos. Logo, é mister projetar um experimento perceptivo para se investigar até que ponto o algoritmo de medida pode tomar a invariância de *loudness* com a distância por garantida, ou se seus ganhos deverão ser escritos como funções da distância fonte-ouvinte.

- Pergunta: *Como a sensação de loudness é afetada pela variação das distâncias fonte-ouvinte?*

Para tanto, no escopo deste experimento, as hipóteses orientadas à invariância de *loudness* com a distância podem ser formuladas como abaixo.

- Hipótese nula: *Sons reproduzidos por alto-falantes situados a diferentes distâncias na mesma sala, ou situados à mesma distância em salas diferen-*

tes, que tenham o mesmo nível de loudness medido na posição do ouvinte, provocarão a mesma sensação de loudness.

- Hipótese alternativa: *Sons reproduzidos por alto-falantes situados a diferentes distâncias na mesma sala, ou situados à mesma distância em salas diferentes, que tenham o mesmo nível de loudness medido na posição do ouvinte, provocarão diferentes sensações de loudness.*

Note que as hipóteses formuladas apresentam as variáveis independentes a serem tratadas (distância e reverberação) objetivando a observação de seus efeitos na variável de resposta (juízo de *loudness* dos ouvintes). As próximas subseções descreverão as condições experimentais, o planejamento do experimento, e o que foi possível concluir a partir da análise dos dados coletados.

5.3.1 Verificações preliminares

Ao se pensar numa configuração experimental de tratamento da distância como única variável posicional, surgiu a preocupação de que uma disposição dos alto-falantes em linha e níveis intensos de reprodução praticados pudessem prejudicar a verificação dos efeitos observáveis. Para mitigar esse risco, alguns cuidados precisaram ser tomados: i) assegurar de que nenhuma excursão de níveis de sinal fosse intensa o suficiente para que o alto-falante distorcesse sua reprodução ou comprimisse o sinal como forma de proteção dos seus *drivers* e ii) garantir que os eixos acústicos de todos os alto-falantes tivessem visada direta para os ouvintes. Estas verificações foram feitas em concomitância com a elaboração do projeto de experimento a ser descrito na [subseção 5.3.2](#) e serviram para validar tanto os posicionamentos empregados quanto os níveis acústicos praticados.

Distorção harmônica dos alto-falantes

Anteriormente à calibração da instalação aos níveis de pressão sonora pretendidos, esbarrou-se numa insegurança quanto ao desempenho dos alto-falantes em níveis de reprodução intensos. Seria a Distorção Harmônica Total (THD) alta o suficiente para corromper um estímulo ou o sinal seria limitado

para proteger o *driver* e prevenir esse efeito? Para garantir que a distorção do sinal não seja um problema neste e nos próximos experimentos, foram feitas medidas de THD tanto a níveis desejados quanto extremos de pressão sonora.

A distorção harmônica total é definida como o nível RMS de todos os harmônicos numa faixa especificada, relacionada com o nível RMS da fundamental

$$\text{THD} = \frac{L_H}{L_F} \quad (5.1)$$

onde

$$L_H = \sqrt{L_2^2 + L_3^2 + \dots + L_n^2}, \quad (5.2)$$

L_F é o nível RMS da fundamental L_1 , e L_n é o nível RMS do n -ésimo harmônico. A THD possui valor nulo para um tom senoidal puro. A medida instrumental é feita pela filtragem da onda senoidal de um sinal capturado e posterior cálculo do valor RMS do resíduo. O medidor retorna então uma Distorção Harmônica Total mais Ruído (THD+N).

As medidas foram feitas numa sala em conformidade com a Recomendação BS.1116 do ITU-R (2015a) para avaliação subjetiva de pequenas diferenças de qualidade em sistemas de áudio. A relação de equipamentos utilizados é listada abaixo:

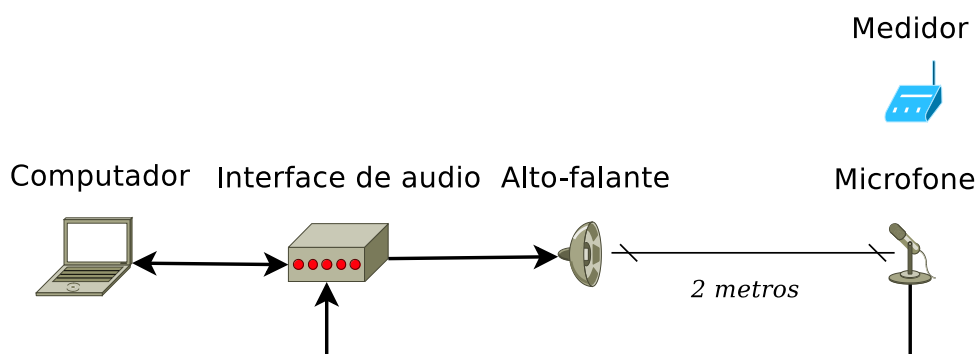
- Um alto-falante *Genelec 8020*;
- Um medidor portátil *NTI AL1 Acoustilyzer*;
- Um microfone de calibração *DPA 4600 XL2*;
- Uma interface de áudio *Focusrite Saffire 24 Pro*;
- Um notebook rodando MATLAB[®] e Audacity;
- Cabos XLR e pedestais.

Um diagrama da instalação de medida está ilustrada na [Figura 5.2](#). Tons senoidais de 997 Hz¹ gerados por computador – usados em conformidade com a Recomendação 17 da [Audio Engineering Society \(1998\)](#) – foram reproduzidos

¹ Um tom senoidal de 997 Hz foi utilizado em conformidade com a Recomendação AES17-1998 (AUDIO ENGINEERING SOCIETY, 1998). 997 é o número primo mais próximo de 1000 e assim mais valores de quantização são usados para representar digitalmente o tom de teste. Uma explicação mais detalhada pode ser encontrada em <<http://www.tonmeister.ca/wordpress/2017/03/03/997-hz/>>

pelo alto-falante sob teste. Medidas de THD+N e THD foram feitas simultaneamente pelo medidor portátil a partir do sinal capturado pelo microfone de calibração, respectivamente.

Figura 5.2 – Configuração de medidas de THD



Fonte: Elaborada pelo autor.

O fabricante do alto-falante especificou o valor de 102 dBSPL (sem ponderação) como máxima pressão sonora de curta duração a uma distância mínima de meio metro (GENELEC, 2018). Tomando este valor como fundo de escala digital, o sistema foi calibrado tal que o som reproduzido a dois metros de distância do microfone, a uma pressão sonora de 84 dBSPL², resultou num sinal de entrada de -18 dBFS, como ilustrado na Figura 5.3.

Algumas amostras de medidas, feitas numa escala variando entre níveis de pressão sonora operacionais e de saturação, estão dispostas na Tabela 5.5. Para uma sala de volume e tempo de reverberação similar à testada, o fabricante especifica uma pressão sonora máxima média sem ponderação de 92 dBSPL a uma distância de dois metros da fonte sonora (GENELEC, 2018), e por isso o valor foi o escolhido para ser o limite superior dos níveis testados.

Note que houve um incremento nas medidas de THD e THD+N para sinais acústicos superiores a 88 dBSPL. Um efeito dente-de-serra pôde ser ouvido nos primeiros segundos de reprodução, sugerindo que as magnitudes dos harmônicos

² Pressão sonora medida numa faixa de 10 Hz a 20 kHz, sem ponderação em frequências, com integração lenta (1 segundo) e nível sonoro contínuo equivalente (*Leq*) calculado ao longo de 20 segundos. Para fins práticos, estas configurações são replicadas a todas medidas de pressão sonora reportadas neste documento com a unidade “dBSPL”. Nos casos em que alguma curva de ponderação em frequências *P* tenha sido utilizada na medição de pressão sonora, esta será indicada na designação da medida (Ex.: *Leq(P)*), ou na própria unidade (Ex.: dBP), ou de forma expressa (Ex.: dBSPL ponderado pela curva *P*).

Figura 5.3 – Calibração de um tom gravado a 84 dB SPL para um sinal de entrada de -18 dBFS.



Fonte: Elaborada pelo autor.

Tabela 5.5 – Níveis de pressão sonora e medidas de distorção harmônica total.

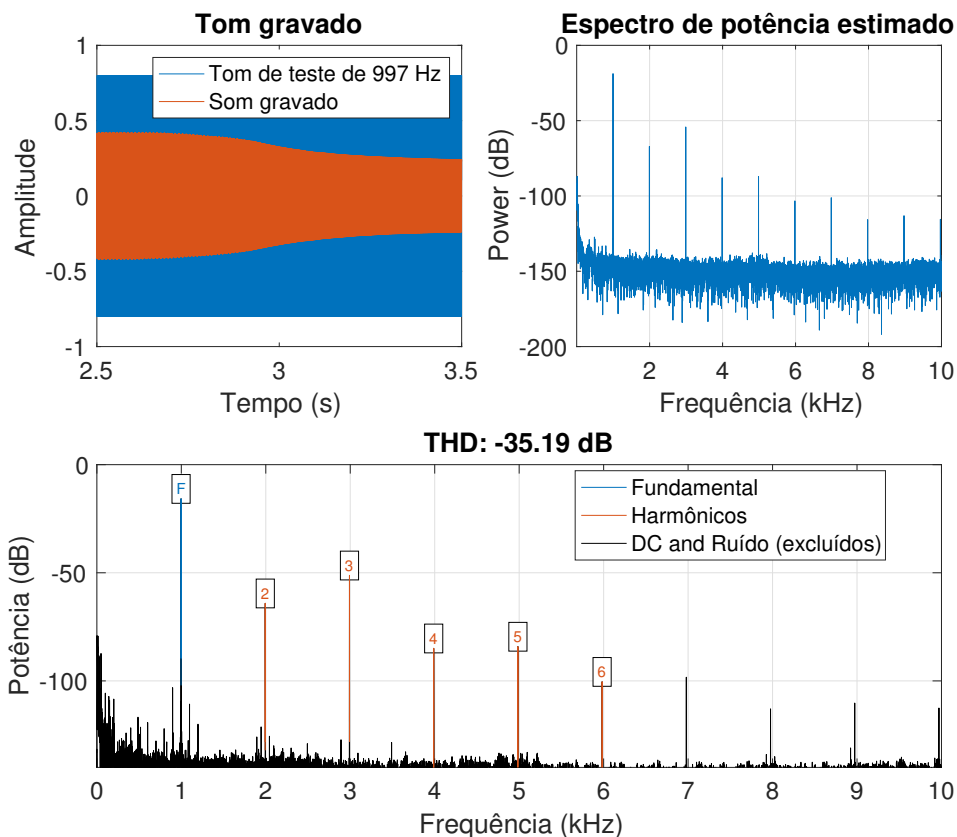
Leq (dB SPL _{slow})	Nível de sinal elétrico (dBu)	THD+N (dB)	THD (dB)
72,8	-53,0	-38,8	-56,93
77,0	-47,6	-38,9	-56,77
84,5	-43,0	-37,5	-55,61
88,0	-35,1	-32,6	-36,51
92.0 (88.0)	-35,3	-33,7	-35,19

ímpares não foram desprezíveis até que uma compressão interna protegeu o driver da saturação, como mostra a [Figura 5.4a](#).

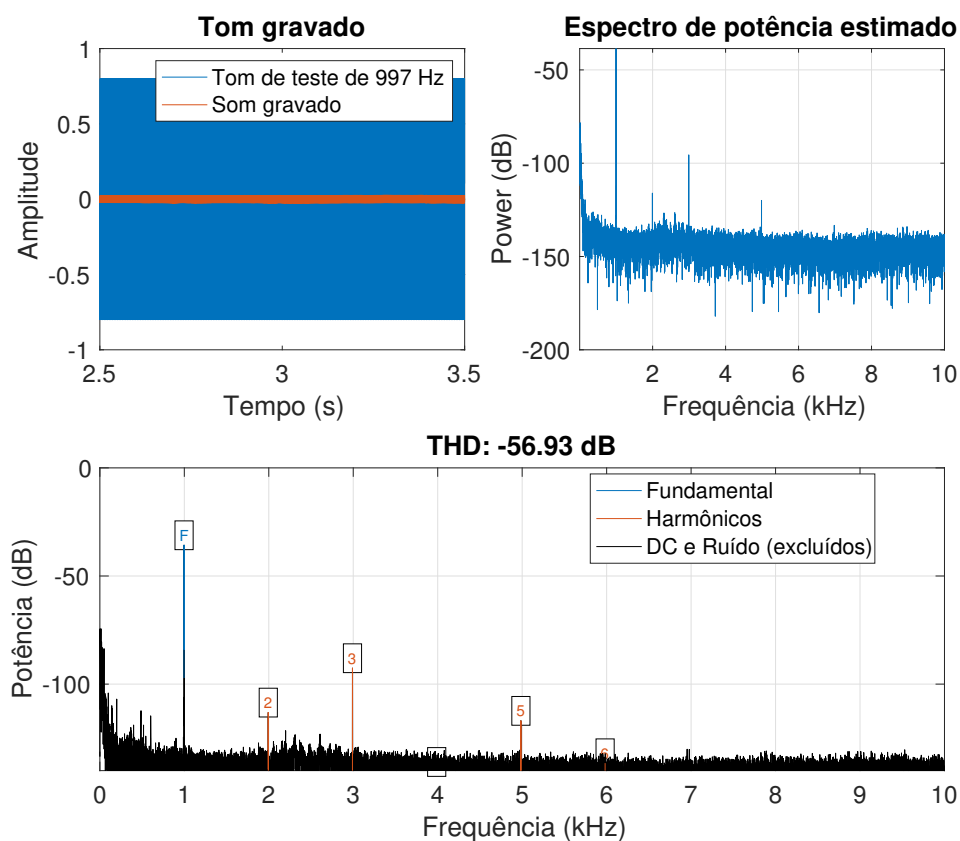
Medidas de THD feitas nesta região de saturação do alto-falante são maiores que os valores típicos de 0,001% (-50 dB) e os primeiros harmônicos ímpares são mais pronunciados. Por outro lado, para níveis de reprodução próximos de um nível de *loudness* de 8 sones / 70 phon, a distorção foi inferior a 0,001% e os primeiros harmônicos tiveram magnitudes imperceptíveis, como exemplificado na [Figura 5.4b](#).

Nos experimentos perceptivos propostos neste trabalho, julgou-se pertinente usar níveis de referência associados às sensações de *loudness* mapeadas aproximadamente pela curva de ponderação B (8 sones / 70 phon), mais próxima da curva RLB usada na radiodifusão, conforme visto na [subseção 3.3.2](#). Para tons de teste próximos a 1 kHz, com níveis de pressão sonora no entorno de 70 dB SPL, a THD observada foi insuficiente para representar problemas de

Figura 5.4 – Distorção harmônica total de sinais acústicos capturados a dois metros de distância da fonte sonora.



(a) Sinal acústico de 92 dB SPL comprimido a um nível 88 dB SPL



(b) Sinal acústico de 72 dB SPL

distorção de estímulos neste modelo específico de alto-falante. Os resultados obtidos sugerem que o monitor *Genelec 8020* é adequado para as tarefas que se seguirão.

Posicionamento e ângulos de espalhamento

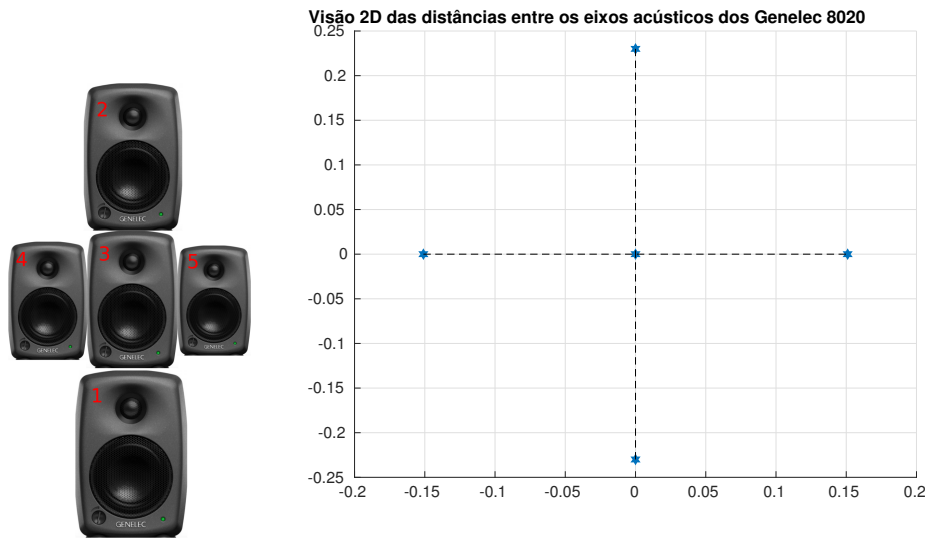
Para posicionar alto-falantes ao longo de uma linha reta e garantir que todos os eixos acústicos dos alto-falantes estejam visíveis para o ouvinte, é preciso que o espalhamento angular entre eles seja mínimo tal que as direções marginalmente diferentes entre si sejam encaradas como ruído experimental, sem efeito na variável de resposta do teste de escuta. Contudo, ater-se a diferenças no limite do observável de aproximadamente um grau de azimute e três graus de elevação, é preocupante do ponto de vista do planejamento. Um posicionamento prático considerando uma distância do alto-falante mais próximo inferior a 1 metro e todos os eixos acústicos desobstruídos, inevitavelmente resultará em ângulos de espalhamentos mais abertos do que os limites de discriminação angular. Torna-se então necessário avaliar como diferentes mudanças angulares são mapeadas nas diferenças entre as medidas de *loudness* ITU-R BS.1770 e o impacto destas diferenças no *setup* de calibração.

Montar um arranjo de alto-falantes em linha reta com mínimos deslocamentos em azimute e elevação, consiste basicamente em não deixar que os alto-falantes mais próximos obstruam as linhas de visada do ouvinte para os alto-falantes mais distantes. Representações bidimensionais desta ideia estão ilustradas na [Figura 5.5](#). À esquerda, alto-falantes mais próximos diferem em elevação (1 e 2), alto-falantes mais distantes diferem em azimute (4 e 5), e o alto-falante no meio é centrado a uma distância de referência. Mas para fins trigonométricos, a representação à direita será usada: alto-falantes serão tratados como fontes pontuais centradas nos seus eixos acústicos³.

Ter uma distância fonte-ouvinte inicial inferior a um metro tem por objetivo observar o efeito da difração de cabeça quando este é dominante. Por outro lado, uma distância inicial desta magnitude pode resultar num ângulo de elevação muito aberto. Conhecida a distância vertical entre dois eixos acústicos

³ O eixo acústico do alto-falante *Genelec 8020* está a uma altura de 146 mm, entre o cone de graves com 105 mm de diâmetro e o *tweeter* de 19 mm ([GENELEC, 2018](#)).

Figura 5.5 – Arranjo de alto-falantes na perspectiva do ouvinte: (a) ilustração crua da perspectiva do participante (b) representação 2D por pontos centrados nos eixos acústicos dos alto-falantes.



Fonte: Elaborada pelo autor.

d_V , o ângulo de espalhamento vertical ϕ_0 pode ser escrito como uma função da distância do primeiro alto-falante da forma:

$$\phi_0 = \arctan \left(\frac{d_V}{d_1} \right). \quad (5.3)$$

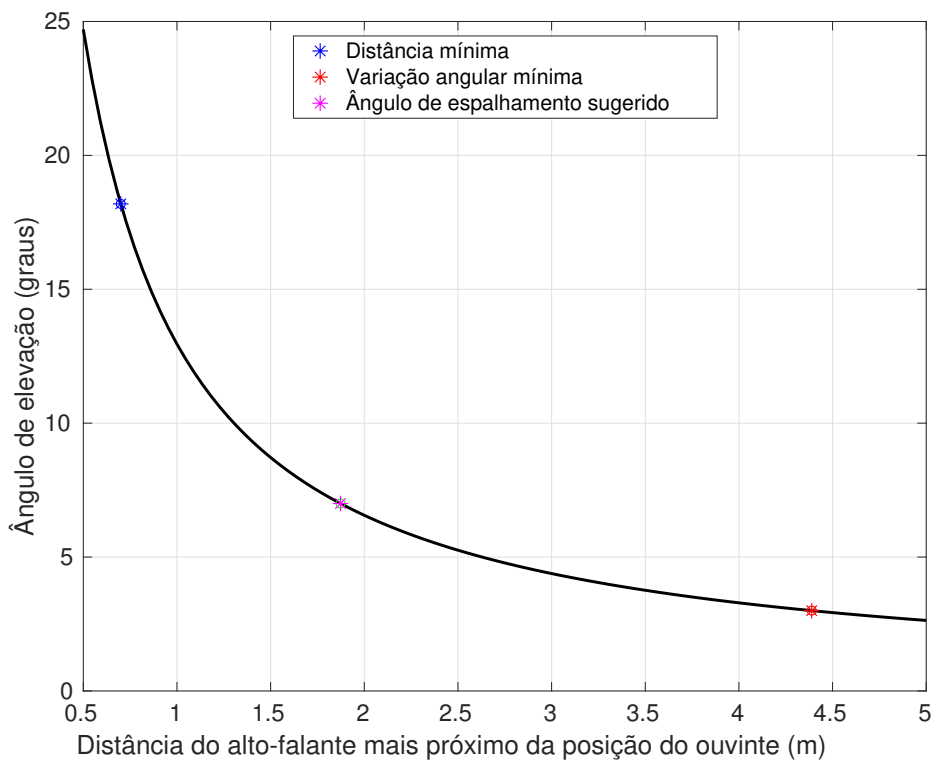
Da mesma maneira, conhecida a distância horizontal entre dois eixos acústicos d_H , o ângulo de espalhamento horizontal θ_0 pode ser escrito como uma função da distância do quarto alto-falante da forma:

$$\theta_0 = \arctan \left(\frac{d_H}{d_4} \right) \quad (5.4)$$

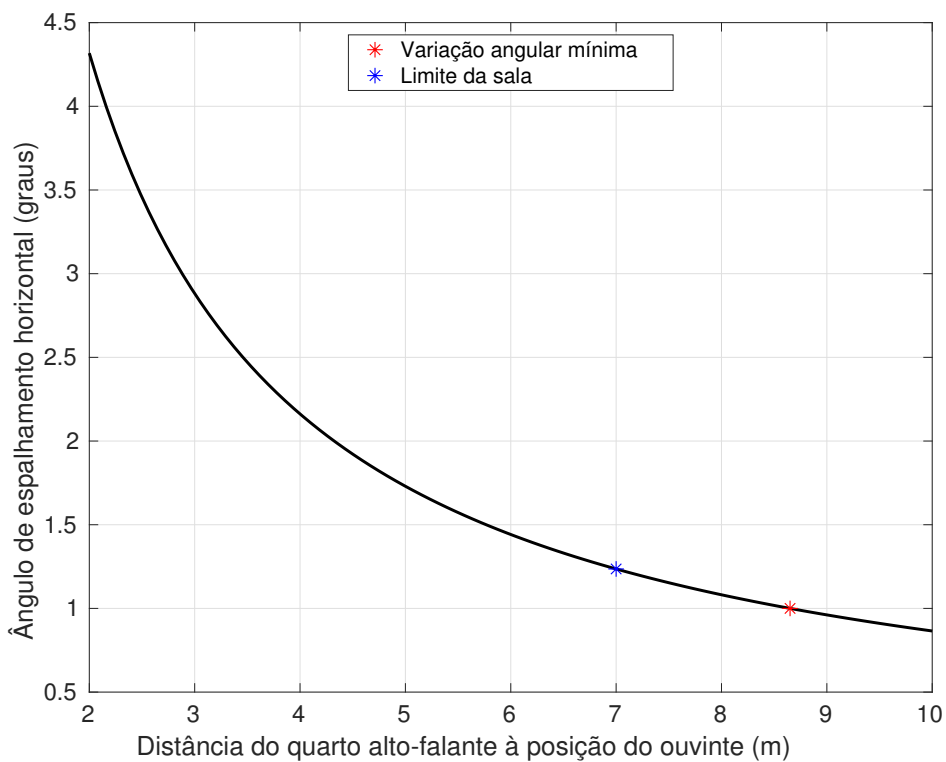
Plotar a [Equação 5.3](#) e a [Equação 5.4](#) para uma faixa de distâncias deu uma ideia melhor dos possíveis ângulos de espalhamento. Isto foi feito na ?? e na ??. No primeiro gráfico, é possível ver que i) a distância inicial deve ser de quase 4,5 metros para funcionar nos limites da elevação percebida, ii) uma distância mínima de 0,7 metro requereria um ângulo de espalhamento vertical de 18 graus e iii) um espalhamento de 7 graus – valor de trabalho imaginado a princípio – resultou numa distância próxima a 1,8 metro. No segundo gráfico, o quarto alto-falante deveria ser posicionado para além dos limites físicos da sala para garantir indiscriminação horizontal.

Embora um posicionamento prático muito provavelmente situar-se-á em ângulos de espalhamentos maiores que os de mínima discriminação, poderia ser

Figura 5.6 – Ângulos e posições para uma distância fixa entre eixos acústicos de dois alto-falantes Genelec 8020.



(a) Distância fixa vertical.



(b) Distância fixa horizontal.

Fonte: Elaborada pelo autor.

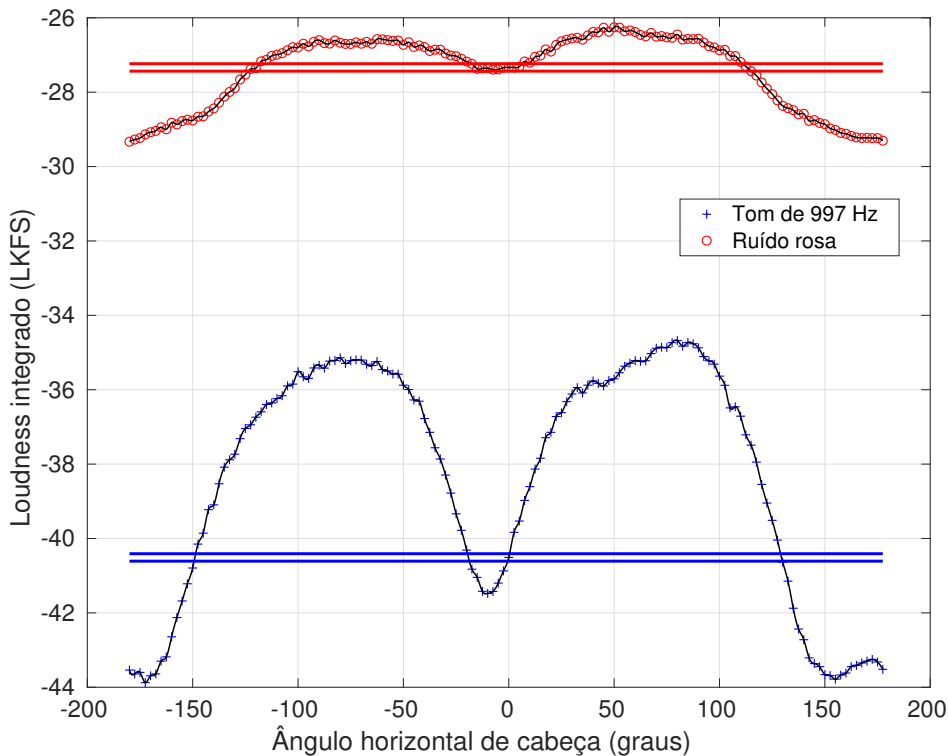
possível encontrar mais espaço para operar confiando não nas mudanças angulares mínimas perceptíveis, mas sim nas tolerâncias de um real medidor de *loudness*. O Relatório ITU-R BS.2217-2 estabelece uma tolerância de ± 0.1 LKFS para um medidor em conformidade com a Recomendação BS.1770 (ITU-R, 2016a). Optou-se então por melhor avaliar esta possibilidade antes de prosseguir com o experimento.

Uma maneira de se fazer verificações de incidência angular, é fazendo uso da Resposta Biauricular da Sala ao Impulso (BRIR) da sala do teste de escuta, medida para diferentes ângulos de cabeça e posições de alto-falantes por Francombe (2015). A ideia consiste em fazer a convolução de sinais de teste com as BRIRs, calcular o *loudness* BS.1770 médio dos sinais biauriculares e observar as diferenças entre medidas feitas a partir de diferentes ângulos de incidência. A Figura 5.7a e a Figura 5.7b ilustram essas medidas de *loudness* para um tom senoidal de 997 Hz e para o ruído rosa. As linhas coloridas correspondem aos valores de tolerância de ± 0.1 LKFS no entorno das medidas a um ângulo de cabeça de zero grau. As curvas de ajuste foram plotadas a partir de interpolações *spline* dos valores medidos.

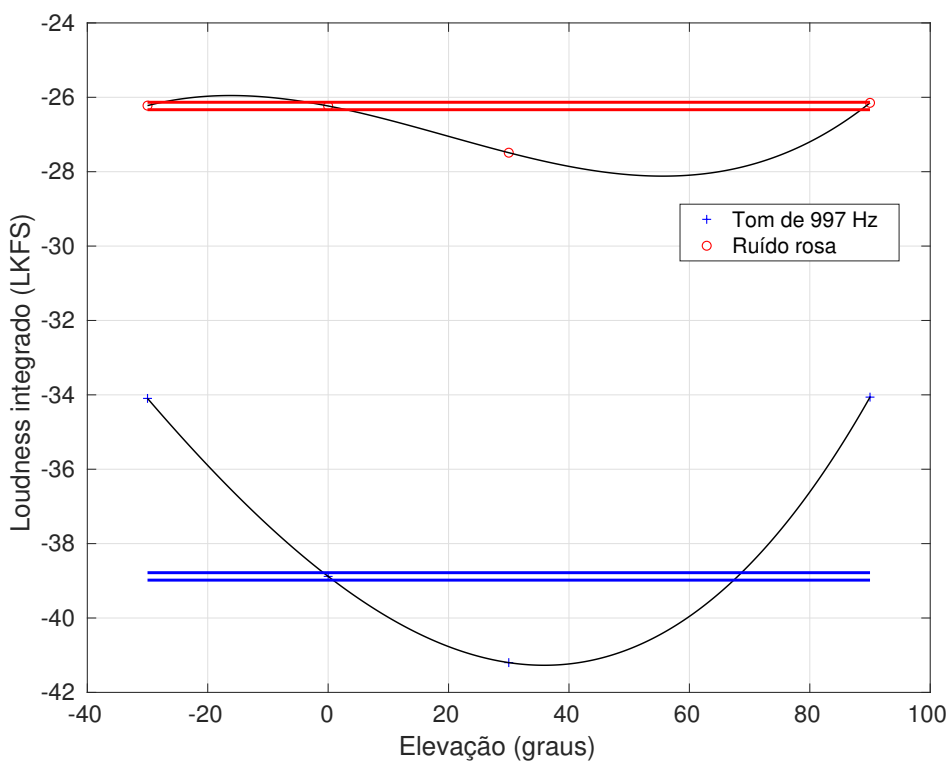
A Figura 5.7a refere-se a convoluções com BRIRs de incidências variando em azimute e a Figura 5.7b refere-se a convoluções com BRIRs de fontes variando em elevação. Em ambos os gráficos, é possível notar que as variações de *loudness* com as mudanças angulares são menores no ruído rosa do que no tom senoidal, pois as baixas frequências sofrem menos o efeito de sombreamento de cabeça e torso. Esta observação sugere que tons senoidais não devam ser usados no experimento principal, dado que seria mais difícil isolar os efeitos de distância, de desejável observação, dos efeitos causados por mudanças angulares, considerados como ruído experimental.

A Figura 5.8a e a Figura 5.8b trazem um olhar mais próximo das medidas de *loudness* perto dos ângulos zero de cabeça e elevação para o ruído rosa de teste. Na Figura 5.8a, há uma indicação clara de que as incidências frontais numa faixa angular de 20° resultaram em níveis de *loudness* dentro da faixa de tolerância do medidor, garantindo assim uma faixa de espalhamento horizontal mais confortável para se trabalhar. Por outro lado, não foi possível depreender um entendimento similar do gráfico na Figura 5.8b em função de um número

Figura 5.7 – Medidas de *loudness* BS.1770 feitas em sinais de teste (tom senoidal e ruído rosa) em convolução com as Respostas Biauriculares da Sala ao Impulso (BRIRs) da sala de escuta crítica.



(a) Para o alto-falante frontal e diferentes ângulos de cabeça.



(b) Para os alto-falantes de zero azimuth e ângulo de cabeça de zero grau.

menor de medidas: somente dos quatro alto-falantes de azimute zero (elevações de -30° , 0° , 30° e 90°). Ainda assim, a interpolação *spline* resultou num espalhamento vertical de aproximadamente seis graus, que dobraria a margem de trabalho na elevação.

Essas observações deram segurança de que seria possível manter um conjunto viável de distâncias fonte-ouvinte, alinhando os alto-falantes a pequenos deslocamentos verticais e horizontais para garantir que os eixos acústicos de todos os alto-falantes ficassem visíveis para os participantes, sem prejuízo para as comparações de suas respostas com os valores de *loudness* medidos pelo modelo ITU-R BS.1770 nas mesmas condições.

Teste da configuração experimental

O *setup* experimental ilustrado na [Figura 5.5](#) foi montado na sala de escuta crítica conforme com a Recomendação BS.1116. A lista definitiva de equipamentos é relacionada abaixo:

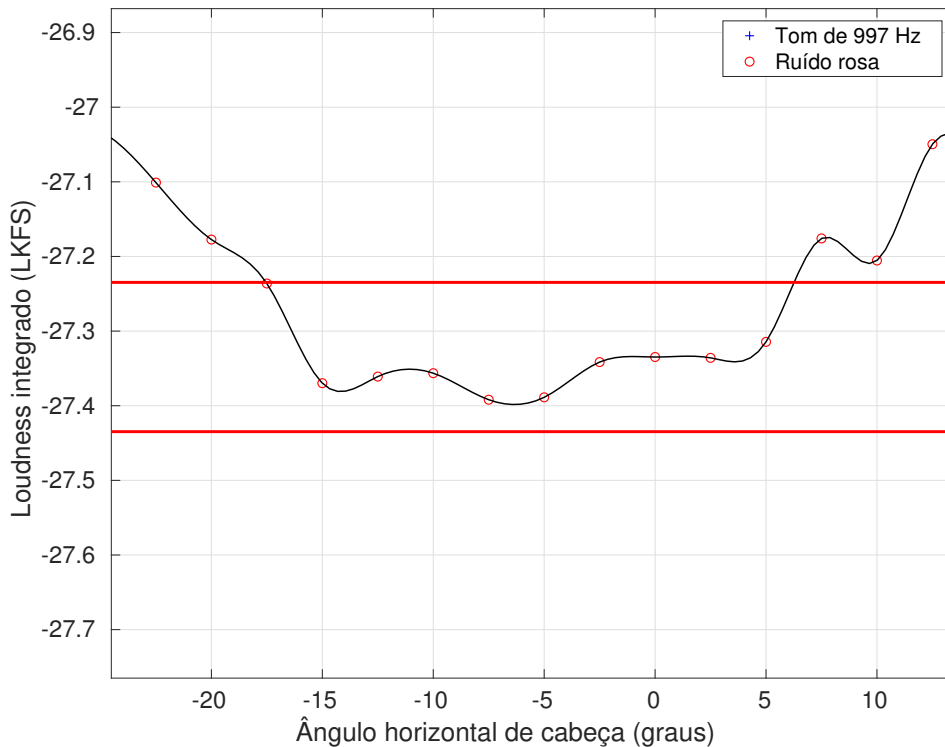
- Cinco alto-falantes *Genelec 8020*;
- Um Simulador de Cabeça e Torso (HATS) *Cortex MK2*;
- Um medidor SPL portátil *N05CC*;
- Um microfone de calibração *DPA 4006 X2*;
- Uma interface de áudio *RME Fireface 800*;
- Um notebook executando MATLAB[®], MaxMSP[®] e Audacity;
- Cabos XLR e pedestais.

Um esquema do *setup* de medida é ilustrado na [Figura 5.9](#)⁴. Um tom senoidal de 997 Hz a -18 LKFS foi reproduzido pelos alto-falantes *Genelec 8020*. Medidas de pressão sonora sem ponderação foram feitas pelo medidor portátil e pelo HATS. Cômputos de espectro e medidas BS.1770 de *loudness* momentâneo, curta duração e integrado, foram feitos por um script MATLAB[®] em duas rodadas: uma com o arranjo em linha, e outra com cada alto-falante livre de obstáculos em sua linha de visada.

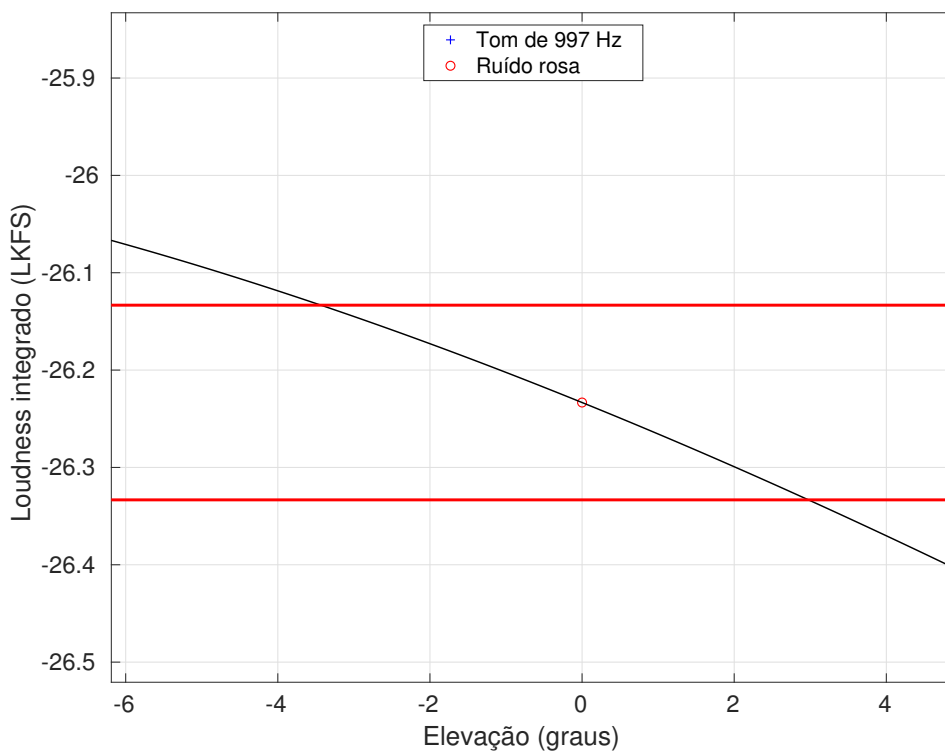
O sistema foi calibrado de tal forma que o tom senoidal a -18 LKFS capturado pelo HATS a uma pressão sonora média de ≈ 70 dB SPL sem pondera-

⁴ O posicionamento dos alto-falantes e a escolha das distâncias serão abordados na [subseção 5.3.2](#)

Figura 5.8 – Gráficos ampliados no zero de variação angular para o ruído rosa: medidas horizontais próximas a um ângulo de cabeça de 0°.



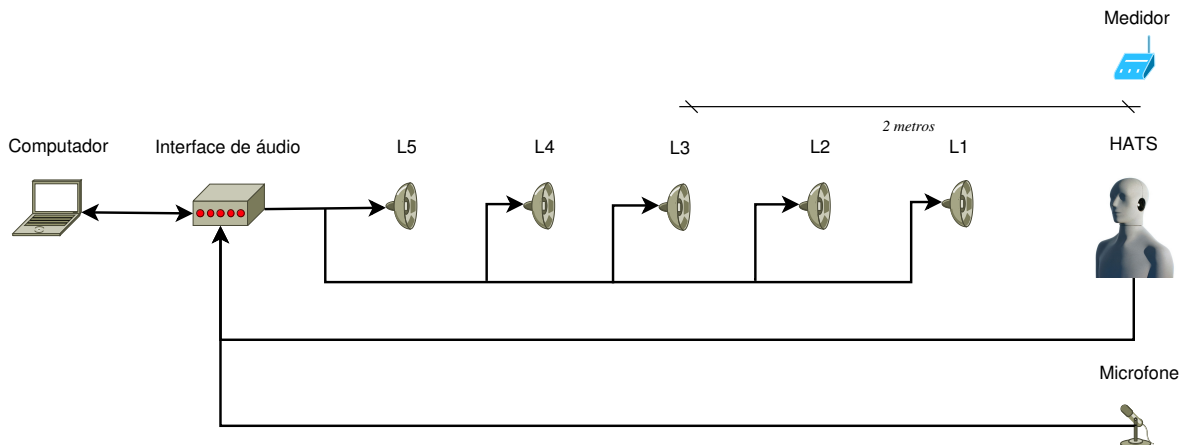
(a) Ampliação da Figura 5.7a



(b) Ampliação da Figura 5.7b

Fonte: Elaborada pelo autor.

Figura 5.9 – Setup de medidas de loudness do arranjo de alto-falantes em linha.



Fonte: Elaborada pelo autor.

ção incidente nos dois ouvidos do manequim, resulte num nível de sinal digital de entrada de -18 dBFS, e um *loudness* de -18 LKFS medido em tempo de execução implementado no MATLAB[®] (ver Figura 5.10)

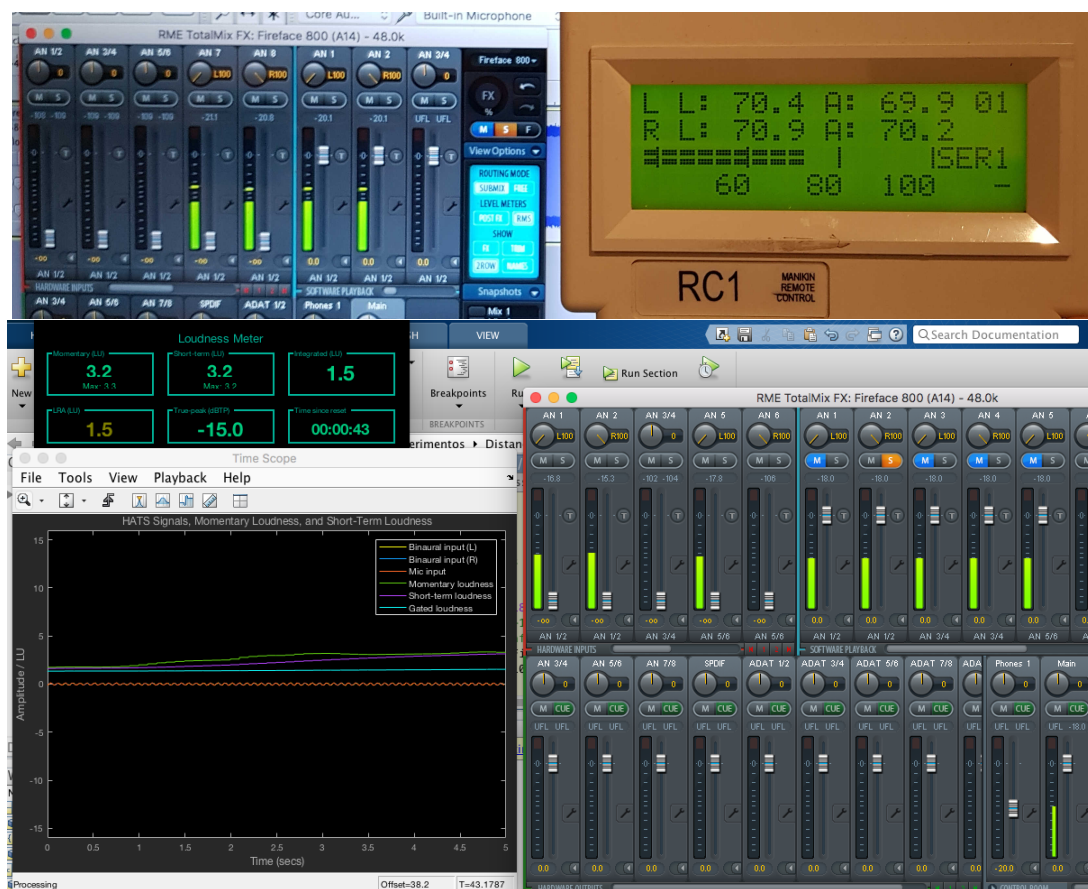
Foram feitas medidas amostrais de níveis de pressão sonora para cada alto-falante no arranjo em linha com mínimos ângulos de espalhamento, e fora do arranjo a cada posição desobstruída. Valores para o tom de teste estão listados na Tabela 5.6 considerando o mesmo sinal senoidal como referência de calibração.

Tabela 5.6 – Medidas de nível de pressão sonora do tom de teste de 997 Hz (dB SPL)

	Setup 1: Reprodução em linha					Setup 2: Reprodução desobstruída				
	0,8 m	1,3 m	2,0 m	3,0 m	4,5 m	0,8 m	1,3 m	2,0 m	3,0 m	4,5 m
HATS ouvido esquerdo	69,4	69,2	71,3	69,1	69,6	71,1	69,8	70,8	68,8	68,7
HATS ouvido direito	69,6	70,2	72,3	70,3	68,5	70,9	69,7	70,6	68,1	68,1
Medidor SPL ($Leq_{(20\ s)}$)	77,6	76,5	76,4	66,9	65,3	67,8	64,5	78,8	63,4	63,4

As diferenças entre níveis de pressão sonora incidente no HATS e no medidor SPL se devem ao efeito de difração do torso e da cabeça do manequim, que não acontece na sonda omnidirecional do medidor. Note que os valores para o terceiro alto-falante a uma distância de referência de dois metros são aproximadamente os mesmos nos cenários “em linha” e “desobstruído”, porque não há espalhamento em nenhum dos cenários e somente a atenuação esperada do próprio manequim pôde ser observada. Por outro lado, diferenças de SPL entre o HATS e o medidor omnidirecional para os dois alto-falantes mais próximos foram positivas no Setup 1 e negativas no Setup 2 devido a diferenças na elevação. Por outro lado, as mesmas diferenças de SPL entre os cenários “em linha” e

Figura 5.10 – Calibração do arranjo experimental



Fonte: Elaborada pelo autor.

Nota – Um sinal senoidal digital de saída num nível de -18 dBFS, reproduzido a uma distância de dois metros do HATS, incidindo com $Leq = 70$ dB SPL nos microfones intra-auriculares do manequim, resultou num sinal senoidal digital de entrada num nível de -18 dBFS, com um nível de *loudness* de -18 LKFS.

“desobstruído” para os dois alto-falantes mais distantes são menores, o que sugere que as diferenças de azimute tiveram menos influência do que as diferenças de elevação quando da calibração dos *setups*.

Se por um lado, as diferenças entre as medidas de SPL omnidirecionais e via manequim foram importantes para elucidar ao efeito do espalhamento angular na incidência sonora, o importante nesta verificação foi que a calibração do sistema de reprodução em linha foi feita sem prejuízos em relação ao cenário ideal no qual todos os alto-falantes estivessem desobstruídos. Portanto, este arranjo em linha, com pequenos *tilts* de azimute e elevação para que os eixos acústicos de todos os alto-falantes estejam na linha de visada do ouvinte, é considerado adequado para se testar as hipóteses formuladas no início desta

seção.

5.3.2 Projeto de experimento

As verificações conduzidas na subseção anterior foram motivadas por preocupações surgidas durante a etapa de planejamento desta investigação sobre os efeitos da distância da fonte sonora nas sensações de *loudness* provocadas nos ouvintes. Esta subseção descreve as condições iniciais do ambiente de experimentação, a metodologia experimental utilizada, além de projeto e análise estatísticos.

Participantes

O experimento foi feito com alunos do curso de graduação *Tonmeister* da Universidade de Surrey, não só acostumados com testes perceptivos sendo muitos deles participantes de painéis de escuta, isto é, pertencentes a grupos semipermanentes treinados para participarem em testes de escuta regularmente (BECH; ZACHAROV, 2007). Dispor de ouvintes treinados tem a vantagem de resultar numa variabilidade dos dados potencialmente reduzida, o que sugere que um critério de potência comum pode ser atingido com um número de participantes logisticamente viável.

Estímulos

Estímulos sintéticos podem dar-se na forma de sinais de ruído rosa entrecortados (*gated*) em intervalos de subida e descida de 200 ms cada, e limitados em faixa entre 200 Hz e 15 kHz. Devido às suas características de mesma energia por oitava que os fazem igualmente audíveis em todo espectro, espera-se que ajudem a revelar pequenas diferenças entre as sensações de *loudness* provocadas. As características espectrais do ruído rosa o torna adequado para testes de escuta que tratem somente as características espaciais das fontes sonoras, como é o caso deste. Já os entrecortes abrem espaço para que a reverberação seja ouvida pelos participantes.

Alguns tipos de conteúdo de especial interesse para os radiodifusores foram considerados. Uma gravação de sons ambientais, um trecho musical

utilizado para testes de identificação de fonte sonora disponível no *handbook* de usuário nº 3253 da EBU (2008) e uma coleção de frases foneticamente balanceadas faladas em língua portuguesa gravadas por Follador, Silva e Yehia (2017) – juntamente com o ruído rosa – compuseram o conjunto de itens de programação. Os sinais possuem uma taxa comum de 48.000 amostras por segundo e foram normalizados em -23 LKFS, tal como exigido pela norma brasileira de *loudness* para a radiodifusão digital (MC, 2012).

Salas de reprodução

Para investigar a influência da reverberação na relação do *loudness* com a distância, o experimento foi conduzido em duas localizações diferentes, sendo a primeira uma sala em conformidade com as especificações da Recomendação BS.1116 do ITU-R (2015a) quanto à avaliação de pequenas diferenças em sistemas de áudio (doravante "Sala 1"), e a segunda uma sala de aula comum (doravante "Sala 2"). As avaliações dos tempos de reverberação das salas serão detalhadas no contexto da relação entre *loudness* e reverberação, objeto de investigação da seção 5.4.

Calibração

A calibração nas salas de reprodução foi realizada com um Simulador de Torso e Cabeça (HATS), manequim com microfones intra-auriculares embutidos projetado para reproduzir as propriedades acústicas da cabeça e do torso de um adulto médio. O manequim é colocado na posição futuramente ocupada pelos participantes do teste.

O volume de cada alto-falante é calibrado tal que um sinal de ruído rosa com um nível de *loudness* integrado de -23 LKFS, quando reproduzido, resulte num nível de pressão sonora de 70 dB SPL incidente nos microfones intra-auriculares do HATS, e num nível de *loudness* de -23 LKFS de sinal biauricular capturado. O procedimento foi testado durante as verificações feitas na subseção 5.3.1.

Apresentação dos estímulos

A variável de resposta dá-se na forma de marcações dos participantes numa tarefa de casamento de *loudness*. Este foi feito pelo método de estimação de Limiares de Diferença (DL) denominado “Método de Ajuste”, no qual os participantes são solicitados a ajustar o *loudness* de cada estímulo reproduzido por um dado alto-falante, conforme um nível de pressão sonora de referência. Os sons de teste e de referência são de mesma categoria (ruído rosa, fala, música ou som ambiental) e podem ser alternados a qualquer tempo durante a tarefa de ajuste.

Distâncias avaliadas

Diferenças no limite do observável entre distâncias de fontes sonoras variam entre 5% e 25% da distância de referência (KOLARIK *et al.*, 2016). Tomando-se incrementos/decrementos geométricos a partir de uma distância de referência de 2 metros da fonte sonora (ITU-R, 2015a), um quociente $q = 1,5$ resultou num vetor de distâncias fonte-ouvinte $D = [0,88 \ 1,33 \ 2,00 \ 3,00 \ 4,50]$ m. Além de garantir uma boa distribuição em linha dos alto-falantes nas salas de reprodução, com distâncias inferiores a dois metros seria possível observar o efeito da incidência de frentes de onda não planas na cabeça do ouvinte (BLAUERT, 1997).

Equipamentos

Os equipamentos separados para uso nesta série de testes estão abaixo relacionados:

- Cinco alto-falantes *Genelec 8020* para reprodução de estímulos;
- Um Simulador de Cabeça e Torso (HATS) *Cortex Mk2* para calibração de sinais de referência;
- Uma interface de áudio *RME Fireface 800*;
- Um controlador USB *Griffin Powermate* para ajustes de nível sonoro;
- Um notebook executando MaxMSP[®], MATLAB[®] e *Audacity*;
- Um segundo notebook posicionado fora do ambiente de teste para operar remotamente o primeiro;

- Cabos e pedestais.

Avaliação de riscos

Os riscos identificados como significativos em experimentos desta natureza são referentes a ruído, segurança elétrica e manuseio.

O nível de reprodução dos estímulos foi pensado considerando os limites da norma 1999 da [ISO \(2013\)](#): 140 dBC de pico e 85/80 dBA de exposição ocupacional. Como medida de controle, as sessões dos participantes serão limitadas e a exposição sonora dar-se-á num nível confortável de 70 dBSPL na posição do ouvinte.

Com relação aos demais riscos identificados, todos os equipamentos alimentados foram checados quanto aos selos de aprovação em testes de segurança elétrica. Guias de saúde e segurança no trabalho referente a manuseio de cargas foram lidos antes do transporte e manipulação de alto-falantes, mesas, cadeiras, cabos, tripés e *cases*. O relatório de avaliação de riscos correspondente pode ser encontrado no [Apêndice B](#).

Teste de hipóteses

Durante o casamento de *loudness*, no qual os participantes ajustam o nível sonoro de um item de programação reproduzido à distância de teste “A”, baseados em um nível de *loudness* de referência reproduzido à distância “B”, a variabilidade devido aos diferentes itens de programação pode ser considerada como uma fonte de variação espúria na avaliação das distâncias. Uma solução possível para eliminar essa inconveniência seria o pareamento dos ajustes por item de programação, no qual as observações são examinadas em pares (A,B) para cada ocorrência e o teste de hipóteses é feito com as amostras das diferenças ([BECH; ZACHAROV, 2007](#)).

Sejam $Y_{A,j}$ e $Y_{B,j}$ pares de observações de ajustes de *loudness* às distâncias A and B, para cada item de programação j . Então as diferenças pareadas das observações são $D_j = Y_{A,j} - Y_{B,j}$.

Se modelarmos nossas observações como um processo aditivo:

$$Y_{i,j} = \underbrace{\mu + \tau_i}_{\mu_i} + \beta_j + \varepsilon_{i,j}, \quad (5.5)$$

onde μ é o valor da média geral, τ_i é o efeito de tratamento da i -ésima distância na média, β_j é o efeito do j -ésimo item de programação, e $\varepsilon_{i,j}$ é o resíduo do modelo, então:

$$D_j = (\mu_A + \beta_j - \mu_B - \beta_j) + \tau_A - \tau_B + \varepsilon_{Aj} - \varepsilon_{Bj} \quad (5.6)$$

$$= \mu_D + \varepsilon_j \quad (5.7)$$

Nossas hipóteses de interesse, textualmente descritas na abertura desta seção, podem ser definidas em termos das médias das diferenças μ_D :

$$\begin{cases} H_0 : \mu_D = 0 \\ H_1 : \mu_D \neq 0 \end{cases} \quad (5.8)$$

que agora podem ser tratadas como um teste de hipóteses de única amostra: a das diferenças nos ajustes de *loudness* para as distâncias fonte-ouvinte sob investigação.

A estatística de teste para estes casos é dada por:

$$T_0 = \frac{\bar{D}}{\sqrt{S_D^2 \times \frac{1}{n}}}, \quad (5.9)$$

distribuída sob a hipótese nula como uma variável t de Student com $n - 1$ graus de liberdade, onde n é o número de diferenças observadas no experimento, e \bar{D} e S_D^2 são a média e a variância estimada das diferenças pareadas, respectivamente (BECH; ZACHAROV, 2007, Eq (6.21)).

Todavia, ao considerar que as variáveis independentes do experimento são um número de pares de distâncias N_d , um número de programas N_p e um número de salas N_s , esta abordagem pareada para o projeto de experimento levaria a um número de $N_d \times N_p \times N_s$ testes t a serem feitos. O melhor seria ter um único teste estatístico com todas as médias para verificar se a hipótese nula será rejeitada em pelo menos uma delas, evitando testes t pareados desnecessários.

Sejam $Y_{i,j,s}$ as observações de ajuste de loudness para cada i -ésimo par de distâncias (A,B), j -ésimo item de programação e s -ésimo participante. Nosso processo aditivo de Análise de Variâncias (ANOVA) seria da forma:

$$Y_{i,j,s} = \underbrace{\mu + \tau_i}_{\mu_i} + \beta_j + \alpha_s + \varepsilon_{i,j,s} \quad (5.10)$$

onde μ é a média geral, τ_i é o efeito de tratamento do i -ésimo par de distâncias na média, β_j é o efeito do j -ésimo item de programação, α_s é o efeito do s -ésimo participante, e $\varepsilon_{i,j,s}$ é o resíduo do modelo.

A introdução da variável “participante” tem por objetivo reduzir a soma dos quadrados dos erros e, conseqüentemente, aumentar a resolução dos F -scores da ANOVA (BECH; ZACHAROV, 2007). Dado que os ouvintes neste experimento não são de forma nenhuma inexperientes, o efeito dos participantes é introduzido no modelo como uma variável fixa, de modo a simplificar a análise.

Um modelo ANOVA que leve em conta as interações entre os efeitos observáveis, pode ser formulado como:

$$Y_{i,j,s} = \underbrace{\mu + \tau_i}_{\mu_i} + \beta_j + \gamma_{i,j} + \alpha_s + \delta_{i,s} + \eta_{j,s} + \varepsilon_{i,j,s} \quad (5.11)$$

onde μ é a média geral, τ_i é o efeito de tratamento do i -ésimo par de distâncias na média, β_j é o efeito do j -ésimo item de programação, $\gamma_{i,j}$ é o efeito de interação entre distâncias e programas, α_s é o efeito do s -ésimo participante, $\delta_{i,s}$ é o efeito de interação entre distâncias e participantes, $\eta_{j,s}$ é o efeito de interação entre programas e participantes, e $\varepsilon_{i,j,s}$ é o resíduo do modelo.

O processo acima pode ser simplificado ao testar-se uma interação entre efeitos por vez e observar possíveis incrementos no erro médio quadrático. Este procedimento foi desenvolvido para ser executado após o experimento para avaliar se a variância removida de cada interação compensa a perda de graus de liberdade – resolução do teste – causada por sua introdução.

Ao se pensar na ANOVA como uma “relação sinal-ruído” – ou uma relação efeito-aleatoriedade de variâncias – com as suas variáveis mapeadas e levadas em consideração, pôde-se finalmente escrever as hipóteses das diferenças das

médias causadas pelos efeitos de tratamento da forma a seguir:

$$\begin{cases} H_0 : \tau_i = 0, \forall i \in \{1, 2, \dots, N_d\} \\ H_1 : \exists \tau_i \neq 0 \end{cases} \quad (5.12)$$

onde τ_i é o efeito do tratamento dos N_d pares de distâncias fonte-ouvinte avaliadas.

Estimação de tamanho de amostra

Num cenário no qual a hipótese nula seja rejeitada por um ou mais efeitos, o caminho a se seguir seria fazer comparações par a par das observações com testes t pareados. Para tanto, é desejável ter alguma ideia antecipada do número de observações necessário, considerando o tamanho de efeito que se quer observar. Assume-se aqui:

- Grau de significância $\alpha = 0,05$;
- Potência estatística $(1 - \beta) = 0,8$;
- Mínimo tamanho de efeito de interesse $\delta^* = 0,5 \text{ dB}$ ⁵;
- Variância estimada das observações $S_D = 1,0 \text{ dB}$ ⁶.

Se a variância da distribuição do projeto de experimento pareado não é conhecida, a distribuição t de Student pode ser substituída pela distribuição normal $N(0,1)$ e os valores das estatísticas de testes podem ser calculadas diretamente a partir dos valores estabelecidos de α e β (BECH; ZACHAROV, 2007, Eqs. (6.10) e (6.11)).

Reescrevendo a estatística de teste da [Equação 5.9](#) para a hipótese nula H_0 :

$$N(0,1)_{1-0.05} = \frac{D}{\sqrt{1.0 \times \frac{1}{n}}} = 1,65; \quad (5.13)$$

⁵ O valor representa um meio termo entre o passo mínimo na escala de magnitude para o método de ajuste empregado (0,1 dB) e o ponto de igualdade subjetiva de 1 dB empregado em métodos de escolha forçada. Mais sobre isso na [seção 5.5](#).

⁶ Esta é uma extrapolação crua feita a partir dos gráficos em barra de casamentos de *loudness* no trabalho de [Francombe et al. \(2015a\)](#).

e para a hipótese alternativa H_1 :

$$N(0, 1)_{0.20} = \frac{D - 0.5}{\sqrt{1.0 \times \frac{1}{n}}} = -0,84 \quad (5.14)$$

Resolvendo as duas equações anteriores eliminando a variável D , obtem-se o número de observações pareadas para este teste:

$$\frac{D - 0.5}{\sqrt{\frac{1}{n}}} = 1.6449 - (-0.84162) \quad (5.15)$$

que resulta num número de observações pareadas necessário para garantir os níveis de α e β estipulados de $n \approx 25$. Este valor não é interpretado como absoluto, mas sim como um norte para o recrutamento e o planejamento neste e nos demais experimentos daqui por diante, de forma a se buscar um número de participantes \times repetições \times sessões considerado confortável para análises “post-hoc”.

Análise estatística

Subsequentemente à análise de variâncias, uma série de comparações múltiplas será feita na forma de testes t para as diferenças das médias estimadas, com a preocupação de se evitar que a probabilidade de erro do Tipo-I cresça substancialmente no processo, ao se aplicar alguma correção dos valores de α usando o teste Diferença Honestamente Significativa (HSD) de Tukey (FIELD, 2013). Os objetivos da análise se encontram abaixo relacionados:

1. Inspeção exploratória por meio de funções de diagrama de caixa (*boxplot*) para visualizar se houve detecção de grandes variações para certos níveis de estímulos ou de distâncias fonte-ouvinte.
2. Ajustar os dados a um modelo linear generalizado (ANOVA) e observar se a hipótese nula é rejeitada a um intervalo de confiança de 95% em algumas das distâncias fonte-ouvinte testadas, para uma dada categoria de estímulos.

3. Fazer comparações par-a-par do tipo “todos contra todos” nas distâncias fonte-ouvinte de interesse, e procurar por influências observáveis de algum dos efeitos.
4. Avaliar a força destas conclusões por meio de pequenas probabilidades de se obter os mesmos resultados na ausência do efeito testado (p -valores) e por meio de medidas de magnitude dos efeitos observados, quando presentes.

5.3.3 Experimento principal

O teste auditivo de casamento de *loudness* foi então conduzido para quantificar o ajuste do nível de pressão sonora necessário de modo a obter-se uma sensação de *loudness* comum entre sinais reproduzidos a diferentes distâncias do ouvinte. Para tanto, o sistema eletroacústico de cinco canais da [Figura 5.9](#) foi utilizado para reproduzir segmentos sonoros de fala, música, ruído sintético e sons ambientais nas salas 1 e 2, sendo a primeira uma sala em conformidade com as especificações da Recomendação BS.1116 do [ITU-R \(2015a\)](#) quanto à avaliação de pequenas diferenças em sistemas de áudio, e a segunda uma sala de aula comum (ver [subseção 5.3.2](#)).

Coleta de dados

O experimento contou com um grupo de sete ouvintes treinados, composto por alunos da graduação *Tonmeister* do Instituto de Gravação Sonora da Universidade de Surrey no Reino Unido, realizado durante estágio doutoral do autor. Os colaboradores levaram cerca de 50 minutos para completar as tarefas, e entre 20 e 30 segundos por casamento. Cada ouvinte executou esta tarefa em duas sessões – uma em cada sala de reprodução – para os quatro itens de programação (“ruído”, “pop”, “fala” e “ambiental”), e os cinco alto-falantes posicionados às distâncias fonte-ouvinte $d \in D = [0,88 \ 1,33 \ 2,00 \ 3,00 \ 4,50]$ m numa apresentação aleatorizada do tipo “todos contra todos”, resultando em 100 ajustes por participante e 1400 ajustes no total.

Como consideração ética, teve-se como boa prática garantir que os voluntários atestassem que consentiam com a participação no experimento. A

assinatura de um formulário de consentimento foi precedida de uma folha de informações na qual constava a natureza do experimento, omitindo menções aos efeitos de observação desejável de modo a evitar quaisquer vieses na execução. A folha de informações e o formulário de consentimento foram versões atualizadas dos exemplos apresentados por [Bech e Zacharov \(2007, Fig. 9.2 e 9.3\)](#). Estes documentos estão disponíveis para consulta no [Apêndice B](#).

O computador de uso dos participantes foi conectado a uma interface de áudio digital *RME Fireface 800*, que encaminhava o sinal de áudio aos cinco canais para reprodução individual. A interface de usuário na [Figura 5.11](#), desenvolvida em *Max/MSP[®]*, continha tão somente as instruções do teste e um contador de tarefas executadas. Os níveis iniciais das reproduções de teste eram aleatorizados entre ± 10 dB sem qualquer indicação em tela. Os ajustes foram feitos usando um potenciômetro ilimitado e não rotulado (*Griffin Powermate*), de tal modo que os participantes não introduzissem quaisquer vieses decorrentes de noções intuitivas de escala. O programa também controlava as reproduções e os níveis dos sinais, além de registrar os tempos de cada tarefa, os *offsets* aleatórios iniciais e os ajustes dos participantes para cada par teste-referência de alto-falantes na tarefa.

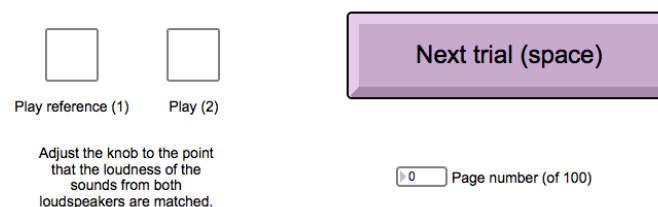
Figura 5.11 – Interface gráfica do experimento.

When you click on "Play reference", or press "1" on the keyboard, you will hear an audio sample being reproduced by one of the loudspeakers ahead of you. A second click (or press) will stop sound reproduction.

When you click on "Play", or press "2" on the keyboard, you will hear the same audio sample being reproduced by a different loudspeaker. The sound volume can be adjusted using the silver knob on your left.

Please tweak the knob to the point that the loudness of the sounds from both loudspeakers are matched.

There will be 100 pages in total.



Nota – Contém somente as instruções do teste e um contador de tarefas executadas, sem *sliders*, *faders*, VUs ou qualquer outro controle/monitor rotulado que pudesse introduzir vieses decorrentes de noções intuitivas de escala.

Fonte: Elaborada pelo autor.

Diferenças médias entre os níveis dos ajustes e os níveis de referência,

com intervalos de confiança de 95%, entre todos os participantes, para cada par teste/referência são ilustradas na [Figura 5.12a](#) e na [Figura 5.12b](#), um gráfico para cada sala. As barras coloridas indicam os ajustes feitos pelos participantes nos níveis dos alto-falantes de teste nas abscissas. Cumpre notar que i) os ajustes nos (ou referentes aos) níveis do alto-falante mais próximo do ouvinte têm magnitudes destacadas se comparados com os ajustes de nível dos demais alto-falantes, e ii) nos casos em que os alto-falantes de teste e de referência são os mesmos, todas as médias e intervalos de confiança são inferiores a 1 dB, o que sugere diligência dos participantes no cumprimento das tarefas do experimento.

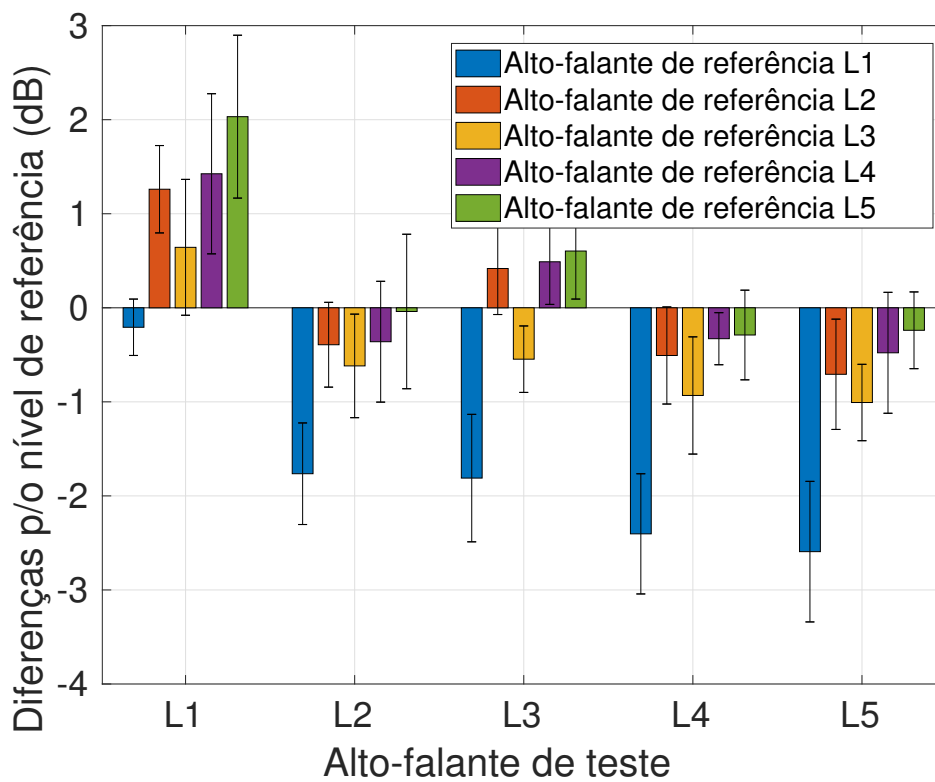
Médias e intervalos de confiança das diferenças dos ajustes dos participantes em relação ao nível de *loudness* nominal, agrupados por alto-falante de referência e quebrados por item de programação são ilustrados na [Figura 5.13](#). Por inspeção visual, não é possível identificar tendência alguma no comportamento dos ajustes para cada estímulo. A ausência de um padrão observável no comportamento dos ajustes de nível sonoro sugere que o efeito da variável “item de programação” nas respostas dos participantes não tenha sido observado neste experimento.

Análise exploratória

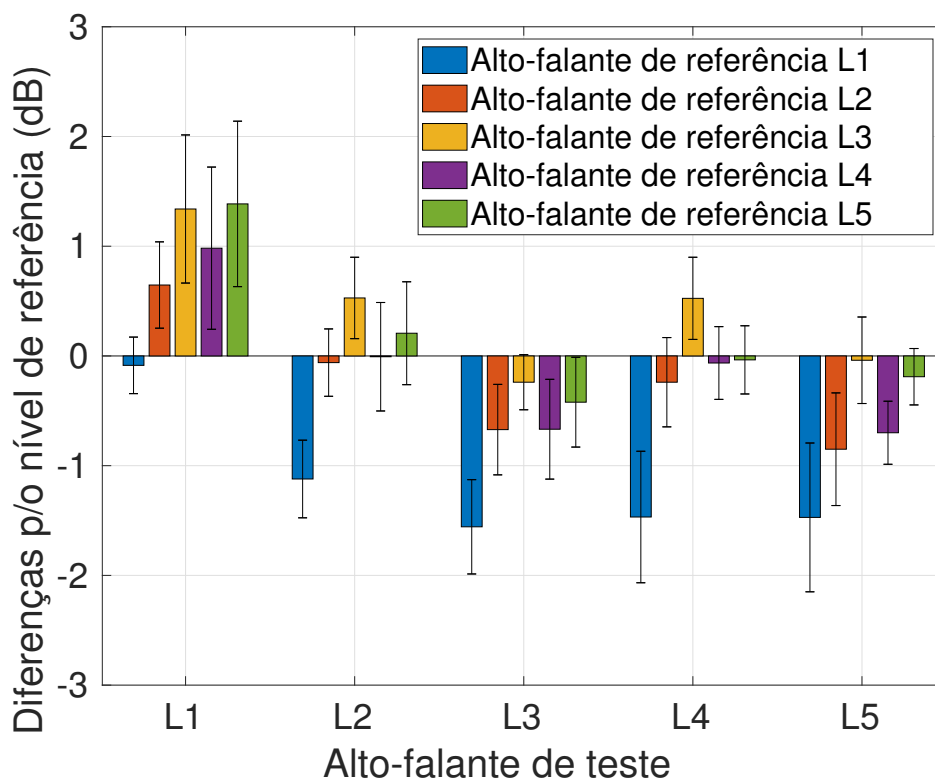
O diagrama de caixa de diferenças de nível por distâncias de referência – superposto pelas distribuições da variável de resposta agrupadas por alto-falante de referência – ilustrado na [Figura 5.14a](#) indica que a maioria dos casamentos de *loudness* com o alto-falante mais próximo deu-se em ajustes inferiores ao nível de referência, e a maioria dos casamentos de *loudness* com o alto-falante mais distante deu-se em ajustes superiores aos níveis de referência. Ainda para os alto-falantes extremos, note que suas medianas (marcadas pelas linhas centrais em cada uma das caixas) estão deslocadas das médias de suas distribuições estimadas (marcadas pelos picos das curvas em formato de sino) sugerindo uma não-normalidade mais pronunciada na referência mais próxima, e menos pronunciada na referência mais distante do ouvinte.

Objetivando verificar tanto a linearidade quanto a homogeneidade de variância dos dados experimentais, estes foram ajustados a um Modelo Linear Generalizado (GLM) com os fatores “distância do alto-falante de teste”, “distân-

Figura 5.12 – Médias e intervalos de confiança entre usuários por pares de distâncias de teste/referência: ajustes dos participantes em relação ao nível de referência de 70 dB SPL (0 dB).



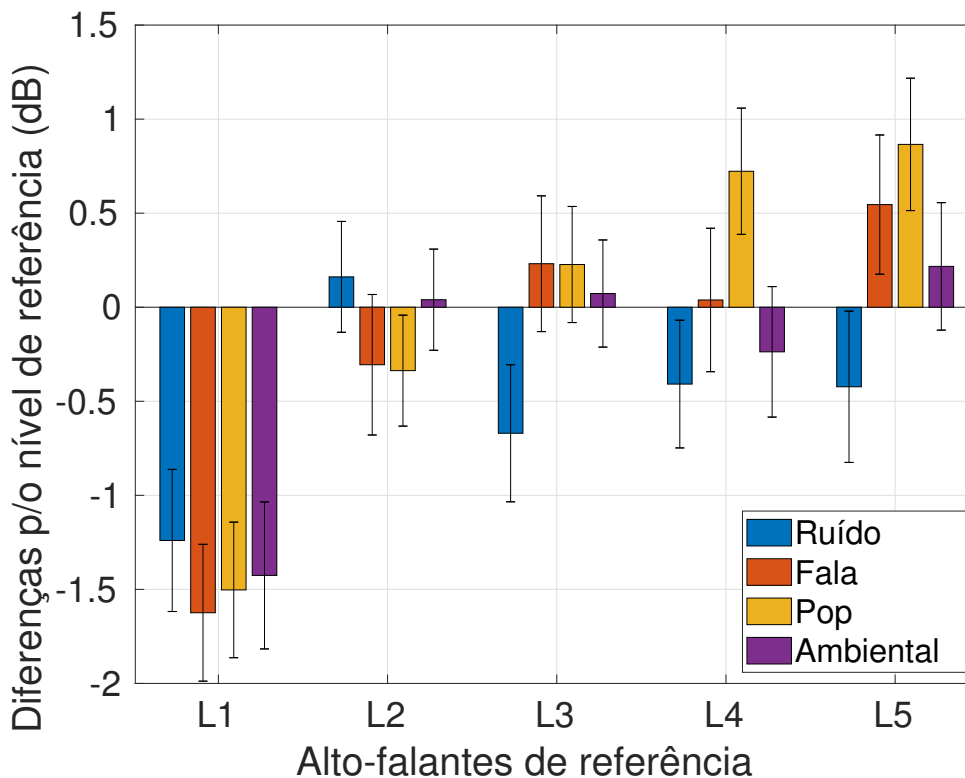
(a) Sala 1.



(b) Sala 2.

Fonte: Elaborada pelo autor.

Figura 5.13 – Médias e intervalos de confiança agrupados por distância do alto-falante de referência ao ouvinte, referentes aos ajustes feitos pelos participantes do experimento nos níveis de reprodução dos alto-falantes de teste, de tal forma que a sensação de *loudness* resultante fosse casada com a sensação de *loudness* produzida por cada alto-falante numa dada distância de referência.

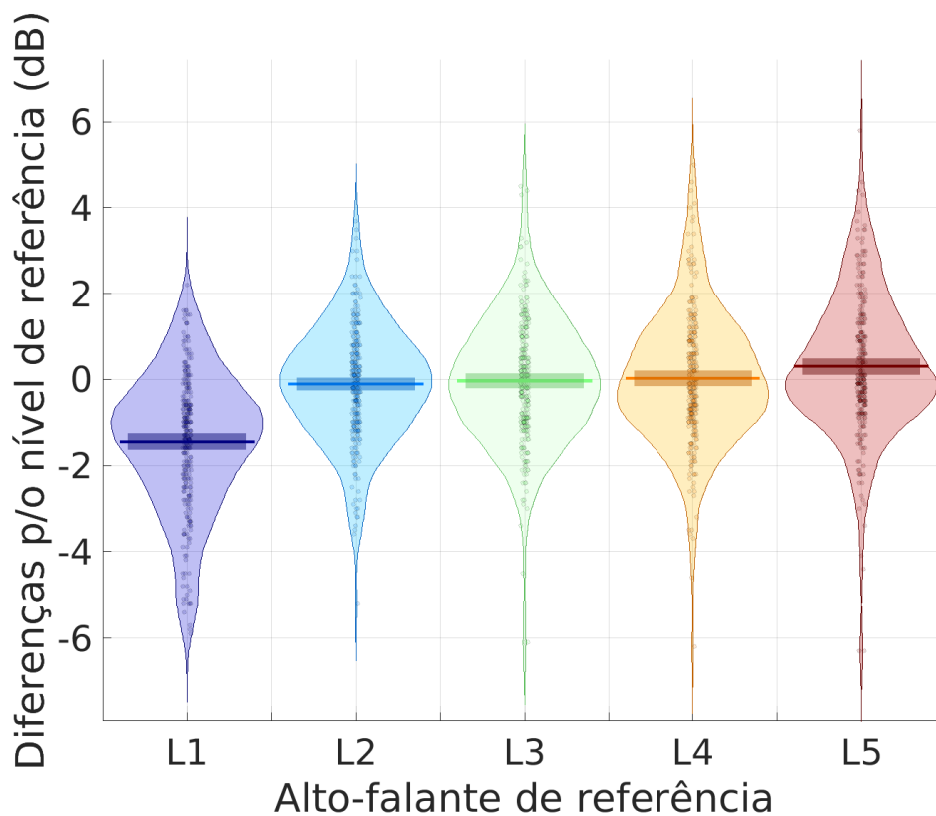


Fonte: Elaborada pelo autor.

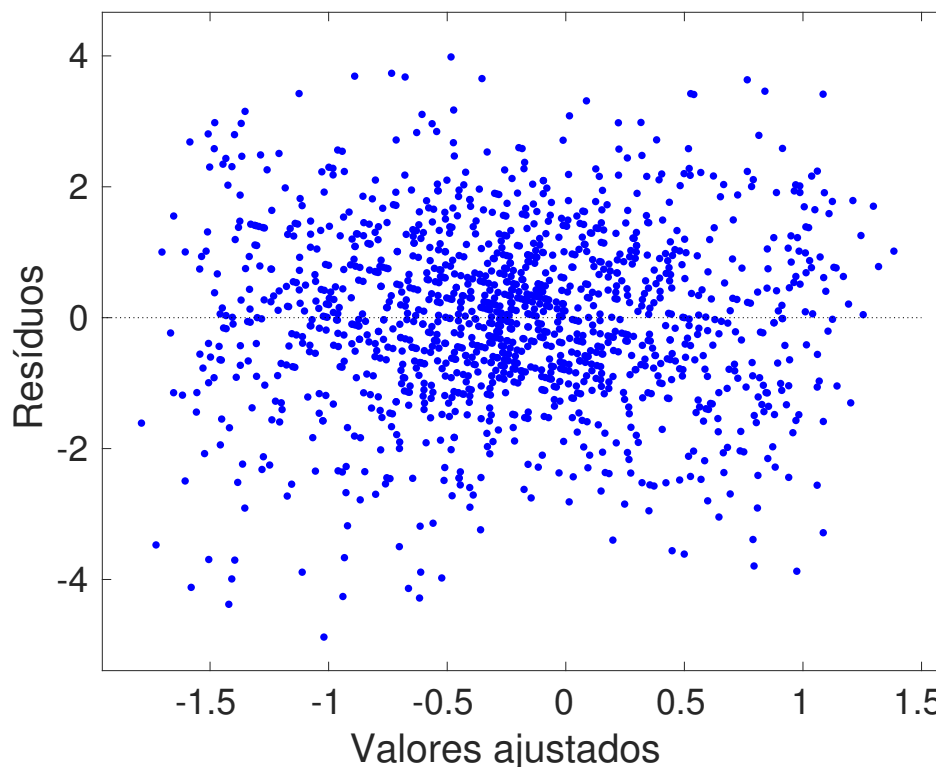
Nota – Dados experimentais segmentados por estímulo. A não-identificação de tendências no comportamento dos ajustes de nível sugere que o efeito da variável “item de programação” nas respostas dos participantes não tenha sido observado neste experimento.

cia do alto-falante de referência”, “sala de reprodução”, “itens de programação”, “participantes” e suas interações. Um gráfico do tipo “Ajustes versus Resíduos” é ilustrado na [Figura 5.14b](#). A partir do espalhamento dos dados, é possível identificar uma relação sistemática entre as previsões do modelo linear e seus resíduos a partir da diminuição da variância dos dados no entorno da origem, o que sugere problemas de heteroscedasticidade (heterogeneidade de variâncias entre os grupos observados) (NOBRE; SINGER, 2007).

Figura 5.14 – Análise exploratória dos ajustes de nível realizados pelos participantes do experimento.



(a) Diagramas de caixa e distribuições estimadas. O descolamento entre médias/medianas sugere distribuições não-normais nas distâncias extremas.



(b) Ajustes contra resíduos de um GLM. A concentração do espalhamento dos dados na origem sugere problemas de heteroscedasticidade.

Estatística descritiva

Para fins de tratamento de "pontos fora da curva" (*outliers*), os dados foram divididos em grupos por distância da fonte sonora / sala de reprodução. Ajustes superiores a $\pm 2,5 \times \sigma$ foram considerados como erros na execução da tarefa e então removidos (*trimmed*) do conjunto. *Outliers* remanescentes nestes grupos foram recodificados (*winsorized*) pelos dados não-*outliers* mais extremos.

Os dados distribuídos nos grupos de alto-falantes de referência ilustrados na Figura 5.14a têm valores de curtose entre 2,77 e 3,19. E como a figura sugere, os dados do alto-falante número 1 (L1), situado a uma distância $d = 0,8$ m, são os mais assimétricos à esquerda ($-0,4921$), e os dados do alto-falante número 5 (L5), situado a uma distância $d = 4,5$ m, são os mais assimétricos à direita ($0,3703$). Testes de Kolmogorov-Smirnov ajustados – ou testes de Lilliefors – para avaliação de normalidade dos mesmos dados agrupados por alto-falante de referência rejeitaram a hipótese nula num intervalo de confiança de 95% tanto em L1 [$D^*(277) = 0,0827, p < 0,001$] quanto em L5 [$D^*(272) = 0,0991, p < 0,001$].

Com base na suspeita levantada pelo gráfico da Figura 5.14b, testes de Fligner-Killeen foram feitos para avaliar a homogeneidade de variâncias entre grupos. Apesar de no agrupamento ilustrado na Figura 5.14a – cinco alto-falantes de referência com dados de ambas as salas – não se haver rejeitado a hipótese nula de homoscedasticidade [$FK(4) = 8,7614, p = 0,067$], esta foi rejeitada em todos os agrupamentos por sala de reprodução, ou por alto-falante de referência e sala de reprodução. Uma vez que seria desejável observar, além do efeito da distância, o efeito da reverberação nas salas, faz-se necessária uma transformação dos dados antes de se prosseguir com a ANOVA.

Estatística inferencial

Para que as premissas da ANOVA fossem satisfeitas, os dados passaram por uma transformação *Fisher-z* com o objetivo de torná-los normalmente distribuídos:

$$z = \operatorname{artanh}(r), \quad (5.16)$$

onde r são os coeficientes de correlação produto-momento de Pearson entre os níveis sonoros ajustados e os níveis sonoros de referência. Os dados transformados foram testados para homoscedasticidade com o teste de Fligner-Killeen mediana χ^2 e a hipótese nula de homogeneidade de variância não foi rejeitada para um intervalo de confiança de 95% [$FK(49) = 57,03, p = 0,201$].

Uma ANOVA de quatro vias foi empregada nos dados *Fisher-z* transformados com os fatores “distância do alto-falante de teste”, “distância do alto-falante de referência”, “sala de reprodução”, “itens de programação” e suas interações⁷. Com a análise, foi possível observar um efeito muito significativo da interação das distâncias dos alto-falantes de teste com as distâncias dos alto-falantes de referência [$F_{(16,136)} = 5,98, p < 0,001, \eta_p^2 = 0,413$]. Outros fatores com tamanho de efeito significativo foram a distância do alto-falante de teste [$F_{(4,136)} = 3,85, p = 0,005, \eta_p^2 = 0,102$] e a sala de reprodução [$F_{(1,136)} = 6,38, p = 0,013, \eta_p^2 = 0,045$]. O efeito do fator “item de programação” não foi observado de modo significativo [$F_{(3,136)} = 0,99, p = 0,116$].

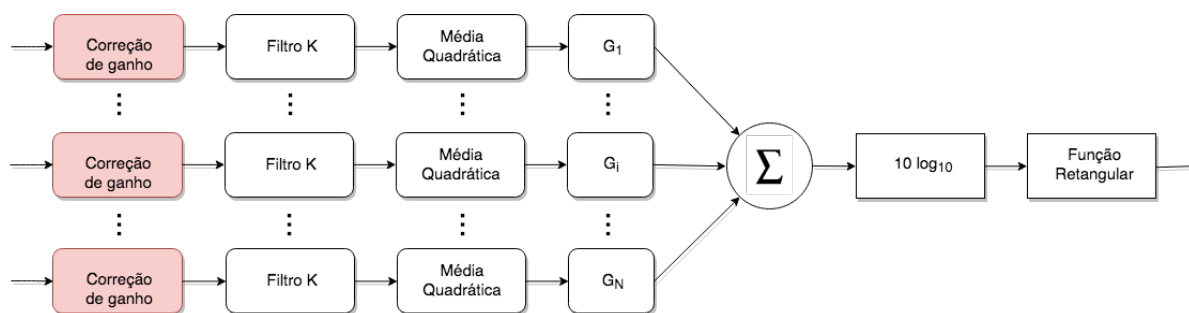
5.3.4 Modelo de loudness ITU-R como função da distância

Dado que os tamanhos dos efeitos relacionados à distância respondem por metade da variância dos dados experimentais, considerar o fator numa modificação do algoritmo de *loudness* BS.1770 é algo bastante razoável. Como visto na [subseção 3.3.3](#), sua ponderação K consiste numa filtragem em duas etapas: a primeira contabiliza o sombreamento da cabeça nas altas frequências quando da incidência de ondas planas (BROWN; DUDA, 1998) e a segunda pondera o sinal em frequência com a curva *RLB* ilustrada na [Figura 3.11](#). Portanto, o ponto identificado para a modificação foi o anterior ao filtro K , no início da cadeia de processamento (ver [Figura 5.15](#)).

Para o bloco de correção de ganho, as médias das respostas dos participantes por distância de referência foram ajustadas a um modelo exponencial de dois termos [$SSE = 0,001335, R\text{-quadrado} = 0,9993, R\text{-quadrado ajustado} = 0,9972,$

⁷ A fórmula desta ANOVA de quatro vias difere da fórmula do GLM da análise exploratória no fator "participantes" e suas interações, visto que os coeficientes de correlação necessários para a transformação dos dados foram calculados entre participantes.

Figura 5.15 – Diagrama em blocos da modificação proposta no algoritmo BS.1770 (adaptado de ITU-R (ITU-R, 2015b)). Blocos de correção de ganho em função da distância foram inseridos na cadeia de processamento multicanal.



Nota – Somente após ter suas magnitudes corrigidas em distância, o sinal terá então suas magnitudes corrigidas em frequência para um mesmo nível de *loudness*.

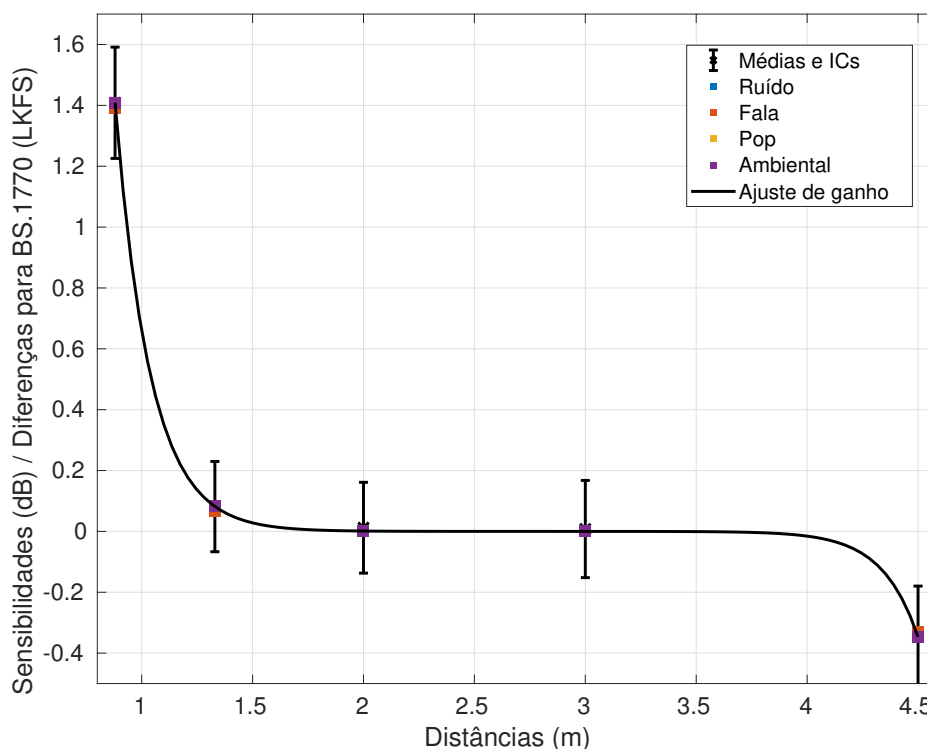
Fonte: Elaborada pelo autor.

$RMSE = 0,03653$] tal que este definisse uma expressão do ganho em função da distância. Somente após ter suas magnitudes corrigidas em distância, o sinal terá então suas magnitudes corrigidas em frequência para um mesmo nível de *loudness*.

Já na Figura 5.16, têm-se as mesmas médias e intervalos de confiança cortados pela curva de ganho resultante de seus ajustes ao modelo exponencial de dois termos, além das diferenças em Unidades de *Loudness* (*Loudness Units* – LU) entre as medidas dos itens de programação feitas pelos algoritmos BS.1770 modificado como função da distância e BS.1770 original. Note que, para distâncias fonte-ouvinte típicas de uma sala de estar (entre 1,30 e 3,00 metros), tanto as diferenças entre os valores de *loudness* medidos quanto o nível de *loudness* de referência (0 LU) se encontram no interior dos intervalos de confiança das diferenças de nível ajustadas pelos participantes. Esta verificação sugere que o modelo ITU-R original seja adequado para este intervalo de distâncias de fonte sonora. Entretanto, para as posições $d = 0,88$ m e $d = 4,50$ m, nas quais os efeitos de distância são mais pronunciados, somente as medidas do algoritmo BS.1770 modificado foram aderentes aos casamentos de *loudness* executados pelos ouvintes.

Por fim, este experimento possibilitou uma melhor compreensão das influências da energia reverberante e da distância auditiva percebida na sensação de *loudness*, dicas importantes de localização espacial não contempladas pela Recomendação BS.1770 para o cômputo do nível de *loudness* no áudio digital.

Figura 5.16 – Médias e intervalos de confiança agrupados por distância do alto-falante de referência ao ouvinte, referentes aos ajustes feitos pelos participantes do experimento nos níveis de reprodução dos alto-falantes de teste, de tal forma que a sensação de *loudness* resultante fosse casada com a sensação de *loudness* produzida por cada alto-falante numa dada distância de referência.



Fonte: Elaborada pelo autor.

Nota – Curva de correção de ganho e diferenças de medidas de *loudness* dos estímulos entre os algoritmos BS.1770 modificado e original (0 LU). Nas distâncias extremas, somente as medidas feitas pelo novo modelo se encontram no interior dos ICs dos ajustes feitos pelos participantes.

O modelo de *loudness* proposto como função da distância da fonte sonora, além de compatível com o método vigente entre distâncias de 1,30 a 3,00 metros – distâncias tipicamente encontradas na sala de reprodução do consumidor de *home audio* e usuário do serviço de radiodifusão digital – trouxe resultados comparáveis com os dos testes subjetivos também nas distâncias de 0,88 e 4,50 metros. Estes resultados foram divulgados por Pires *et al.* (2018) no XXVIII Encontro da Sociedade Brasileira de Acústica.

Conduzir sessões do experimento em salas com tempos de reverberação distintos foi importante para se identificar uma tendência à invariância de *loudness* com a distância da fonte conforme o aumento da energia reverberante,

manifestada por uma redução geral das magnitudes ajustadas pelos participantes do experimento da Sala 1 para a Sala 2, e verificada na análise de variâncias por um fator “sala de reprodução” com apenas dois níveis (Salas 1 e 2) ser responsável por 4,5% da variância dos dados. A investigação sobre a significância deste efeito experimental será detalhada na seção seguinte.

5.4 Relação entre *Loudness* e Reverberação

No experimento anterior, o fator “sala de reprodução” teve um efeito significativo na variável de resposta [$F_{(1,136)} = 6,38$, $p = 0,0127$, $\eta_p^2 = 0,0448$]. O mesmo não pôde ser dito com relação aos efeitos de suas interações. Este comportamento das respostas dos participantes em função das salas de reprodução sugere que o efeito da energia reverberante no *loudness* é significativo de tal maneira que é indiferente à significância, ou até mesmo à presença, de interações deste com os demais fatores experimentais presentes no modelo.

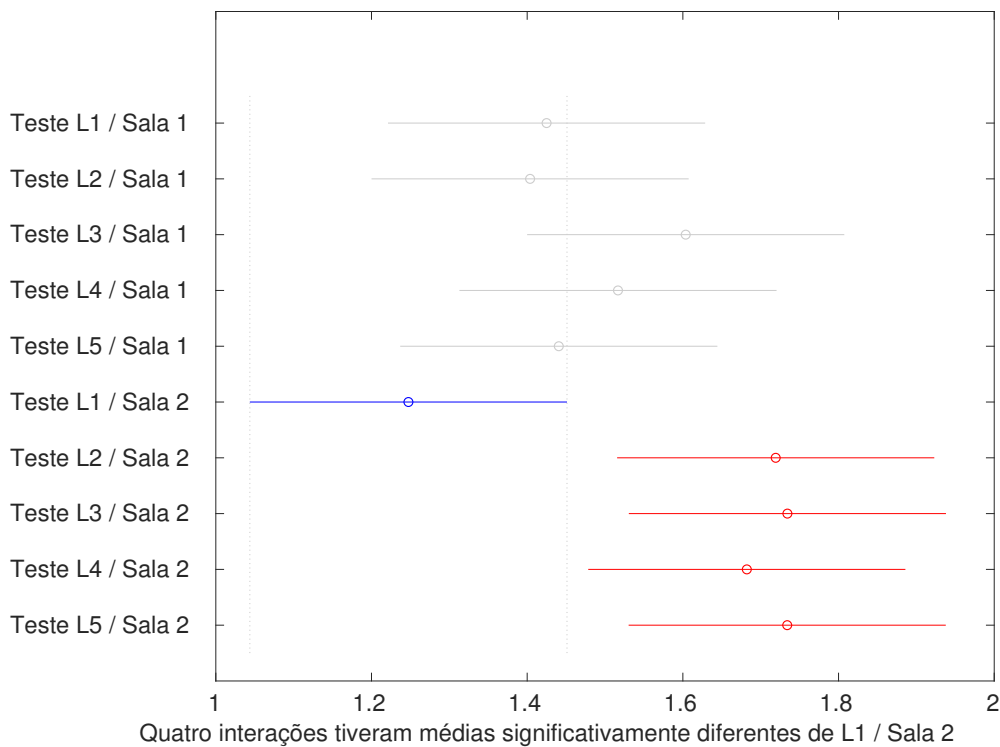
Tamanha indiferença pode ser notada na [Figura 5.17](#), na qual o incremento no tempo de reverberação da sala de escuta crítica BS.1116 para o tempo de reverberação da sala de aula comum levou as comparações par-a-par a uma condição de invariância de *loudness* com a distância. Note que as diferenças entre alto-falantes na sala de aula foram menos significativas do que na sala de escuta crítica como um todo exceto para o alto-falante mais próximo, cuja sensação de *loudness* provocada é muito mais dependente da onda direta do que de suas reflexões.

Muito embora o tempo de reverberação não seja descrito nos metadados do modelo de definição de áudio BS.2076 do [ITU-R \(2017\)](#), é válido investigar seu efeito nas sensações de *loudness* provocadas na posição do ouvinte.

5.4.1 Efeito da reverberação

De modo a conseguir-se uma interpretação física do efeito da reverberação no *loudness*, medidas de resposta ao impulso foram feitas em ambas as salas do experimento anterior, às mesmas cinco distâncias em relação à posição do ouvinte, com o manequim (HATS) no seu lugar.

Figura 5.17 – Comparações nível a nível dos fatores “alto-falante de teste” e “sala de reprodução”.



Nota – As diferenças entre alto-falantes na sala de aula (Sala 2) foram menos significativas do que na sala de escuta crítica (Sala 1) como um todo exceto para o alto-falante mais próximo, cuja sensação de *loudness* provocada é muito mais dependente da onda direta do que de suas reflexões.

Fonte: Elaborada pelo autor.

A técnica de medida empregada é denominada “varredura senoidal”, que usa uma varredura em frequência exponencialmente crescente no tempo da forma:

$$x(t) = \text{sen} \left[\frac{\omega_1 \cdot T}{\ln \left(\frac{\omega_2}{\omega_1} \right)} \left(e^{\frac{t}{T} \ln \left(\frac{\omega_2}{\omega_1} \right)} - 1 \right) \right], \quad (5.17)$$

que tem início numa frequência angular ω_1 e termina numa frequência angular ω_2 em T segundos. Quando este sinal é reproduzido pelo alto-falante, o sinal capturado resultante exhibe os efeitos da reverberação na sala, que espalha o sinal no tempo. O sinal capturado passa então por um processo de deconvolução (operação inversa da convolução), resultando na resposta ao impulso do sistema linear (FARINA, 2007).

Dez varreduras senoidais de 10 Hz a 22 kHz, com durações de 6 segundos e períodos de silêncio de 4 segundos, foram reproduzidas em cada um dos alto-

falantes e as respostas correspondentes foram capturadas pelo HATS como o objetivo de se calcular os tempos de reverberação em ambas as salas, conforme padrão 3382-1 da ISO (2009). Tempos médios de reverberação entre as faixas de oitava de 1000 Hz a 4000 Hz, calculados a partir de impulsos capturados a dois metros de distância da fonte sonora, resultaram num $RT_{60} \approx 0,22$ s para a sala de escuta crítica BS.1116 e num $RT_{40} \approx 1,30$ s para a sala de aula comum⁸. Estimções das curvas de decaimento de energia e ajustes lineares dos dados à inclinação entre 0 dB e 60 dB na primeira sala e à inclinação entre 0 dB e 40 dB na segunda sala, são ilustrados na Figura 5.18 e na Figura 5.19, respectivamente.

De modo a traduzir estas Respostas Biauriculares das Salas ao Impulso (BRIRs) na forma de métricas de desempenho, computaram-se as Razões entre Energia Direta e Energia Reverberante (DRRs) dos sinais capturados para cada distância fonte-ouvinte em ambas as salas. Segundo Zahorik, Brungart e Bronkhorst (2005), a DRR é amplamente aceita como métrica de predição de distância percebida de fonte sonora e é definida por:

$$DRR = 10 \log_{10} \frac{\int_{T_0-c}^{T_0+c} s^2(t) dt}{\int_{T_0+c}^{\infty} s^2(t) dt}, \quad (5.18)$$

onde o sinal $s(t) = h_l(t)$ é a resposta ao impulso relativa ao l -ésimo alto-falante, e T_0 é a duração do som direto. Nos cálculos deste trabalho, a constante de tempo utilizada foi $c = 2,5$ ms, mesmo valor usado pelo próprio Zahorik (2002).

Os valores de DRR foram então agrupados por sala de reprodução e associados às distâncias físicas correspondentes. Para cada sala, os pares (d, DRR) foram ajustados a um modelo não-linear de intensidade proporcional ao inverso do quadrado da distância ($I \propto 1/d^2$), e a qualidade de ajuste foi medida pelas estatísticas enumeradas a seguir:

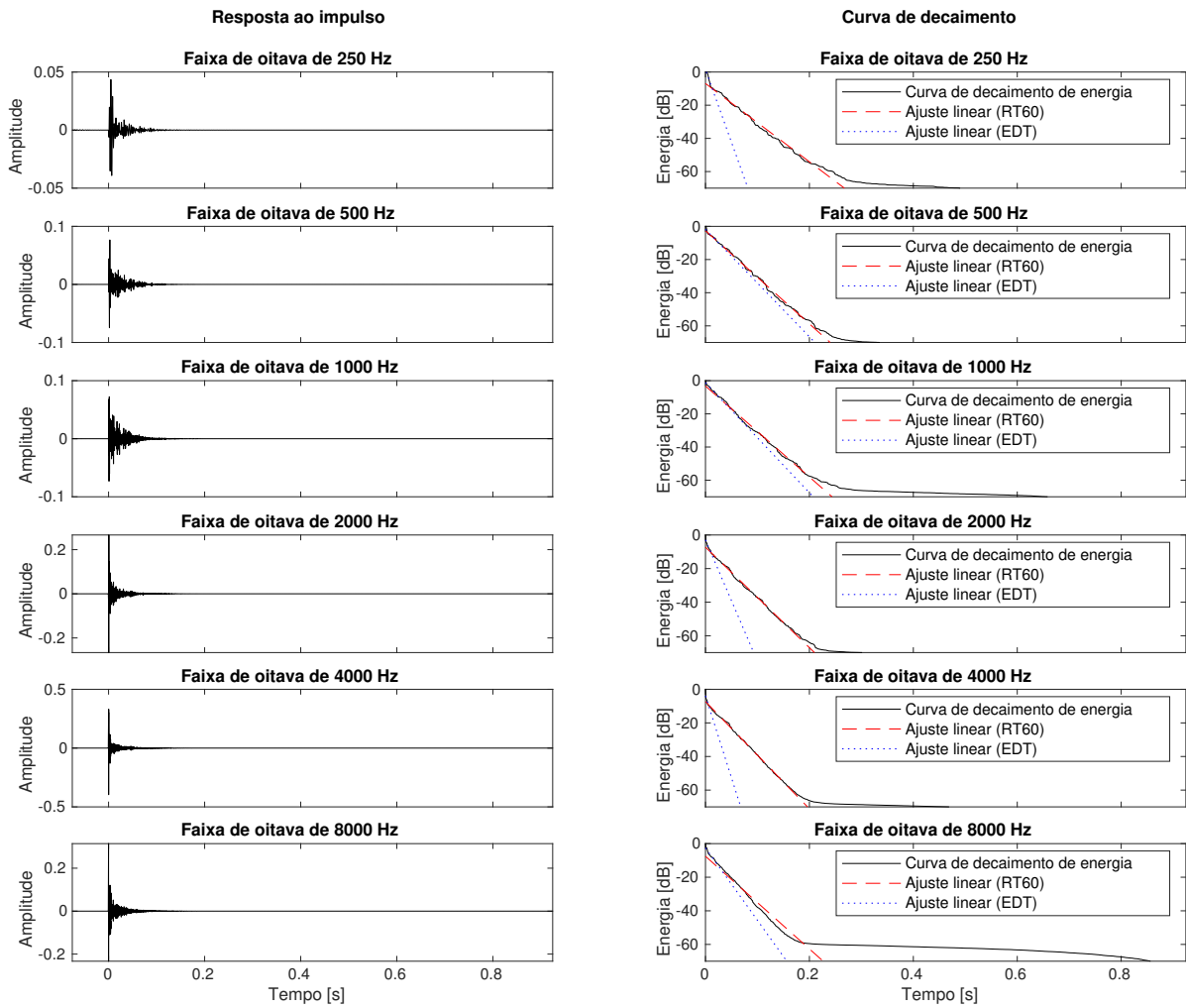
1. Soma dos Quadrados dos Erros (SSE),

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2; \quad (5.19)$$

onde para cada i -ésimo elemento do conjunto de medidas, y_i é o valor de DRR calculado e \hat{y}_i é o valor correspondente no modelo não-linear

⁸ O piso de ruído na sala de aula tornou proibitiva sua caracterização por um decaimento de energia de 60 dB (RT_{60}).

Figura 5.18 – Respostas ao impulso e tempos de reverberação por oitava na sala de escuta crítica BS.1116.



Nota – A curva de decaimento de energia é acompanhada do ajuste linear correspondente ao cálculo do tempo de reverberação (RT60) e do ajuste linear do Tempo de Decaimento de Energia (EDT), correspondente à taxa de queda nos primeiros 10 dB, perceptivamente mais importante.

Fonte: Elaborada pelo autor.

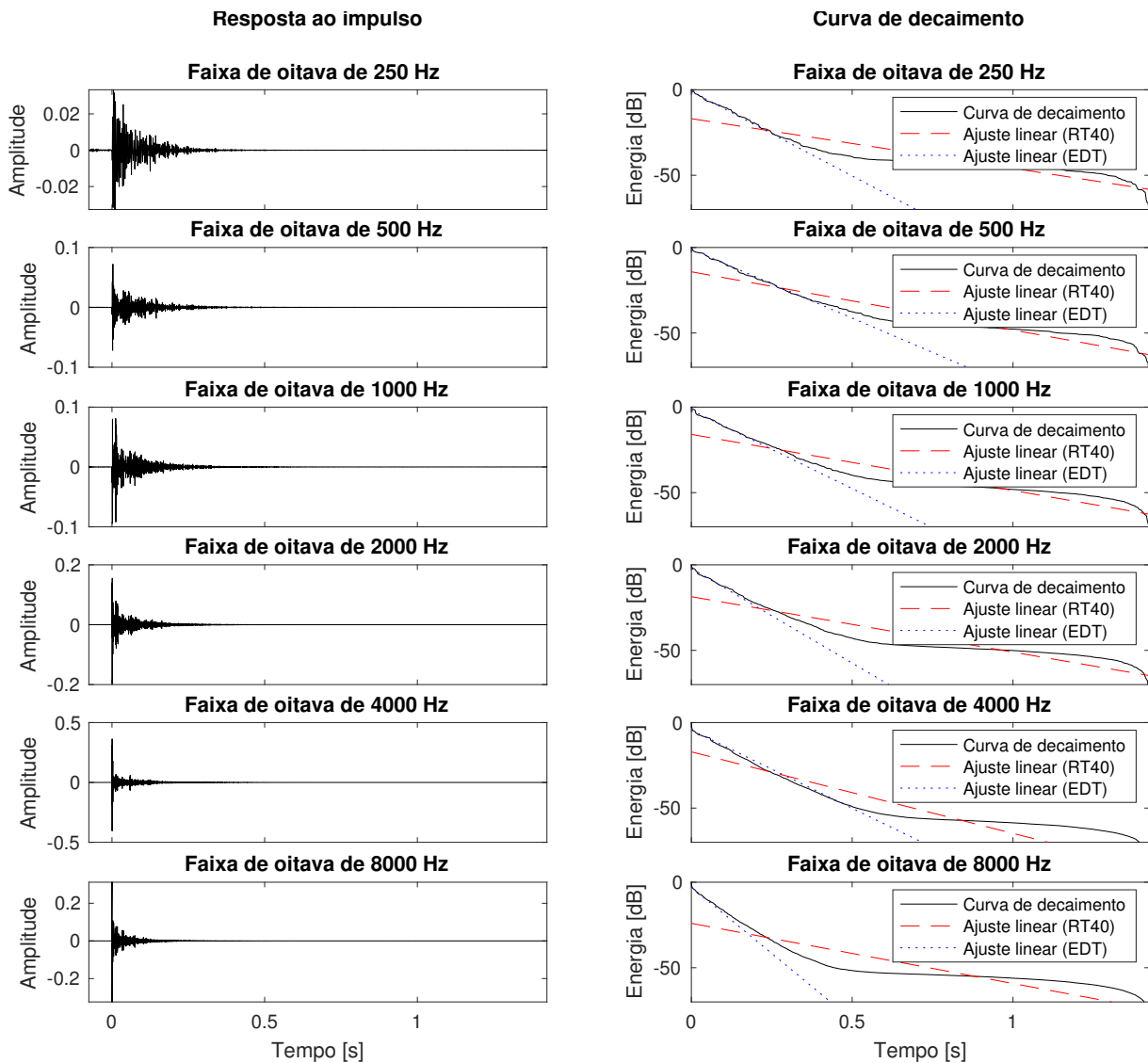
de $I \propto 1/d^2$. Um valor próximo de 0 indica que o modelo possui uma componente de erro aleatório pequena.

2. R-Square

Definida como a razão da Soma dos Quadrados da Regressão (SSR) e a Soma dos Quadrados dos Totais (SST):

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2; \quad (5.20)$$

Figura 5.19 – Respostas ao impulso e tempos de reverberação por oitava na sala de aula comum.



Nota – A curva de decaimento de energia é acompanhada do ajuste linear correspondente ao cálculo do tempo de reverberação (RT40) e do ajuste linear do Tempo de Decaimento de Energia (EDT), correspondente à taxa de queda nos primeiros 10 dB, perceptivamente mais importante.

Fonte: Elaborada pelo autor.

$$SST = \sum_{i=1}^n (y_i - \bar{y}_i)^2; \quad (5.21)$$

$$Rsquare = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}; \quad (5.22)$$

onde \bar{y}_i é a média entre o valor calculado de DRR e o valor estimado pelo modelo para o i -ésimo elemento. Seus valores situam-se entre 0 e 1, sendo

que a proximidade de 1 indica que uma maior proporção da variância é levada em consideração pelo modelo.

3. *R-Square* Ajustado

Estatística *R-Square* ajustada pelos graus de liberdade dos resíduos (df = 3):

$$df = n - m, \tag{5.23}$$

onde $n = 5$ é a quantidade de valores de respostas e $m = 2$ é o número de coeficientes ajustados

$$\text{AdjRsquare} = 1 - \frac{SSE(n - 1)}{SST(df)}, \tag{5.24}$$

onde valores próximos de 1 indicam um melhor ajuste.

4. RMSE (Raíz Média Quadrática do Erro)

$$RMSE = s = \sqrt{MSE} = \sqrt{\frac{SSE}{df}}, \tag{5.25}$$

onde valores próximos de 0 indicam um melhor ajuste.

As estatísticas de ambas as salas estão relacionadas na [Tabela 5.7](#).

Tabela 5.7 – Estatísticas de qualidade de ajuste dos dados de Razões entre Energia Direta e Energia Reverberante (DRRs) ao modelo de intensidade proporcional ao inverso do quadrado da distância.

Faixas de oitava	Sala 1 – Qualidade ajuste				Sala 2 – Qualidade ajuste			
	SSE	R^2	Adj. R^2	RMSE	SSE	R^2	Adj. R^2	RMSE
250 Hz	4,12	0,79	0,72	1,17	–	–	–	–
500 Hz	10,99	0,60	0,47	1,91	27,96	0,22	0,03	3,05
1000 Hz	8,54	0,89	0,86	1,68	7,81	0,77	0,69	1,61
2000 Hz	15,52	0,90	0,86	2,27	40,81	0,65	0,53	3,68
4000 Hz	3,18	0,97	0,96	1,03	–	–	–	–
8000 Hz	32,64	0,73	0,65	3,29	–	–	–	–

As Relações entre Energia Direta e Energia Reverberante (DRRs) com a distância estão melhor ajustadas ao modelo não-linear de inverso do quadrado da distância na sala de escuta crítica (Sala 1) do que na sala de aula comum (Sala 2). Isso reforça a noção de que o relacionamento do *loudness* com a distância seja mais próximo do relacionamento da intensidade sonora com a distância na sala

menos reverberante. Por outro lado, pode ser dito que a relação do *loudness* com a distância esteja mais próxima de uma condição de invariância na sala mais reverberante.

Muito embora esta conclusão seja corroborada pela significância do efeito da sala de reprodução na variável de resposta do experimento anterior e pelas diferenças de qualidade de ajuste das DRRs a um modelo de $I \propto 1/d^2$, seria desejável trabalhar com mais que dois níveis no fator “sala de reprodução”, mas isso não pôde ser feito no experimento da [seção 5.3](#) por razões logísticas quanto à reserva de salas, ao transporte de equipamentos, e à montagem/desmontagem do arranjo experimental. Uma possibilidade de superação desta limitação seria via simulação de salas, como no experimento da [seção 4.3](#), de tal forma que um participante pudesse executar tarefas de casamento de *loudness* em várias “salas” durante uma única sessão. Um experimento foi então elaborado com este racional.

Pergunta de pesquisa e hipóteses a testar

Ao se destacar o efeito da reverberação na sensação de *loudness*, a pergunta de pesquisa pode ser formulada da forma abaixo:

- *Como a sensação de loudness é afetada pela variação dos tempos de reverberação nas salas de reprodução?*

E as hipóteses orientadas à reverberação seriam da forma:

- Hipótese nula: *Sons reproduzidos por alto-falantes situados a diferentes direções na mesma sala, ou situados à mesma direção em salas diferentes, que tenham o mesmo nível de loudness medido na posição do ouvinte, provocarão a mesma sensação de loudness.*
- Hipótese alternativa: *Sons reproduzidos por alto-falantes situados a diferentes direções na mesma sala, ou situados à mesma direção em salas diferentes, que tenham o mesmo nível de loudness medido na posição do ouvinte, provocarão diferentes sensações de loudness.*

Note que as hipóteses formuladas contemplam um efeito posicional além do próprio efeito da reverberação. Isto se deve à metodologia experimental escolhida. A próxima subseção descreve os procedimentos e métodos usados nesta investigação.

5.4.2 Projeto de experimento

Objetivando testar as hipóteses apresentadas, um experimento de casamento de *loudness* deveria ser desenvolvido tendo em mente um maior número de níveis do fator experimental “sala de reprodução”. Isto foi possível com sínteses binauriculares de sons reproduzidos em diferentes ambientes, caracterizados pelos seus tempos de reverberação.

O que se segue é um relato sobre capturas preliminares de BRIRs, condições iniciais, métodos procedurais e projeto estatístico.

Respostas Binauriculares das Salas ao Impulso (BRIRs)

As respostas ao impulso foram capturadas com o uso do Simulador de Cabeça e Torso (HATS). As produções das BRIRs seguiram o mesmo procedimento descrito anteriormente: a partir de varreduras senoidais reproduzidas por um alto-falante ativo Genelec 8020A e subsequente deconvolução dos sinais capturados. As gravações foram realizadas a uma taxa de amostragem de 48.000 amostras/s. O alto-falante foi posicionado ao redor do HATS num arco no plano médio com 1,5 m de raio entre $\pm 90^\circ$ e as medidas foram feitas em intervalos de 15° . O eixo acústico do alto-falante foi alinhado à mesma altura das orelhas do manequim. Um resumo das propriedades acústicas de cada sala é listado na [Tabela 5.8](#). O RT_{60} geral de cada sala foi calculado pela média nas faixas de oitava centradas em 500 Hz e 1 kHz. Outros parâmetros foram calculados *a posteriori* diretamente das respostas ao impulso.

Para a condição anecoica, foi utilizado um método pseudo-anecoico nos quais as respostas foram capturadas numa sala ampla e truncadas no instante anterior à primeira reflexão, de tal forma que as reflexões subsequentes não colorissem a resposta em frequência. A sala possuía dimensões de $17,04 \times 14,53 \times 6,5 \text{ m}^3$ ($c \times l \times a$); o HATS e o alto-falante foram posicionados no

Tabela 5.8 – Propriedades acústicas das salas, incluindo RT_{60} , Lacuna Inicial de Retardo Temporal (ITDG), Relação entre Energia Direta e Energia Reverberante (DRR) e índice de clareza C_{te} .

Sala	RT (s)	ITDG (ms)	DRR (dB)	C_{te} (50 ms) (dB)
A	0,32	8,72	6,09	16,50
B	0,47	9,66	5,31	11,40
C	0,68	11,90	8,82	17,40
D	0,89	21,60	6,12	9,43

Fonte: [Hummerson \(2011, p.2\)](#).

centro da sala a uma altura de 2,8 m e separados por 1,5 m.

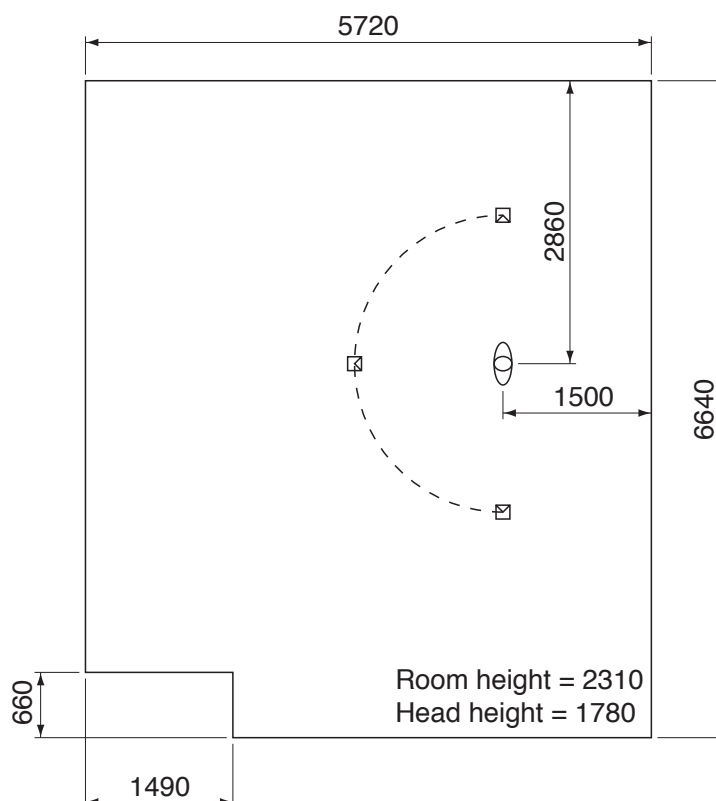
A Sala A era a própria sala dos alunos de pós-graduação e pesquisadores do Instituto de Gravação de Som (IoSR) da Universidade de Surrey: um escritório de tamanho médio com capacidade para 8 pessoas e com um tempo de reverberação RT_{60} considerado baixo para sua extensão. A sala possuía dimensões de $6,64 \times 5,72 \times 2,31 \text{ m}^3$ ($c \times l \times a$); o HATS e o alto-falante foram posicionados no centro da sala a uma altura de 1,7 m e separados por 1,5 m. A planta da sala e a localização do HATS estão ilustrados na [Figura 5.20](#).

A Sala B era uma sala de aula de tamanho médio para pequeno. Apesar do formato “caixa de sapatos”, a construção da sala lhe conferiu um RT_{60} relativamente longo para o seu tamanho. A sala possuía dimensões de $4,65 \times 4,65 \times 2,68 \text{ m}^3$ ($c \times l \times a$); o HATS e o alto-falante foram posicionados no centro da sala a uma altura de 1,9 m e separados por 1,5 m. A planta da sala e a localização do HATS estão ilustrados na [Figura 5.21](#).

A Sala C era um auditório amplo estilo cinematógrafo com capacidade para 418 pessoas. Contudo, a abundância de estofados e o teto baixo resultou num RT_{60} pequeno para o tamanho da sala. A sala possuía dimensões de $18,80 \times 23,50 \times 4,60 \text{ m}^3$ ($c \times l \times a$); o HATS e o alto-falante foram posicionados no centro da sala a uma altura de 1,9 m e separados por 1,5 m. A planta da sala e a localização do HATS estão ilustrados na [Figura 5.22](#).

A Sala D era um espaço de seminários e apresentações de tamanho médio para amplo e com um teto muito alto. A sala possuía dimensões de $8,00 \times 8,70 \times 4,20 \text{ m}^3$ ($c \times l \times a$); o HATS e o alto-falante foram posicionados no centro da sala a uma altura de 1,7 m e separados por 1,5 m. A planta da sala e a localização do HATS estão ilustrados na [Figura 5.23](#).

Figura 5.20 – Sala A: Planta e localização do HATS.



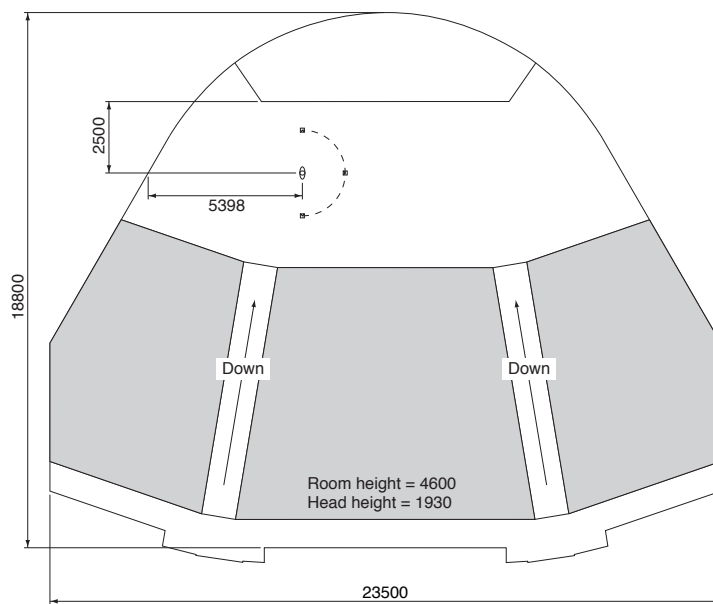
Fonte: [Hummersone \(2011, p.3\)](#).

As BRIRs anteriores foram obtidas pela equipe do IoSR ([HUMMERSONE, 2011](#)). Minha contribuição, à época do estágio doutoral, foi obter as BRIRs da sala de escuta crítica BS.1116 ([Figura 5.24](#)), de dimensões $8,00 \times 8,70 \times 4,20 \text{ m}^3$ ($c \times l \times a$) e $RT_{60} = 0,22 \text{ s}$ e incluí-las no conjunto de dados a serem utilizados nas sínteses biauriculares deste experimento.

Participantes

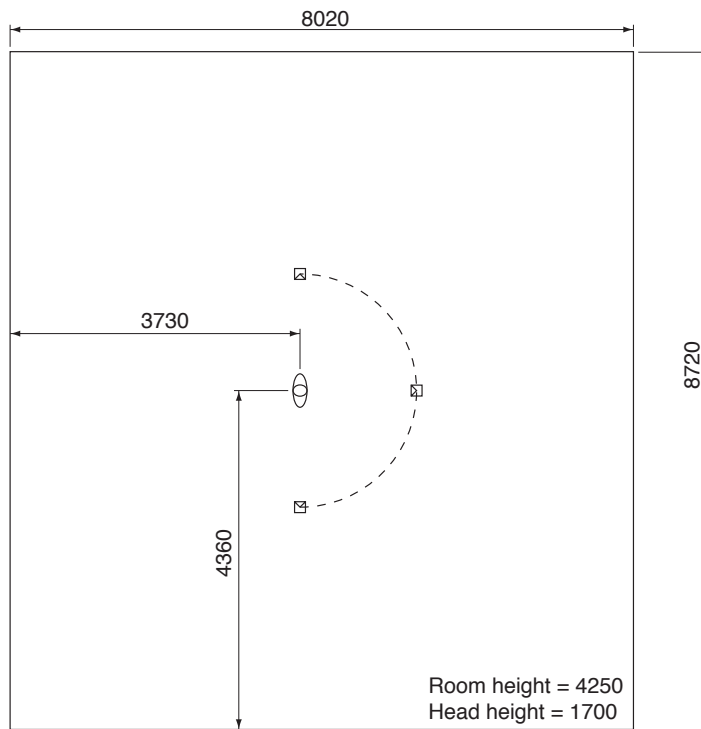
Desta vez, os participantes – em sua maioria estudantes ligados ao Centro de estudos da fala, acústica, linguagem e música (Cefala) da Universidade Federal de Minas Gerais – poderiam ser classificados como “inexperientes”, dado que não tinham treinamento de escuta crítica nem estavam familiarizados com testes perceptivos. Considerando uma maior variabilidade da variável de resposta entre participantes, seria preciso um número maior de observações para melhorar a relação de variâncias efeito/ruído experimental. O esforço de recrutamento foi maior neste experimento do que no anterior.

Figura 5.22 – Sala C: Planta e localização do HATS.



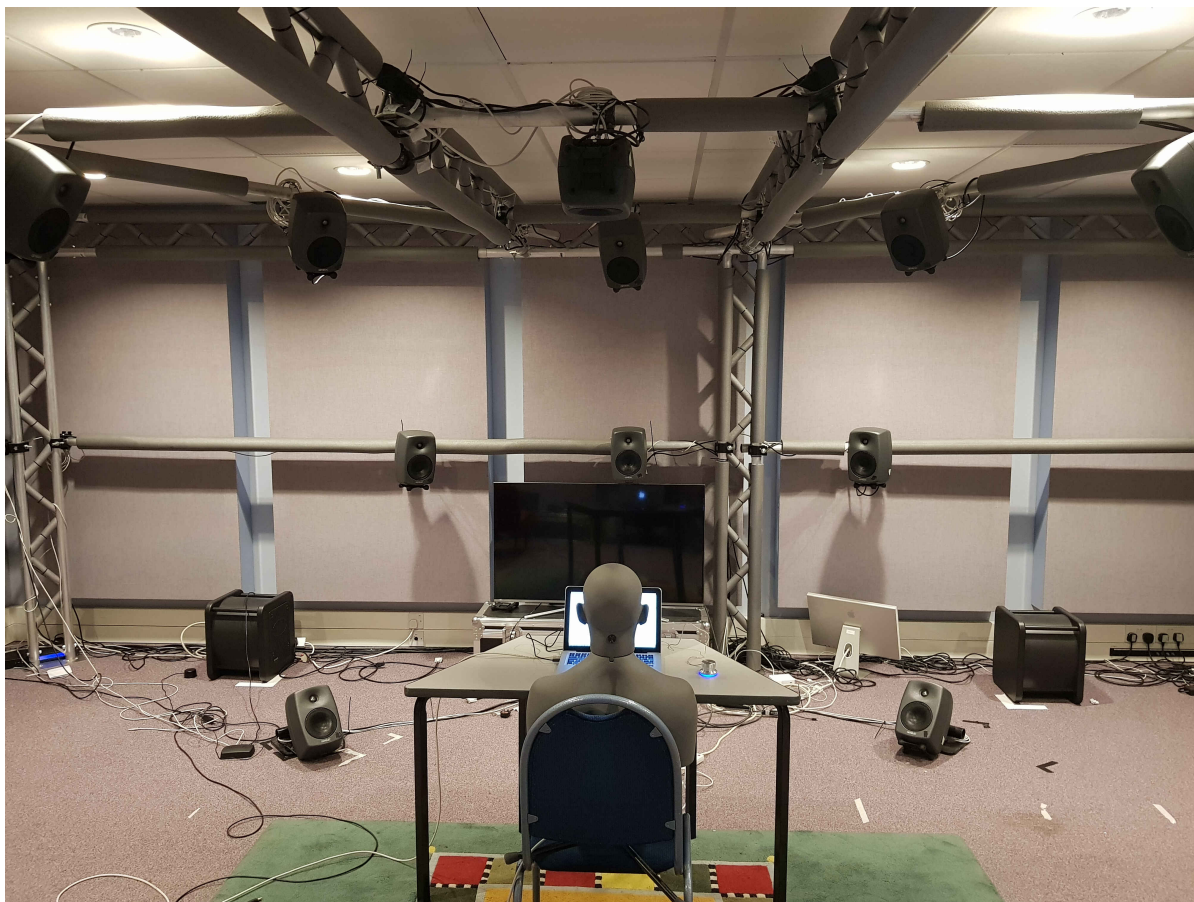
Fonte: Hummersone (2011, p.5).

Figura 5.23 – Sala D: Planta e localização do HATS.



Fonte: Hummersone (2011, p.6).

Figura 5.24 – Sala de escuta crítica em conformidade com a Recomendação BS.1116 do ITU-R (2015a).



Fonte: Elaborada pelo autor.

itens de programação na variável de resposta não foram significativos. Com base nesta conclusão, considerar aqui o estímulo como um fator experimental seria provavelmente inócua, desde que todos os seus níveis fossem sinais de faixa larga.

Muito embora escolher apenas um estímulo dentre os possíveis siga um racional claro, a escolha em si não foi fácil. Foi visto na [subseção 5.3.1](#), durante as verificações preliminares do experimento de distância, que efeitos direcionais tendem a ser melhor observados quando estreitada a largura de faixa dos estímulos, o que é consistente com as escolhas de experimentadores pioneiros e contemporâneos (ROBINSON; WHITTLE, 1960; SHAO; MO; MAO, 2015). Contudo, aqui preferiu-se o estímulo de banda larga – ao custo de tamanhos menores de efeito – porque são mais próximos de sinais reais vindos de diferentes direções, prováveis de serem ouvidos numa transmissão de radiodifusão.

Calibração

A calibração foi realizada tal como na [subsecção 5.3.1](#) e na [subsecção 5.3.2](#), desta vez usando o mesmo *Cortex MK 2* HATS combinado com um par de fones de ouvido de resposta plana *BeyerDynamics DT-150*. Os controles de volume foram calibrados de tal forma que o ruído rosa entrecortado com nível de *loudness* de -23 LKFS, e com uma pressão sonora de 65 dB SPL incidente nos microfones intra-auriculares do manequim, resultou num nível de *loudness* médio de -23 LKFS no sinal biauricular do HATS (ver [Figura 5.25](#)).

Figura 5.25 – Calibração de fones de ouvido em estúdio com um Simulador de Cabeça e Torso (HATS).



Fonte: Elaborada pelo autor.

Equipamentos

Os equipamentos usados nesta série de medidas e no presente experimento são abaixo listados:

- Um par de fones de ouvido de resposta plana *BeyerDynamic DT-150* para reprodução dos estímulos;
- Um Simulador de Cabeça e Torso (HATS) *Cortex MK2* para calibração dos sinais de referência;
- Uma interface conversora analógico-digital *RME Audio Fireface 800* para gravação dos sinais do HATS e calibração;
- Um controlador USB *Griffin Powermate* para ajustes de nível;
- Um *MacBook Pro* executando *MaxMSP[®]*, *MATLAB[®]* e *Audacity*;
- Um *MacBook Pro* para operação remota do anterior.

Apresentação dos estímulos

Como no experimento anterior, o formato das respostas é do tipo escala, produto do método de ajuste da psicoacústica. Os participantes são solicitados a casar o *loudness* de um som de teste com o *loudness* de um som de referência. Este último corresponde à “incidência frontal”, isto é, a partir do ruído rosa monofônico entrecortado e limitado em faixa, é feita uma síntese biauricular com a $HRTF(0^\circ, 0^\circ)$. Já o som de teste corresponde à “incidência lateral”: a síntese biauricular feita com a $HRTF(\vartheta, 0^\circ)$ para um dado $\theta = \vartheta$.

Os sinais podem ser intercambiados a qualquer tempo com algum *cross-fade* e reproduzidos continuamente durante o ajuste com entrecortes de 500 ms. A ordem com que cada participante executa a tarefa para cada fonte sonora específica de uma sala/direção é aleatorizada. Para cada tentativa de casamento de *loudness*, o par de sons de (teste, referência) são da mesma sala virtual. Um projeto fatorial completo composto de 8 direções em 6 salas com uma repetição resulta em 96 ajustes de nível por participante, obtidos numa sessão de 50-60 minutos.

Análise estatística

Sejam $Y_{i,j,s}$ as observações dos ajustes de *loudness* para cada i -ésima sala virtual, j -ésimo azimuth e s -ésimo participante. O processo aditivo de análise de variância é então escrito da forma:

$$Y_{i,j,s} = \underbrace{\mu + \tau_i}_{\mu_i} + \beta_j + \alpha_s + \varepsilon_{i,j,s} \quad (5.26)$$

onde μ é a média geral, τ_i é o efeito do tratamento da i -ésima sala de reprodução na média, β_j é o efeito do j -ésimo azimuth, α_s é o efeito do s -ésimo participante, e ε_{ij} é o resíduo do modelo.

Assim como no planejamento do experimento anterior, a introdução da variável “participante” tem por objetivo reduzir a Soma dos Quadrados do Erro (SSE). Porém, dado que nem todos os ouvintes podem ser classificados como “ingênuos”, a decisão sobre o efeito dos participantes ser introduzido no modelo como variável fixa ou aleatória depende da variabilidade das médias e intervalos de confiança das respostas dos ouvintes.

Um modelo ANOVA que leve em conta as interações entre os efeitos pode ser formulado como:

$$Y_{i,j,s} = \underbrace{\mu + \tau_i}_{\mu_i} + \beta_j + \gamma_{i,j} + \alpha_s + \delta_{i,s} + \eta_{j,s} + \varepsilon_{i,j,s} \quad (5.27)$$

onde μ é a média geral, τ_i é o efeito do tratamento da i -ésima sala de reprodução na média, β_j é o efeito do j -ésimo azimuth, $\gamma_{i,j}$ é a interação entre os efeitos da i -ésima sala e do j -ésimo azimuth, α_s é o efeito do s -ésimo participante, $\delta_{i,s}$ é a interação entre os efeitos da i -ésima sala e do s -ésimo participante, $\eta_{j,s}$ é a interação entre os efeitos do j -ésimo azimuth e do s -ésimo participante, e $\varepsilon_{i,j,s}$ é o resíduo do modelo.

Os passos da análise são os mesmos descritos no experimento anterior: análise exploratória, análise de variância para testar as hipóteses formuladas, comparações par a par entre os fatores fixos (salas e azimuths), verificação da força das conclusões via p -valores e tamanhos dos efeitos observáveis e, por fim, estabelecimento de uma relação entre os tempos de reverberação e a variável dependente na forma de uma curva de correção de ganho para o modelo de loudness ITU-R.

Verificação preliminar

A verificação preliminar objetivou avaliar a duração da sessão, a configuração experimental e a clareza das instruções da interface de usuário. Uma visão exploratória dos dados é ilustrada na [Figura 5.26a](#) e na [Figura 5.26b](#). Os dados de resposta obtidos na sala real de escuta crítica foram incluídos a título

de comparação com a versão sintetizada da mesma sala. Pequenas diferenças nas medianas e nas faixas interquartis observadas na [Figura 5.26a](#) sugerem que a síntese binauricular da sala de escuta crítica real – onde se dá a maior parte do trabalho – foi realizada com sucesso.

Os resultados deste teste piloto ilustrados em ambas as figuras apontam para uma tendência cuja observação no experimento principal é desejada. A incidência frontal foi sentida de modo menos intenso nas salas mais “secas” (menos reverberantes) do que nas salas mais reverberantes, o que explica o incremento das medianas com os tempos de reverberação na [Figura 5.26a](#). Adicionalmente, pelo aspecto da [Figura 5.26b](#), esta tendência parece ser insensível às variações de azimuth, o que é consistente com as descobertas feitas no experimento de distância com relação ao efeito de reverberação nos ajustes dos participantes se manifestar de forma independente dos efeitos dos itens de programação, das distâncias (e talvez das direções) dadas suas interações não significativas. Outrossim, a noção intuitiva de que os azimuths laterais necessitariam de ajustes maiores do que os azimuths mais próximos da incidência frontal foi observada somente nas salas menos reverberantes. É possível que esta segunda tendência observada seja também afetada pelo crescimento da energia reverberante a cada sala virtual.

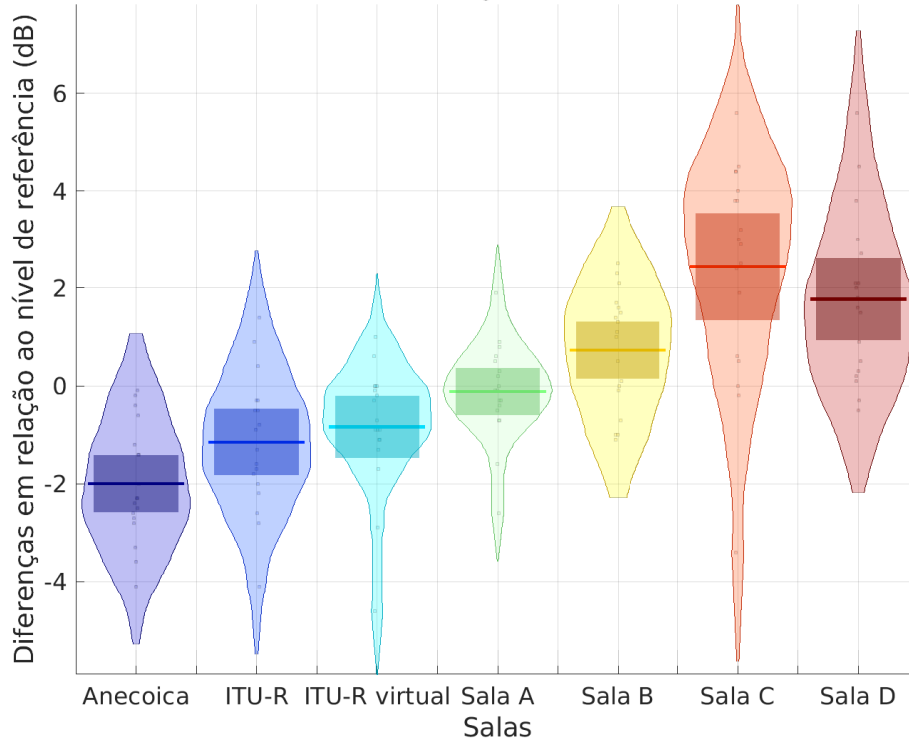
5.4.3 Experimento principal

A coleta de dados deu-se na cabine de áudio do laboratório Cefala na UFMG, cujo piso de ruído é de aproximadamente 33 dBA. Antes de iniciarem o teste perceptivo, os participantes leram uma folha de informações e assinaram um formulário de consentimento, disponíveis para consulta no [Apêndice B](#). Tanto os esclarecimentos fornecidos quanto os consentimentos obtidos foram conformes com os requisitos do Comitê de Ética em Pesquisa da Universidade (Art. 10 da Resolução COEP no. 510/2016). Os níveis de reprodução sonora estiveram dentro dos limites estabelecidos pela prefeitura (Lei Ordinária no. 9.505/2008 do Município de Belo Horizonte) e os dados dos participantes foram processados e guardados conforme a Lei Brasileira de Proteção de Dados (Lei no. 13.709 de 14/08/2018).

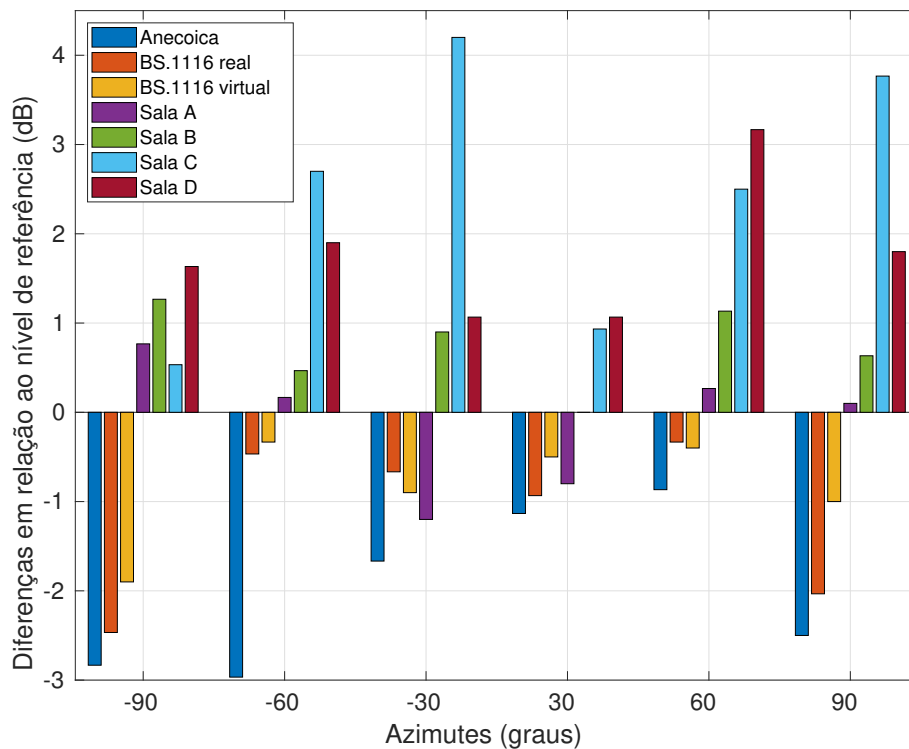
Doze estudantes universitários com acuidade auditiva autodeclarada nor-

Figura 5.26 – Resultados do teste piloto da configuração experimental.

Sensibilidades de loudness nas direções $(\pm 30^\circ, 0^\circ)$, $(\pm 60^\circ, 0^\circ)$ and $(\pm 90^\circ, 0^\circ)$



(a) Diagrama de caixas dos ajustes de nível por salas virtuais de apenas um participante.

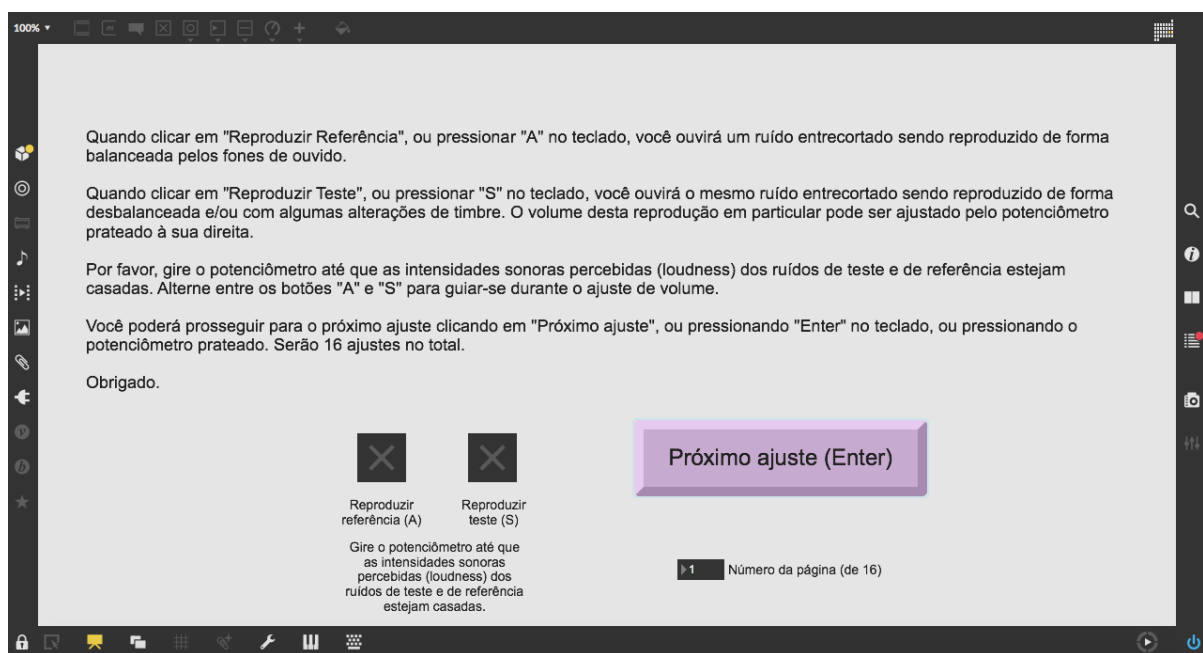


(b) Médias dos ajustes de nível da Figura 5.26a separados por azimute.

Fonte: Elaborada pelo autor.

mal participaram do experimento. Nenhum deles tinha experiência prévia com testes de escuta. Os ouvintes “inexperientes” executaram tarefas de casamento de *loudness* no interior da cabine usando o mesmo par de fones de ouvido da etapa de calibração. Os sinais de teste foram ajustados com o controlador USB não rotulado, conforme instruções dispostas na interface gráfica de usuário. Nenhum indicador visual de volume estava disponível para os participantes para evitar vieses de escala. Uma fotografia da interface de usuário é disposta na [Figura 5.27](#).

Figura 5.27 – Patch de MaxMSP® para o teste de escuta.



Nota – A interface de usuário foi livre de sliders, faders, VUs e qualquer outro indicador de volume para evitar vieses decorrentes de escala.

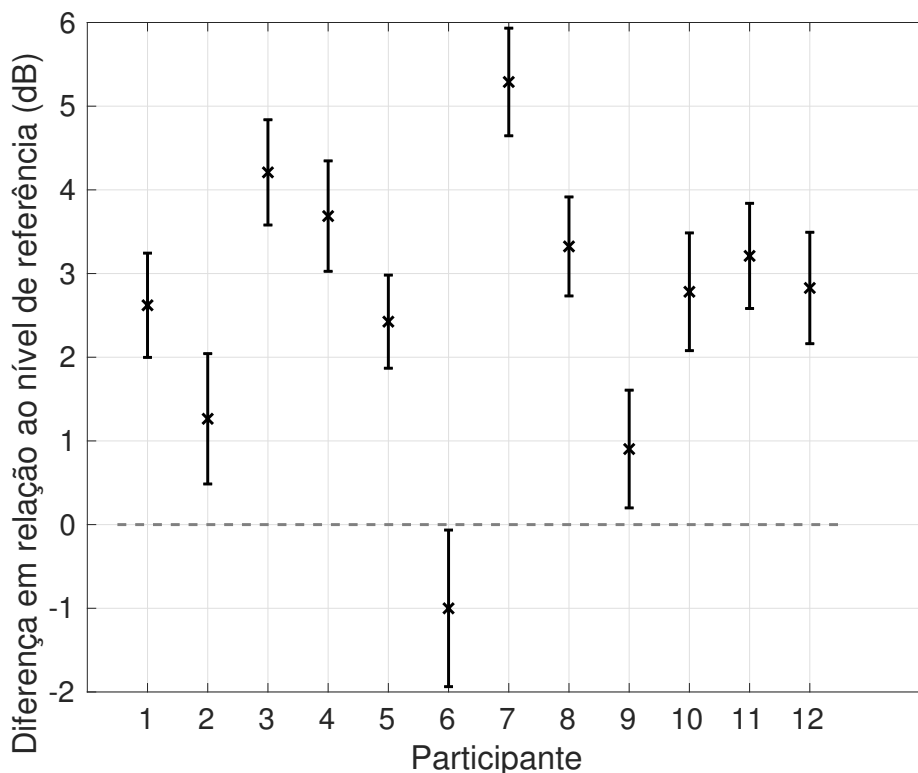
Fonte: Elaborada pelo autor.

Desempenho dos participantes

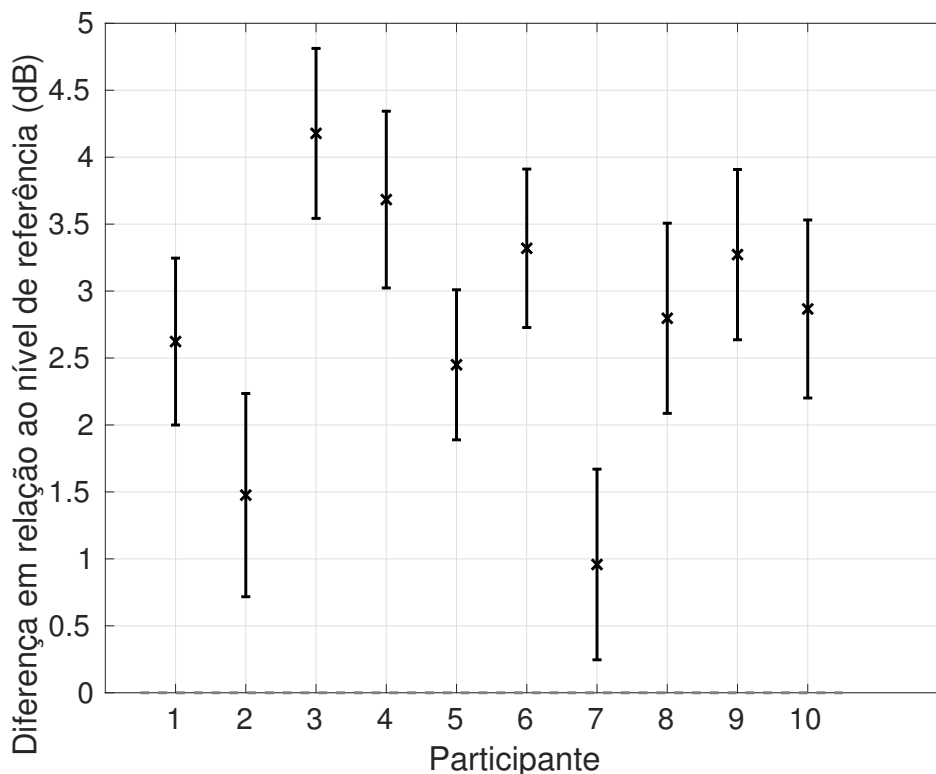
Médias e intervalos de confiança de 95% dos ajustes de nível por participante em todas as salas são dispostos na [Figura 5.28a](#) e na [Figura 5.28b](#). Uma variabilidade mais acentuada do que no experimento anterior era esperada devido às diferenças entre os ouvintes “treinados” dos “inexperientes” e entre as salas reais e virtuais, porém a variabilidade observada foi alta o suficiente para se assumir alguma correlação entre participantes e respostas.

A partir da tendência geral exibida na [Figura 5.28a](#), com base em suas médias distantes e seus intervalos de confiança não sobrepostos com os dos

Figura 5.28 – Médias e intervalos de confiança de 95% dos ajustes de nível executados pelos participantes.



(a) Os participantes de nº 6 e 7 não pareceram entender completamente a tarefa e foram excluídos da avaliação geral.

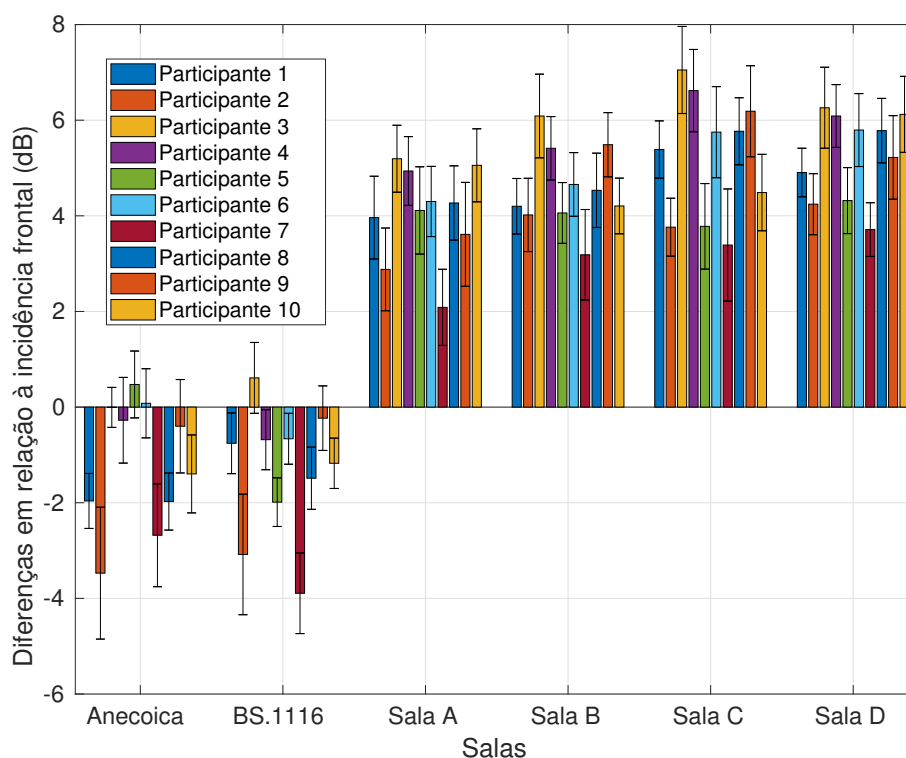


(b) Mesmo depois da remoção dos participantes que fizeram os ajustes mais extremos, a variabilidade geral indica que o efeito “participante” deva ser incluído no modelo linear da ANOVA como uma variável aleatória.

demais, os participantes 6 e 7 foram considerados extremos e então excluídos da análise. O desempenho dos participantes remanescentes está ilustrado na Figura 5.28b.

As mesmas médias agrupadas por sala de reprodução estão ilustradas na Figura 5.29. A variabilidade dos intervalos de confiança dispostos na Figura 5.28b não alterou a tendência geral de respostas com relação às salas de escuta, o que sugere que o efeito do tratamento da reverberação afetou a variável de resposta de modo similar para todos os participantes.

Figura 5.29 – Médias e intervalos de confiança dos participantes agrupados por sala de reprodução.



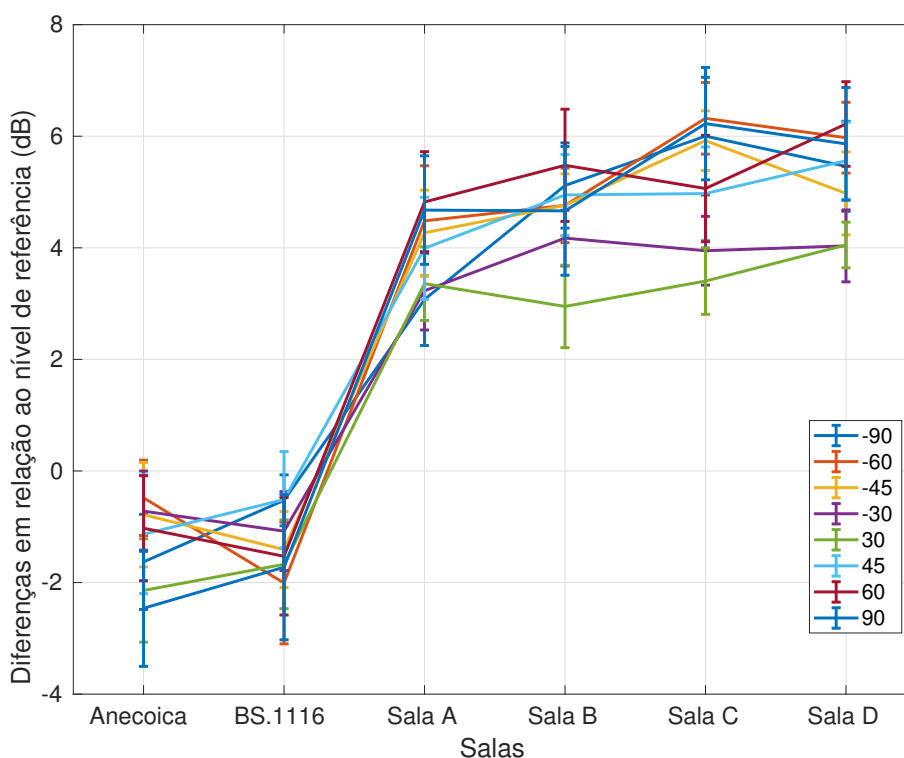
Nota – O efeito dos tempos de reverberação na variável de resposta deu-se de modo similar para todos os participantes.

Fonte: Elaborada pelo autor.

Com um modelo linear composto de somente duas variáveis fixas, faz sentido observar as respostas dos participantes ao longo de todos os níveis destes dois fatores experimentais. Um gráfico em linha caracterizando estas observações é ilustrado na Figura 5.30. Desvios em relação aos níveis de referência da incidência frontal crescendo conforme a lateralização da incidência sonora eram esperados, porém é possível notar que as diferenças de ajustes de nível entre

incidências mais próximas ($\pm 30^\circ$) e mais distantes ($\geq |45^\circ|$) da referência frontal aumentam conforme os tempos de reverberação ficam maiores. Isto implica que a interação entre os efeitos dos tratamentos do “azimute” e da “sala de reprodução” possam ser de tamanho maior do que a interação entre os efeitos da “distância de teste” e da “sala de reprodução” no experimento anterior.

Figura 5.30 – Gráfico em linha das respostas dos participantes por todos os níveis dos fatores experimentais “sala de reprodução” e “azimute”.



Nota – Com tempos de reverberação mais longos, as diferenças de nível entre os azimutes mais próximos e os mais distantes da incidência frontal são crescentes.

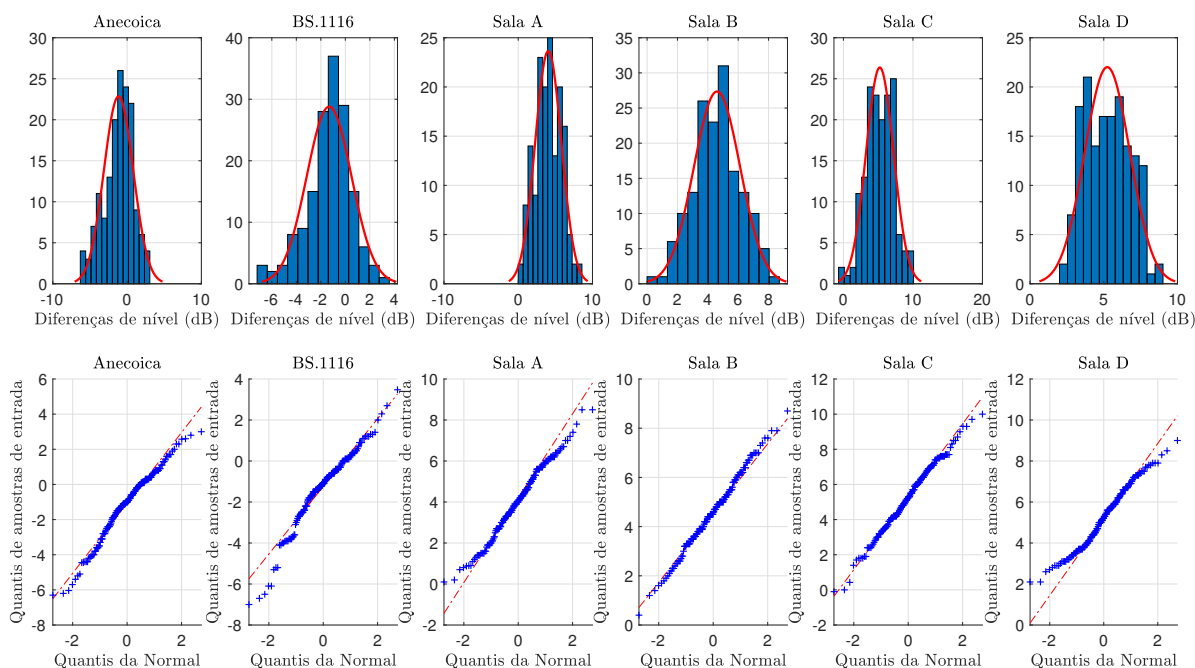
Fonte: Elaborada pelo autor.

Análise exploratória

O total de observações é grande o suficiente para que testes de normalidade como Kolmogorov-Smirnoff e Shapiro-Wilk resultem em significância até mesmo para pequenos desvios (FIELD, 2013). Portanto, os dados gerais podem ser considerados normais para fins práticos. Histogramas e gráficos quantil-quantil dos dados agrupados por salas são ilustrados na Figura 5.31. É possível notar que os desvios de normalidade são facilmente identificados nos gráficos

Q-Q das salas mais “secas” e das salas menos reverberantes mas em padrões diferentes, devido à diferença de curtose entre as distribuições.

Figura 5.31 – Histogramas e gráficos Q-Q para avaliação de normalidade. Os dados estão agrupados por salas sintetizadas.



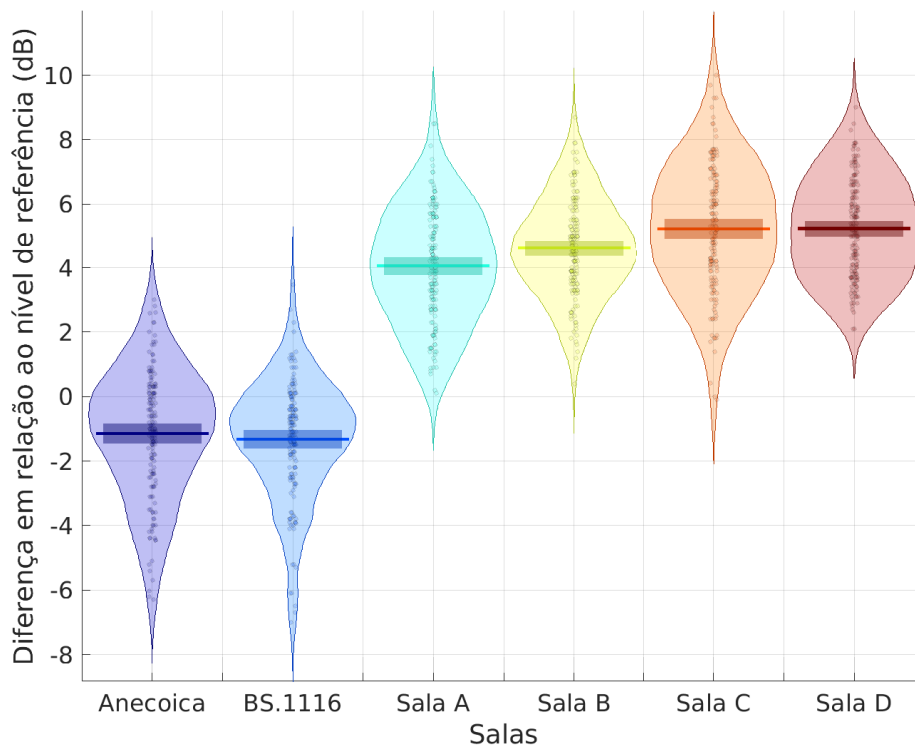
Nota – A maioria dos desvios de normalidade concentram-se nas salas mais “secas” e nas salas mais reverberantes. As distribuições extremas possuem diferenças aparentes de curtose.

Fonte: Elaborada pelo autor.

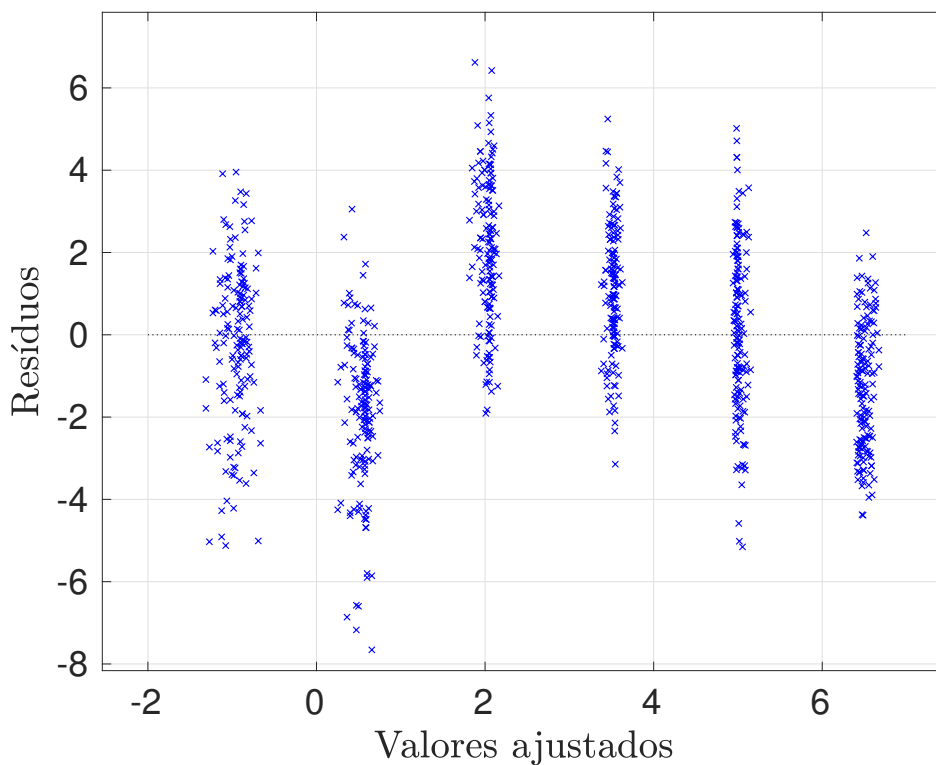
Diagramas de caixa dos ajustes de nível sonoro agrupados por salas de reprodução estão ilustrados na [Figura 5.32a](#). Estes diagramas replicam a mesma tendência observada na [Figura 5.29](#), onde mais intervalos de confiança sobrepostos foram traduzidos para funções de probabilidade mais densas. Note que para as salas virtuais comuns (A, B, C e D), as medianas crescem até certo ponto e as faixas interquartil ficam mais comprimidas, destacando uma tendência à invariância de *loudness* observada no experimento de distância.

Os dados foram então ajustados ao modelo linear formulado na [subseção 5.4.2](#) e um gráfico dos valores ajustados pelos resíduos é ilustrado na [Figura 5.32b](#). O espalhamento dos resíduos para cada valor ajustado sugere normalidade, mas a relação sistemática geral sugere heterogeneidade de variâncias ([FIELD, 2013](#)). A comparação entre várias médias deve ser feita com alguma correção de heteroscedasticidade.

Figura 5.32 – Análise exploratória dos dados obtidos experimentalmente.



(a) Diagrama de caixa dos ajustes agrupados por fator “sala de reprodução”. Tempos de reverberação mais longos concentram funções de probabilidade mais densas.



(b) Gráfico dos valores ajustados pelos resíduos do modelo linear formulado na [subseção 5.4.2](#). O espalhamento dos resíduos para cada um dos valores ajustados sugere normalidade, mas a relação sistemática geral sugere heteroscedasticidade.

Estatística descritiva

Com relação ao tratamento de *outliers*, os dados foram separados em grupos por combinação azimute / sala virtual e então tratados em três rodadas. Primeiro, dois participantes que destoavam do restante dos ouvintes na amostra tiveram suas respostas excluídas da análise (ver [Figura 5.28b](#)). Então os ajustes cujas diferenças em relação ao nível de referência foram duas ou mais vezes maiores do que o desvio padrão estimado dos ajustes de nível, foram considerados extremos e foram removidos (*trimmed*). Os *outliers* remanescentes nestes grupos foram recodificados (*winsorized*) pelos maiores valores não-*outliers*. O tratamento de *outliers* resultou num conjunto de dados desbalanceado com uma tabela de frequências da forma:

	Az-90	Az-60	Az-45	Az-30	Az30	Az45	Az60	Az90
anechoic	20	20	20	19	20	19	20	19
roomA	20	20	20	20	19	20	20	20
roomB	20	19	20	19	19	19	20	20
roomC	19	20	20	20	20	20	19	20
roomD	20	20	20	19	19	20	20	19
BS1116	20	20	20	20	19	20	20	20

Anteriormente, comentou-se a não necessidade de testes estatísticos para assegurar normalidade para um conjunto grande de dados. Mas quando os dados são separados por grupos, estes testes podem ser usados para observar a normalidade intra-grupos. As distribuições por sala de reprodução são na sua maioria assimétricas negativamente, isto é, valores frequentes da variável de resposta estão concentrados numa faixa de valores mais alta que a média da distribuição. As distribuições dos dados da sala pseudoanecoica e das salas B, C e D são moderadamente assimétricas à direita ($-1 < \text{assimetria} < -0,5$), e as distribuições da sala de escuta crítica e da sala A são aproximadamente simétricas ($-0,5 < \text{assimetria} < 0$).

Todas as salas com exceção da sala C apresentaram distribuições leptocúrticas leves ($-1 < \text{curtose} < 0$). Em paralelo, testes de normalidade Kolmogorov-Smirnoff ajustados (testes de Lilliefors) por sala de reprodução rejeitaram a hipótese nula de distribuição normal a um intervalo de confiança de 95%: i)

de modo significativo na câmara pseudoanecoica e na sala D de seminários e apresentações, e ii) de modo muito significativo na sala de escuta crítica em conformidade com a Rec. ITU-R BS.1116.

Estatística inferencial

O teste de Levene para os dados gerais rejeitou a hipótese nula de homoscedasticidade de modo muito significativo ($p = 0,0057$). Transformações de dados do tipo \log , $1/x$, $\sqrt{(x)}$ e Fisher- z também violaram esta premissa para a análise padrão de variâncias. Os dados foram então ajustados ao modelo linear de efeitos mistos de duas vias da [Equação 5.27](#), mas com matrizes de covariância com correções “White-Huber” de heteroscedasticidade (LONG; ERVIN, 2000).

O teste resultou num efeito muito significativo do fator “sala de reprodução” e de tamanho muito grande [$F_{(5,898)} = 100,32$, $p < 0,001$, $\omega^2 = 0,721$]. Por outro lado, o efeito do fator “azimute” foi observado como sendo significativo, porém com um tamanho de efeito muito menor [$F_{(7,898)} = 2,88$, $p = 0,006$, $\omega^2 = 0,018$]. A interação entre os efeitos dos dois fatores experimentais foi muito significativa, porém com um tamanho de efeito muito menor que o do fator “sala de reprodução” [$F_{(35,898)} = 3,86$, $p < 0,001$, $\omega^2 = 0,023$]. A significância de ambos os efeitos e sua interação já era esperada com base na análise exploratória, mas a comparação entre as médias ao longo de todos os níveis dos fatores experimentais atesta para a dominância dos tempos de reverberação das salas sobre os azimutes (e sobre suas interações) neste experimento.

Comparações par-a-par entre os fatores experimentais ao longo de seus diferentes níveis foram consistentes com os tamanhos de efeito observados. Para o fator “sala de reprodução”, apenas as salas com tempos de reverberação próximos aos extremos foram observadas com diferenças não significativas na variável de resposta: BS.1116 vs. anecoica ($p = 0,734$) e sala C vs. sala D ($p = 0,133$). Já para o fator “azimute”, muito poucas comparações reportaram diferenças significativas na variável de resposta: $+90^\circ / -60^\circ$ ($p = 0,0180$), $+90^\circ / -45^\circ$ ($p = 0,0327$), $+90^\circ / -30^\circ$ ($p = 0,0159$) e $+90^\circ / +60^\circ$ ($p = 0,0375$).

5.4.4 Modelo de loudness ITU-R como função da reverberação

Curva de correção de ganho

Muito embora uma interação entre reverberação e azimuth tenha sido observada de modo muito significativo, o efeito da sala de reprodução teve um *F-score* e um tamanho de efeito bem maiores se comparados aos do efeito do azimuth e interações. Logo, é razoável considerar uma curva de correção baseada somente nos tempos de reverberação e nas respostas dos participantes. Pares de dados da forma (tempos de reverberação, médias dos participantes) foram ajustados a uma interpolação *spline* cúbica, resultando numa função polinomial por partes da forma:

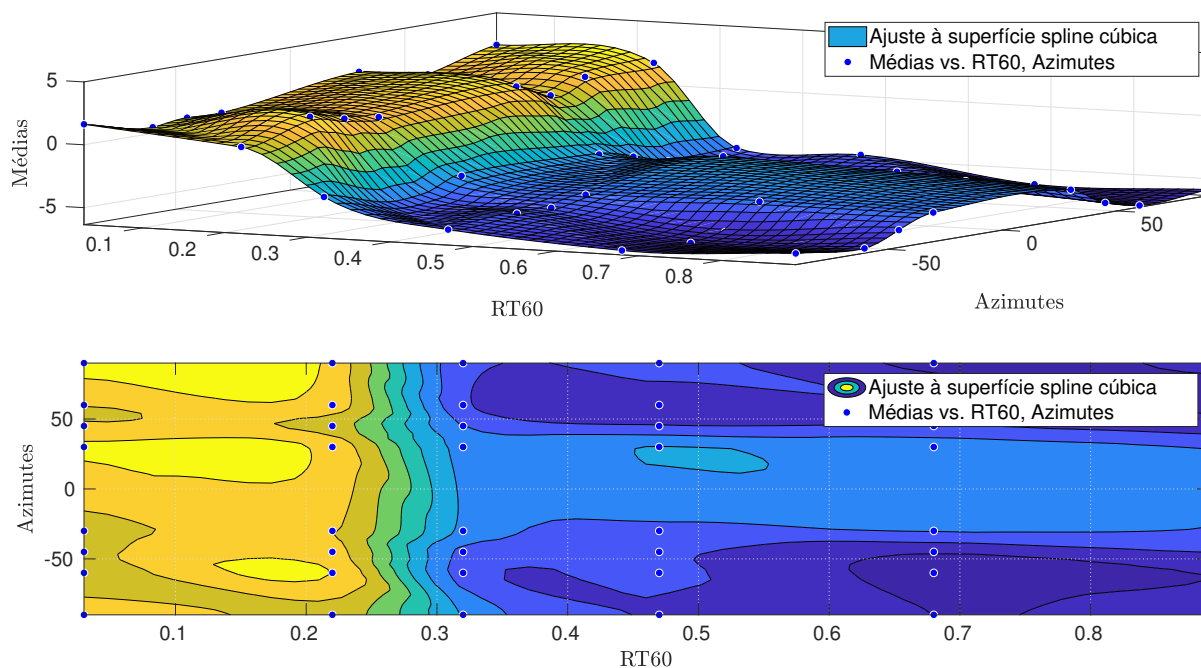
$$\begin{cases} 202,7089t^3 + 0,9983t - 1,5823, & 0,03 \leq t < 0,22, \\ -710,9374t^3 + 115,5441t^2 + 22,9517t - 0,0023, & 0,22 \leq t < 0,32, \\ 159,7207t^3 + 97,7371t^2 + 24,7324t + 2,7374, & 0,32 \leq t < 0,47, \\ 37,1978t^3 - 25,8628t^2 + 6,1924t + 4,7872, & 0,47 \leq t < 0,68, \\ 3,8542t^3 - 2,4281t^2 + 0,2513t + 5,2916, & 0,68 \leq t < 0,89, \end{cases} \quad (5.28)$$

onde t é o tempo de reverberação em segundos. Um conjunto de métricas de qualidade de ajuste retornou os seguintes resultados: $SSE = 3,3927$, $R^2 = 0,9327$, $Adj. R^2 = 0,7375$ e $RMSE = 1,6266$.

No espírito desta investigação, ao se considerar também o fator azimuth na correção, tem-se a superfície de ganho ilustrada na [Figura 5.33](#), a partir da qual se observa quão menor foi a influência geral do azimuth nas respostas dos participantes se comparada à influência dos tempos de reverberação.

Assim como na correção de ganho baseada em distância, não é o caso aqui de se corrigir os patamares dos filtros de sombreamento de cabeça e RLB porque a largura de faixa não foi um fator avaliado experimentalmente aqui. A correção deve então ser feita no nível geral e para isso a curva de ganho é posicionada antes da filtragem K , como ilustrado anteriormente na [Figura 5.15](#).

Figura 5.33 – As médias dos participantes como função dos tempos de reverberação e dos azimutes foram ajustadas a uma superfície *spline* cúbica.



Nota – Pelo gráfico de contorno, é possível notar que a influência da variação de azimuth é bem menor que a dos tempos de reverberação, o que é consistente com os tamanhos de efeito observados na análise de variâncias.

Fonte: Elaborada pelo autor.

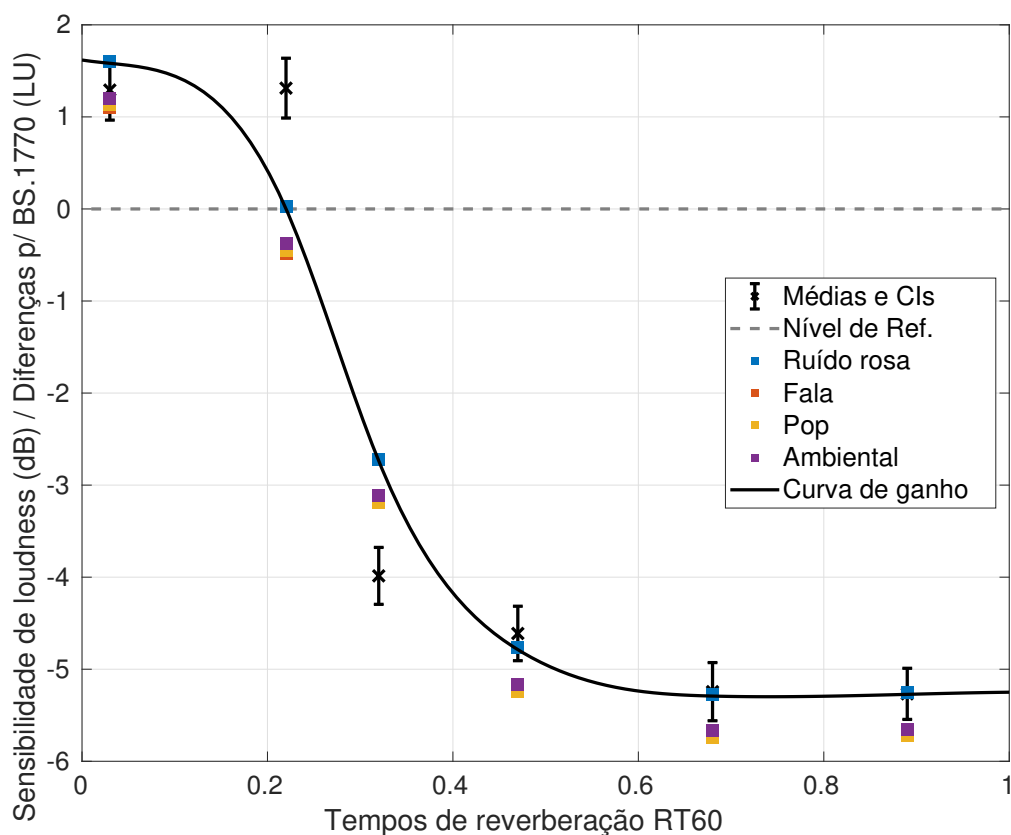
Medidas objetivas

As diferenças entre as medidas de *loudness* feitas pelo modelo ajustado e pelo algoritmo BS.1770 são exibidas na Figura 5.34. Esta avaliação foi feita utilizando-se os mesmos itens de programação do experimento anterior (“ruído”, “fala”, “pop” e “ambiental”), agora auralizados em relação a cada uma das salas de reprodução virtuais.

A variável de resposta não é expressa aqui pelos ajustes de nível realizados pelos participantes, mas sim pelas suas sensibilidades, que podem ser interpretadas como sendo o quanto um sinal sob teste com o mesmo nível do sinal de referência deva ser ajustado para ser percebido como tendo o mesmo *loudness* do sinal de referência. Este é o racional para a operação desejada da curva de correção de ganho.

Gráficos análogos das medidas e das médias dos participantes agrupadas

Figura 5.34 – Diferenças entre as medidas de *loudness* do modelo ajustado e o algoritmo BS.1770, sobrepostas à curva de correção de ganho e às médias de sensibilidade dos participantes.

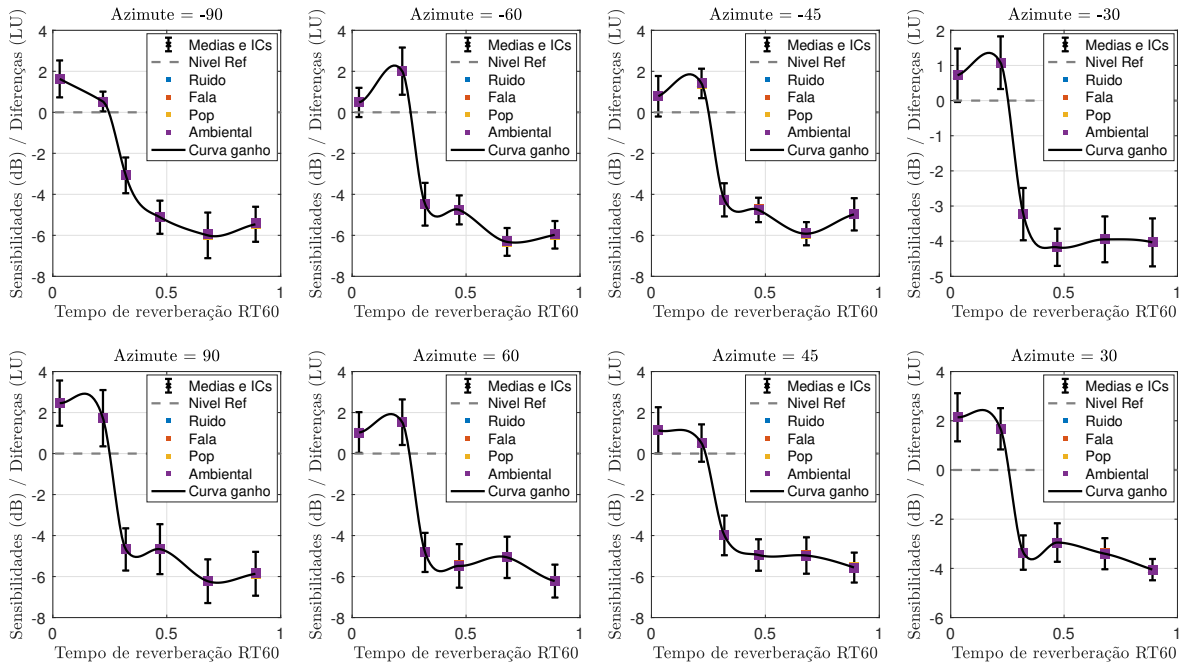


Fonte: Elaborada pelo autor.

por azimute, sobrepostas às curvas de correção de ganho correspondentes derivadas a partir da superfície de ajuste da [Figura 5.33](#), são dispostos na [Figura 5.35](#). Note que a variação de azimute teve pouca influência no aspecto geral das curvas, corroborando o tamanho de efeito muito maior do fator experimental “sala de reprodução”.

Nesta seção, a influência da energia reverberante foi melhor observada em condições experimentais distintas da [seção 5.3](#), nas quais o fator experimental “sala de reprodução” esteve restrito a dois níveis. Esta limitação logística foi superada com a auralização dos estímulos por síntese biauricular feita a partir de respostas ao impulso de salas virtuais, com tempos de reverberação diferentes. Um relato destes resultados foi submetido à apreciação do comitê científico do XXXVII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrT 2019), pendente de publicação até a data desta redação.

Figura 5.35 – Diferenças entre as medidas de *loudness* do modelo ajustado e o algoritmo BS.1770, sobrepostas à curva de correção de ganho e às médias de sensibilidade dos participantes, agrupadas pelos azimutes testados.



Fonte: Elaborada pelo autor.

Além disso, foi observado um efeito direcional significativo na variável de resposta. E apesar de este efeito ter sido de tamanho bem menor que o da energia reverberante, sua presença foi um motivador para se dar continuidade à investigação do efeito direcional na sensação de *loudness*, que, como visto na [subseção 5.2.2](#), é um efeito de especial interesse para o cômputo do *loudness* de objetos sonoros. Esta investigação é o objeto da próxima seção.

5.5 Relação entre *Loudness* e Direção

Na [subseção 5.2.2](#), estabeleceu-se que as investigações da dependência do *loudness* das coordenadas espaciais de um objeto sonoro deve focar em dois objetivos práticos: propor modificações na filtragem K com base nos resultados dos testes subjetivos da relação com a distância, e modificações nos ganhos G_{N_s} baseadas em resultados de testes subjetivos da relação com o azimute e a elevação. Nas seções anteriores conseguiu-se obter curvas de correção para o filtro K com base nas respostas dos participantes em experimentos avaliando

distância e reverberação.

Deste ponto em diante, o foco se volta para o segundo objetivo: investigar o que há para ser feito quanto ao *loudness* de objetos sonoros com localizações específicas. Para tal, faz-se um exame dos passos empregados para a obtenção da ponderação direcional do modelo ITU-R de modo a entendê-la, contemplar possíveis lacunas, formular as hipóteses e planejar o experimento de acordo. Contudo, não sem antes atentar para as seguintes palavras de Georg von Békésy:

Uma quantidade considerável de trabalho científico consiste na repetição de experimentos anteriormente realizados por outros. Essa é uma tarefa particularmente onerosa, em parte porque as condições do experimento original não estão descritas na íntegra e porque a motivação é baixa. O trabalho original foi feito com cuidado e entusiasmo, mas sua repetição pode ser monótona e pouco inspiradora. (BÉKÉSY; WEVER, 1960, p. 7, tradução minha)

Ter este trecho do clássico *Experiments in Hearing* em mente, gerou uma preocupação quanto à pesquisa não cair nesse tipo de armadilha. Espera-se que a adição de novos fatores, procedimentos e estímulos, contribua para não rotular o experimento descrito nesta seção como um caso de mera reprodutibilidade. Deixo este juízo para o leitor.

5.5.1 Loudness direcional no modelo ITU-R

Os ganhos direcionais do modelo de *loudness* do ITU-R foram atualizados em sua versão de 2015 para um número irrestrito de canais. Estas atualizações foram propostas por Komori *et al.* (2015), afiliados à emissora NHK do Japão. Embora o artigo não entre em detalhes sobre como esses cálculos foram feitos, estes podem ser rastreados pelas referências apresentadas abaixo.

A partir dos experimentos de *loudness* direcional feitos por Robinson e Whittle (1960) com medidas de pressão sonora em campo livre e nas orelhas dos ouvintes, os autores desenvolveram uma expressão para o somatório biauricular de *loudness* com um ganho biauricular de 6 dB (ver Equação 2.24) da forma

$$L_{mon} = 6 \times \log_2 \left(2^{\frac{L_{esquerdo}}{6}} + 2^{\frac{L_{direito}}{6}} \right), \quad (5.29)$$

onde L_{mon} é a pressão sonora equivalente necessária para uma estimulação monóptica casada com qualquer combinação biauricular, diótica ($L_{esquerdo} = L_{direito}$)

ou dicótica ($L_{esquerdo} \neq L_{direito}$), dos níveis de pressão sonora incidentes no ouvido esquerdo ($L_{esquerdo}$) e no ouvido direito ($L_{direito}$).

Sivonen e Ellermeier (2006) reproduziram o experimento anterior levando as HRTFs em consideração. O ganho do somatório biauricular de *loudness* g foi estimado via minimização da soma dos quadrados do erro (SSE) entre as Sensibilidades Direcionais de *Loudness* (DLS) medidas e as sensibilidades calculadas a partir das variações entre os níveis de pressão sonora nos ouvidos (L_{mon}), para $i = 4$ ângulos de incidência no plano horizontal ($30^\circ, 60^\circ, 90^\circ$ e 135°) e $j = 16$ repetições, usando a equação a seguir

$$SSE = \sum_{i=1}^4 \sum_{j=1}^{16} \left\{ DLS_{i,j} - \left[L_{mon,comp_i}(g) - L_{mon,ref}(g) \right] \right\}^2, \quad (5.30)$$

onde

$$L_{mon,comp_i}(g) = g \times \log_2 \left(2^{\frac{L_{esquerdo,comp_i}}{g}} + 2^{\frac{L_{direito,comp_i}}{g}} \right), \quad (5.31)$$

e

$$L_{mon,ref}(g) = g \times \log_2 \left(2^{\frac{L_{esquerdo,ref}}{g}} + 2^{\frac{L_{direito,ref}}{g}} \right). \quad (5.32)$$

$L_{esquerdo,comp_i}$ e $L_{direito,comp_i}$ referem-se aos níveis da incidência lateral sob comparação, calculados a partir de HRTFs para os ouvidos esquerdo e direito, enquanto $L_{esquerdo,ref}$ e $L_{direito,ref}$ referem-se aos níveis nos ouvidos esquerdo e direito correspondentes à incidência frontal.

Komori *et al.* (2015), com base nos experimentos de Robinson e Whittle (1960) e Sivonen e Ellermeier (2006), calcularam as ponderações direcionais dos canais na Tabela 5.9. Considerando o efeito da direção no *loudness* ser menor em salas comuns do que numa câmara anecoica e a garantia de compatibilidade reversa, os autores escolheram normalizar os resultados a 1,5 dB e aproximá-los a passos também de 1,5 dB (ITU-R, 2014c). Os ganhos direcionais decididos pelo grupo relator são os constantes da Tabela 5.10.

O estudo de Sivonen e Ellermeier (2006) resultou num ganho biauricular $g \approx 3$ dB para ouvintes experientes, que deu seguimento a um novo estudo com um resultado de $g \approx 6$ dB para ouvintes inexperientes (SIVONEN; ELLERMEIER, 2008). A diferença foi comentada pelos autores:

Note que quanto maior o somatório, maior é o efeito do ouvido com menor SPL no *loudness* biauricular. Portanto, para amostra de ouvintes

Tabela 5.9 – Ganhos de somatório biauricular de *loudness* calculados em contribuição da NHK para o grupo de trabalho de *loudness* do ITU-R (2014c) e os pesos direcionais propostos por Komori et al. (2015).

Ângulo de azimute (θ)	0°	$\pm 30^\circ$	$\pm 60^\circ$	$\pm 90^\circ$	$\pm 110^\circ$	$\pm 135^\circ$	180°
Níveis calculados (dB)	0,00	1,36	4,47	5,22	4,46	0,84	-8,25
Níveis normalizados (dB)	0,00	0,39	1,29	1,50	1,28	0,24	-2,37
Pesos propostos (dB)	0,00	0,00	1,50	1,50	1,50	0,00	-1,50

Fonte: Adaptada de ITU-R (2014c).

Tabela 5.10 – Ponderação de canais dependente da posição na versão de 2015 do modelo de *loudness* ITU-R

Elevação (ϕ)	Azimute (θ)		
	$ \theta < 60^\circ$	$60^\circ \leq \theta \leq 120^\circ$	$120^\circ < \theta \leq 180^\circ$
$ \phi < 30^\circ$	1,00 (± 0 dB)	1,41 (+ 1,50 dB)	1,00 (± 0 dB)
demais	1.00 (± 0 dB)		

Fonte: Adaptada de ITU-R (2015b).

inexperientes, o ouvido (direito) com o menor SPL incidente “puxa para baixo” a curva de *loudness* direcional um pouco mais se comparado com a amostra de ouvintes experientes (...). Segundo a análise estatística, a pequena diferença entre grupos do ganho biauricular (1,7 dB) (...) é mais provável ser devida ao acaso (SIVONEN; ELLERMEIER, 2008, p. 458, tradução minha).

Não ficou claro em que medida os ganhos estimados foram afetados por diferentes condições em cada experimento: HRTFs personalizadas em (SIVONEN; ELLERMEIER, 2006) vs. medidas com HATS em (SIVONEN; ELLERMEIER, 2008), além de ouvintes com diferentes perfis. Porém, as diferentes sensibilidades provocadas pelos mesmos níveis de pressão sonora incidentes nos ouvidos sugerem que somatórios biauriculares de *loudness* são feitos de maneira individual, e um ganho geral pode ser estimado com uma faixa mais ampla de ouvintes de diferentes categorias.

Muito embora esses trabalhos de 2006 e 2008 tenham contemplado direções no plano médio, o objetivo principal foi estimar um ganho biauricular para ser aplicado na fórmula de Robinson (Equação 5.29) para estimar um nível de *loudness* monauricular a partir de sons dicóticos nos ouvidos. Se a extensão do modelo ITU-R para um número irrestrito de canais foi baseada nas suas descobertas, não é surpreendente que os pesos propostos tenham sido calculados

considerando tão somente o plano horizontal.

Adicionalmente, os estímulos dos testes subjetivos nos trabalhos de Sivonen e Ellermeier foram tons monauriculares filtrados em 1/3 de oitava, e nos trabalhos da NHK foram itens de programação multicanal para configurações até 22 canais. Apesar do fato de estímulos de faixa estreita serem mais sensíveis à direção (como verificado por Shao, Mo e Mao (2015) e observado na [subseção 5.3.1](#)), materiais de programação para a radiodifusão são de faixa larga por natureza, embora um teste de escuta com reais programações multicanal não aborde propriamente o efeito de uma fonte de áudio posicionada em localizações específicas. Além do mais, fontes-imagem geradas por balanceamento de alto-falantes – ou fontes fantasmas – são elementos importantes da experiência de *home audio* que carecem de maiores investigações.

Para tanto, faz-se necessário conduzir um experimento com pessoas executando tarefas de casamento de *loudness* em objetos de áudio situados a diferentes localizações numa sala de escuta crítica. Os dados do teste de escuta serão então utilizados para estimar um ganho biauricular e subsequentemente calcular novos pesos de ponderação direcional para o modelo ITU-R como funções de azimute e elevação.

Pergunta de pesquisa e hipóteses a testar

Ao se destacar o efeito da direção na sensação de *loudness*, a pergunta de pesquisa pode ser feita da forma:

- *Como a sensação de loudness é afetada pela variação de azimutes e elevações de uma fonte sonora?*

E as hipóteses orientadas à direcionalidade ficam da forma:

- Hipótese nula: *Sons reproduzidos por fontes sonoras físicas ou fantasmas, situadas a diferentes azimutes e/ou elevações, que tenham o mesmo nível de loudness medido na posição do ouvinte, provocarão a mesma sensação de loudness.*

- Hipótese alternativa: *Sons reproduzidos por fontes sonoras físicas ou fantasmas, situadas a diferentes azimutes e/ou elevações, que tenham o mesmo nível de loudness medido na posição do ouvinte, provocarão diferentes sensações de loudness.*

Note que as hipóteses formuladas contemplam fontes reais de direção definida e “fantasmas”, isto é, cuja direção é gerada pelo balanceamento de alto-falantes vizinhos. A próxima subseção descreve os procedimentos e métodos usados nesta investigação.

5.5.2 Projeto de experimento

Testes piloto devem ser conduzidos antes do experimento principal para obtenção de alguns *insights* sobre as questões apresentadas. A seguir serão relatados o projeto de experimento e as verificações preliminares.

Estímulos

O estímulo base (ruído rosa entrecortado e limitado em faixa) e o processo de calibração são os mesmos usados nos experimentos anteriores. Quanto à apresentação, os estímulos serão apresentados em pares: um som de incidência frontal e um som de comparação incidindo de uma direção específica no espaço. A direção de incidência frontal corresponde ao par de coordenadas de azimute e elevação ($\theta = 0^\circ, \phi = 0^\circ$). Primeiramente, as diferentes direções incluem pares de coordenadas fixas correspondentes às posições individuais dos alto-falantes num sistema de 22 canais, mais fontes fantasmas situadas às direções $(45^\circ, 0^\circ)$, $(120^\circ, 0^\circ)$, $(-45^\circ, 0^\circ)$, $(-120^\circ, 0^\circ)$, $(45^\circ, 15^\circ)$, $(120^\circ, 15^\circ)$, $(-45^\circ, 15^\circ)$ e $(-120^\circ, 15^\circ)$. Todos os estímulos foram sintetizados no *MATLAB*[®] com o método de balanceamento de alto-falantes denominado Sistema Vetorial de Panorama por Amplitude (VBAP) (PULKKI, 1997).

Sistema de reprodução

Muito foi dito sobre a sala de escuta crítica do Instituto de Gravação Sonora da Universidade de Surrey, mas pouco sobre seu sistema de reprodução de 22 canais, a ser usado pela primeira vez neste experimento. Este sistema é o

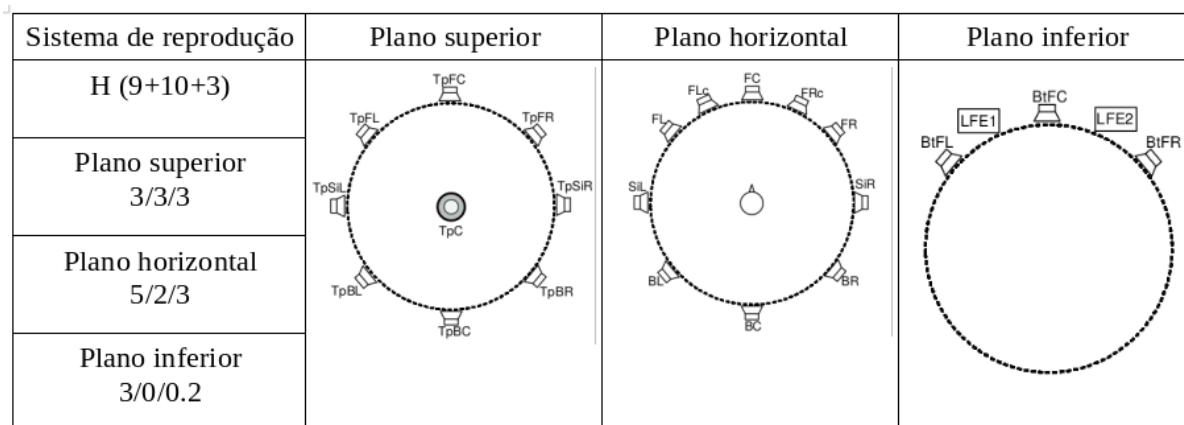
posicionamento identificado como “H” na Recomendação BS.2051 do ITU-R (2014a) para sistemas avançados de som. O plano superior tem 9 alto-falantes, sendo 3 frontais, 3 no plano do ouvinte e 3 traseiros; o plano horizontal possui 5 alto-falantes frontais, 2 laterais e 3 traseiros; e o plano inferior conta com 3 alto-falantes frontais mais os 2 *sub-woofers*, que não serão usados na avaliação pois os canais LFE não entram no cômputo do *loudness*, como visto na subseção 3.3.3 e na subseção 5.2.1. Uma relação dos alto-falantes e seus azimutes e elevações é disposta na Tabela 5.11 e o detalhamento dos planos superior, horizontal e inferior é ilustrado na Figura 5.36. Uma fotografia panorâmica da sala de reprodução com as posições físicas dos alto-falantes é ilustrada na Figura 5.37.

Tabela 5.11 – Azimutes e elevações dos alto-falantes que integram o sistema de reprodução 22.2 da sala de escuta crítica.

Azimute θ ($^{\circ}$)	Elevação ϕ ($^{\circ}$)	Rótulo ITU-R
-45	-30	B-045
0	-30	B+000
45	-30	B+045
-135	0	M-135
-90	0	M-090
-60	0	M-060
-30	0	M-030
0	0	M+000
30	0	M+030
60	0	M+060
90	0	M+090
135	0	M+135
180	0	M+180
-135	30	U-135
-90	30	U-090
-45	30	U-045
0	30	U+000
45	30	U+045
90	30	U+090
135	30	U+135
180	30	U+180
0	90	T+000

O sistema é ligado a uma interface de áudio *MOTU 24Ao* responsável pelo roteamento dos canais também instalada na sala. Cada alto-falante possui três

Figura 5.36 – Planos superior, horizontal e inferior do sistema de reprodução “H” de 22.2 canais, a ser usado neste experimento.



Fonte: Adaptada de ITU-R (2014a, p. 10).

Figura 5.37 – Fotografia panorâmica da sala de escuta crítica capturando os alto-falantes dos planos superior, horizontal e inferior.

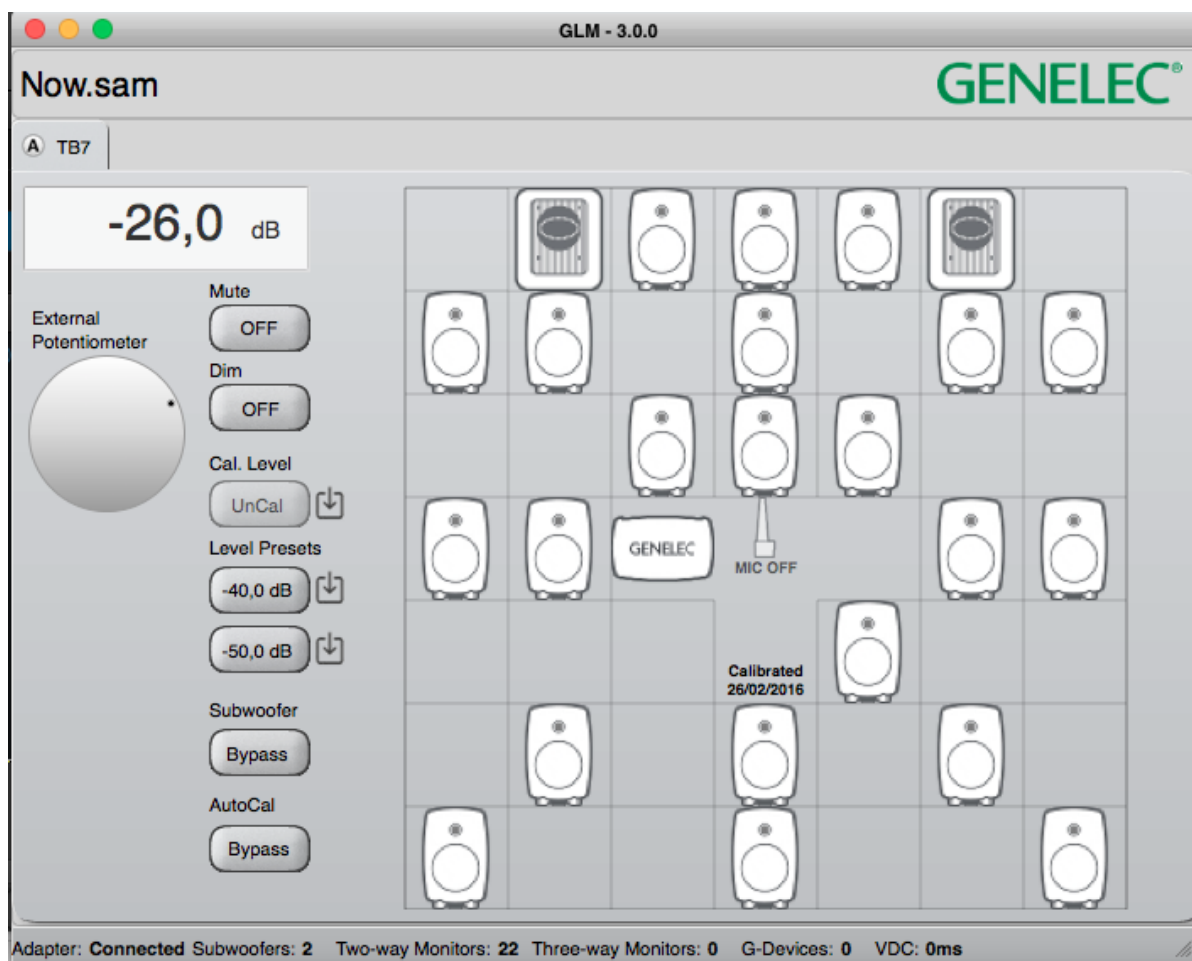


Fonte: Elaborada pelo autor.

conexões: áudio (cabo XLR vindo da interface), energia (proveniente de tomadas rotuladas nas paredes e no chão) e controle (do adaptador de rede *Genelec*). As conexões de controle servem para alinhamento do sistema via dispositivo de gerenciamento de alto-falantes do fabricante. O alinhamento era feito com ruídos rosas de 22 canais girando vertical e horizontalmente via VBAP, carregando-se o arquivo de calibração do sistema em seguida. Uma fotografia do *software* de alinhamento é exibida na [Figura 5.38](#).

Variável de resposta

O formato das respostas dos participantes dá-se na forma de Sensibilidade Direcional de *Loudness* (DLS), que é a diferença de nível entre o som frontalmente incidente (referência) e o som não-frontalmente incidente (teste). Na sua

Figura 5.38 – Fotografia da tela do *software* de operação do *Genelec Loudspeaker Management (GLM)*

Fonte: Elaborada pelo autor.

definição original por [Sivonen e Ellermeier \(2006\)](#), foi incluída a expressão “no Ponto de Igualdade Subjetiva (PSE) em *loudness*” porque o método Escolha Forçada de Duas Alternativas (2AFC) fora usado com a justificativa de prevenir vieses dos próprios participantes, como ajustes influenciados por expectativas ou familiaridade. Por outro lado, enquanto o PSE nos trabalhos de Sivonen e Ellermeier de 2006 e 2008 foi alcançado numa faixa de 1 dB, diferenças práticas em passos de $\pm 0,1$ dB são viáveis no método de ajuste. Dado que pela natureza dos estímulos neste experimento são esperados pequenos tamanhos de efeito, é melhor que o mínimo tamanho de efeito observável seja também menor. Com isso em mente, o casamento de *loudness* neste experimento será executado pelos participantes em tarefas baseadas no método de ajuste. Um total de 29 direções testadas (21 físicas e 8 fantasmas) com três repetições cada resultariam, a princípio, em 87 ajustes por participante para serem executadas em sessões com

duração de 40 minutos a uma hora.

Equipamentos

Os equipamentos usados nesta série de medidas são os relacionados abaixo:

- Vinte e dois alto-falantes *Genelec 8330A* para reprodução de estímulos (os dois sub-woofers *Genelec 7350A* não são usados);
- Um Simulador de Cabeça e Torso (HATS) *Cortex Mk2* para calibração de sinais de referência;
- Uma interface de áudio de dois canais *Focusrite* para calibração e gravação dos sinais biauriculares do HATS;
- Uma interface de áudio *MOTU 24Ao* para o sistema de 22 canais;
- Um controlador USB *Griffin Powermate* para ajustes de nível de áudio digital;
- Um *MacBook Pro* executando *MaxMSP[®]*, *MATLAB[®]* and *Audacity*;
- Um segundo *MacBook Pro* posicionado fora da sala de escuta crítica para operar remotamente o primeiro.

Modelo linear

Para cada grupo dividido entre fontes reais e fantasmas, sejam $Y_{i,s}$ as observações das sensibilidades direcionais de *loudness* para cada i -ésima direção, e s -ésimo participante. O modelo linear da análise de variâncias (ANOVA) pode ser escrito da forma:

$$Y_{i,s} = \underbrace{\mu + \tau_i}_{\mu_i} + \alpha_s + \varepsilon_{i,s} \quad (5.33)$$

onde μ é a média geral, τ_i é o tratamento do efeito da i -ésima direção na média, α_s é o efeito do s -ésimo participante, e $\varepsilon_{i,s}$ é o resíduo do modelo.

Um modelo ANOVA que leve em consideração as interações entre efeitos pode ser formulado como:

$$Y_{i,s} = \underbrace{\mu + \tau_i}_{\mu_i} + \alpha_s + \beta_{i,s} + \eta_{j,s} + \varepsilon_{i,j,s} \quad (5.34)$$

onde μ é a média geral, τ_i é o tratamento do efeito da i -ésima direção na média, α_s é o efeito do s -ésimo participante, $\beta_{i,s}$ é a interação entre os efeitos da direção e do participante, e ε_{ij} é o resíduo do modelo.

Os passos da análise são os mesmos descritos no experimento anterior: análise exploratória, análise de variância para testar as hipóteses formuladas, comparações par a par entre os fatores fixos (salas e azimutes), verificação da força das conclusões via p -valores e tamanhos dos efeitos observáveis e, por fim, estabelecimento de uma relação entre os azimutes e elevações e a variável dependente na forma de ponderações direcionais para o modelo de *loudness* ITU-R.

A descrição de materiais e métodos apresentada se refere à etapa de planejamento, e correções de curso – reportadas ao longo do texto – foram feitas conforme necessidade.

5.5.3 Verificação preliminar

Para este primeiro passo, os estímulos de ruído rosa entrecortado foram sintetizados com *Audacity*, balanceados por VBAP com MATLAB[®], e reproduzidos pelo sistema de 22 canais na sala de escuta crítica diretamente por cada alto-falante (fonte real) e por um trio de alto-falantes adjacente a uma direção específica (fonte fantasma). Os alto-falantes foram calibrados de tal forma que um estímulo a -23 LKFS reproduzido pelo alto-falante frontal mediu 70 dB SPL nos microfones intra-auriculares do *Cortex MK2 HATS*.

Como visto anteriormente, comentários feitos por [Sivonen e Ellermeier \(2008\)](#) sugeriram que as diferenças de DLS por grupo de participantes estariam ligadas ao efeito do nível de pressão sonora no ouvido da incidência contra-lateral, mas não foram feitas investigações mais aprofundadas a esse respeito. Pareceu uma boa ideia, portanto, procurar por correlações entre os dados experimentais e métricas interauriculares de modo a se buscar suporte para esta alegação.

Uma métrica possível seria a Diferença de Intensidade Interauricular Normalizada (ILD_{Norm}) proposta por Ben [Supper \(2005\)](#). Nesta, a potência total dos sinais incidentes é calculada. A diferença nos valores de potência é então dividida pela soma de potências e o resultado é então elevado ao quadrado com

o sinal preservado:

$$\text{ILD}_{Norm} = \pm \left(\frac{\frac{1}{T} \int_0^T s_{esquerdo}^2(t) dt - \frac{1}{T} \int_0^T s_{direito}^2(t) dt}{\frac{1}{T} \int_0^T s_{esquerdo}^2(t) dt + \frac{1}{T} \int_0^T s_{direito}^2(t) dt} \right)^2, \quad (5.35)$$

onde $s_{esquerdo}(t) = h_{esquerdo}(t) * s(t)$ é o sinal do ouvido esquerdo produto da convolução da função de transferência do sinal incidente para o ouvido esquerdo com o sinal sintetizado originalmente, $s_{direito}(t) = h_{direito}(t) * s(t)$ é o sinal do ouvido direito definido de forma análoga, e T é a duração do sinal. A métrica varia entre $[-1, 1]$. Valores negativos indicam sinais mais intensos no ouvido direito e vice-versa.

Uma segunda métrica é a denominada Coeficiente de Correlação Cruzada Interauricular (IACC) vista no livro de [Blauert \(1997\)](#). O IACC é baseado na Função de Correlação Cruzada Interauricular (IACF):

$$\text{IACF}(\tau) = \frac{\int_0^{80 \text{ ms}} s_{esquerdo}(t) s_{direito}(t + \tau) dt}{\sqrt{\left[\int_0^{80 \text{ ms}} s_{esquerdo}^2(t) dt \right] \left[\int_0^{80 \text{ ms}} s_{direito}^2(t) dt \right]}}. \quad (5.36)$$

O IACC é definido como sendo o valor absoluto máximo entre $\tau \pm 1$ ms:

$$\text{IACC} = \max_{\forall \tau \in [-1 \text{ ms}, 1 \text{ ms}]} |\text{IACF}(\tau)|. \quad (5.37)$$

A quantidade $1 - \text{IACC}$ é comumente associada com a magnitude da impressão espacial ([JACKSON *et al.*, 2008](#)), mas para os fins desta verificação preliminar, o IACC é tomado como uma métrica de similaridade entre os sinais dos dois ouvidos ([BLAUERT, 1997](#)).

Os cálculos das métricas ILD_{Norm} e IACC foram feitos com relação às respostas biauriculares ao impulso para cada alto-falante na sala de escuta crítica. Para as fontes fantasmas, as respostas biauriculares ao impulso dos alto-falantes adjacentes balanceados por VBAP foram interpoladas pelo método proposto por Hannes [Gamper \(2013\)](#). Os resultados estão dispostos nas [Tabela 5.12](#) e [Tabela 5.13](#), juntamente com as medidas SPL feitas com os microfones intra-auriculares do HATS.

Para as fontes físicas na [Tabela 5.12](#), é possível notar que as trocas de sinal dos ILDs medidos seguem as dos ILD_{Norm} calculados, e os dois vetores

Tabela 5.12 – Fontes físicas (alto-falantes) discriminadas por azimute e elevação (θ, ϕ).

Medidas de nível feitas pelo HATS (dBSPL) e métricas interauriculares					
(θ, ϕ)	Esq. (dBSPL)	Dir. (dBSPL)	ILD medido (dB)	ILD _{Norm} (-1,1)	IACC (0,1)
(-60°, 0°)	72,6	67,5	5,1	0,1531	0,6319
(60°, 0°)	67,7	72,7	-5,0	-0,2045	0,5895
(0°, 0°)	69,9	70,1	-0,2	-0,0048	0,9539
(-135°, 0°)	70,3	66,8	3,5	0,0709	0,7255
(-135°, 0°)	66,9	71,1	-4,2	-0,1002	0,6395
(-30°, 0°)	72,3	68,4	3,9	0,0645	0,7688
(30°, 0°)	68,2	72,0	-3,8	-0,1323	0,7519
(180°, 0°)	69,6	69,3	0,3	-0,0048	0,9443
(-90°, 0°)	72,7	68,1	4,6	0,1171	0,5011
(90°, 0°)	66,4	72,5	-6,1	-0,1886	0,5069
(-45°, 30°)	72,1	67,2	4,9	0,1527	0,7205
(45°, 30°)	66,4	72,4	-6,0	-0,1991	0,7519
(0°, 30°)	69,4	69,7	-0,3	-0,0045	0,9694
(0°, 90°)	67,7	67,0	0,7	-0,0008	0,9624
(-135°, 30°)	70,3	66,9	3,4	0,0526	0,7662
(135°, 30°)	66,8	70,1	-3,3	-0,0950	0,7833
(-90°, 30°)	71,9	66,9	5,0	0,1255	0,6727
(90°, 30°)	66,7	72,5	-5,8	-0,2181	0,5602
(180°, 30°)	68,7	68,5	0,2	-0,0033	0,9572
(0°, -30°)	68,3	68,5	-0,2	-0,0030	0,9367
(-45°, -30°)	69,7	68,3	1,4	0,1151	0,7027
(45°, -30°)	66,9	70,6	-3,7	-0,1583	0,6577

Tabela 5.13 – Fontes fantasmas (trio de alto-falantes adjacentes) com direções discriminadas por azimute e elevação (θ, ϕ).

Medidas de nível feitas pelo HATS (dBSPL) e métricas interauriculares					
(θ, ϕ)	Esq. (dBSPL)	Dir. (dBSPL)	ILD medido (dB)	ILD _{Norm} (-1,1)	IACC (0,1)
(45°, 0°)	69,0	72,6	-3,6	-0,0719	0,7444
(120°, 0°)	67,4	71,5	-4,1	-0,0662	0,7464
(-45°, 0°)	74,4	68,0	6,4	0,0561	0,8210
(-120°, 0°)	71,3	67,0	4,3	-0,4061	0,6192
(45°, 15°)	67,9	72,5	-4,6	-0,1173	0,7727
(120°, 15°)	67,5	71,7	-4,2	-0,0982	0,6458
(-45°, 15°)	73,4	68,2	5,2	0,0943	0,7843
(-120°, 15°)	71,0	66,4	4,6	0,0001	0,6955

têm forte correlação ($r = 0,9732$). Ademais, os valores absolutos dos ILDs medidos também mostraram um alto grau de correlação negativa com os valores calculados de IACC ($r = -0,8854$), sugerindo que as diferenças de ILD estão associadas a uma maior impressão espacial, sendo possível que as diferenças de nível entre os sinais incidentes em cada ouvido possam explicar as variações de DLS entre alto-falantes.

Já para as fontes fantasmas na [Tabela 5.13](#), embora os valores medidos de ILD façam sentido, considerando-se as medidas feitas nas fontes físicas adjacentes, os resultados são descorrelacionados dos ILD_{Norm} e dos IACC calculados

($|r| < 0,2$ em ambos os casos). Estes resultados pobres podem ter várias explicações: desde erros no procedimento de interpolação das HRTFs até o número pequeno de funções de transferência a interpolar, resultando numa localização fracamente calculada.

Este esforço preliminar lançou alguma luz sobre quando procurar por correlações com métricas de áudio espacial. Embora restrita a fontes físicas reais, pode ser uma boa ideia persegui-las quando as DLSs dos participantes de fato estiverem disponíveis após a conclusão da coleta de dados do experimento principal. Além disso, em conjunto com as medidas de SPL feitas com os microfones intra-auriculares do HATS, as DLSs dos participantes podem alimentar o algoritmo de otimização usado em (SIVONEN; ELLERMEIER, 2006; SIVONEN; ELLERMEIER, 2008) e auxiliar na obtenção de perspectivas para a estimação dos ganhos direcionais.

5.5.4 *Teste piloto*

Após finalização da GUI para este experimento direcional, uma sessão experimental foi executada para mitigar problemas técnicos e ter alguns primeiros resultados da variável de resposta já no formato de DLS: diferenças de nível de áudio digital entre a incidência frontal de referência e a incidência não frontal de teste. Os estímulos foram apresentados aleatoriamente com três repetições por direção da fonte sonora, e o item de programação foi o mesmo ruído rosa entrecortado e limitado em faixa usado na etapa anterior.

As tarefas foram executadas na sala de escuta crítica com o sistema “H” de 22 canais em operação – ilustrado na Figura 5.24 e na Figura 5.37 – e também com fones de ouvido na sala de edição da Figura 5.25. Da mesma maneira que foram feitas as sínteses biauriculares do experimento anterior, estas sínteses foram feitas a partir de BRIRs da própria sala de escuta crítica.

As fontes fantasmas foram redefinidas considerando as direções com as maiores ILDs, e baseadas nas fontes físicas com os menores valores de IACC: $(\pm 90, 0)$, $(\pm 90, 30)$, $(\pm 60, 0)$, $(\pm 45, \pm 30)$ e $(\pm 135, 0)$ em graus. As direções das fontes fantasmas resultantes foram: $(\pm 45, \pm 15)$, $(\pm 52,5, 0)$, $(\pm 75, 0)$, $(\pm 75, 15)$, $(\pm 112,5, 0)$ e $(\pm 112,5, 15)$ em graus.

Os SPLs medidos com os microfones intra-auriculares do HATS – juntamente com a amostra da variável de resposta – alimentou o problema de minimização para estimação de ganhos biauriculares da [Equação 5.30](#). A principal diferença em relação aos trabalhos progressos que fizeram uso desse método de estimação, é o interesse não num único ganho biauricular para aplicação direta na fórmula de Robinson ([Equação 5.29](#)), como visto na [subseção 5.5.1](#), mas sim obterem ganhos distintos por localização como na contribuição da NHK para o grupo de trabalho de *loudness* do ITU-R ([2014c](#)). Para este intento, as SSEs devem ser somadas não mais ao longo das direções, mas sim ao longo do número de participantes – neste caso específico do teste piloto, somente ao longo das repetições de um único participante.

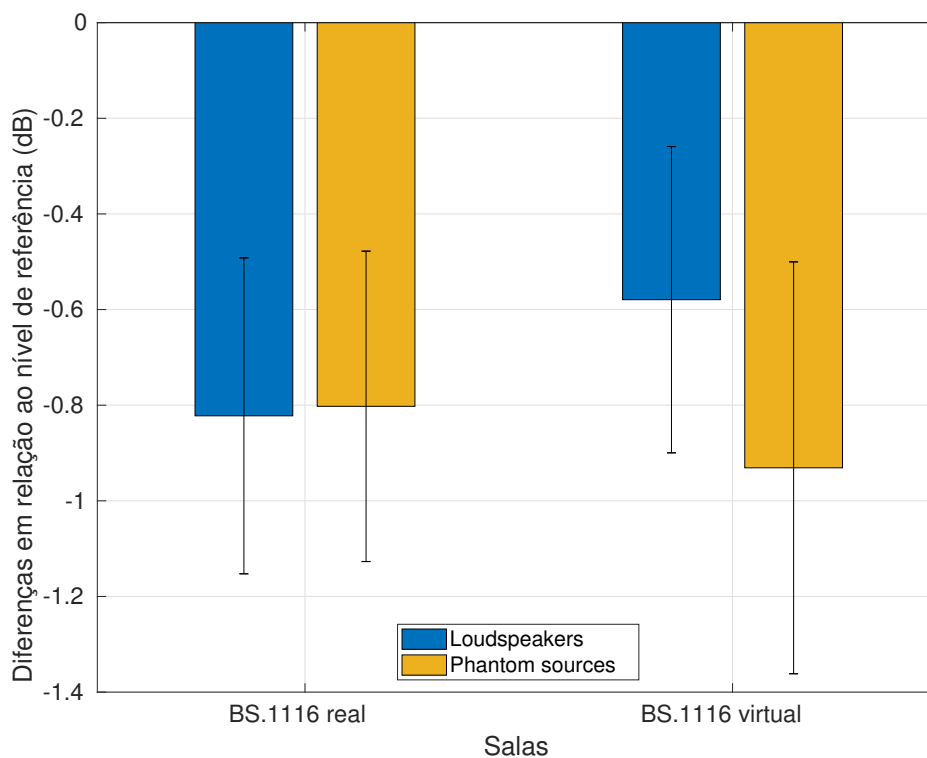
Sala real vs. sala virtual

Os ajustes de nível de áudio digital representados pelas suas diferenças em relação à incidência frontal são exibidos na [Figura 5.39](#). As respostas foram quebradas por sala real e virtual, como também em fontes reais e fantasmas.

As médias e os intervalos de confiança foram calculados com 63 pontos de dados de alto-falantes (21 fontes \times 3 repetições) e 42 pontos de dados de fontes fantasmas (14 fontes \times 3 repetições). As diferenças entre as médias são maiores entre os sinais biauriculares, e embora os intervalos de confiança possuam interseções em ambas as salas, o que sugere que as diferenças entre as médias não sejam estatisticamente significativas; isto pode mudar com mais dados à disposição.

Uma maior diferença geral de níveis de áudio digital entre fontes reais e fantasmas via fones de ouvido levanta a preocupação de que a síntese biauricular de fontes balanceadas pode não ser uma boa ideia. Todos somos familiares com a reprodução musical estereofônica em fones de ouvido: muito embora consigamos uma imagem sonora espacial de uma banda com balanceamento, também temos a impressão que a banda está tocando “dentro das nossas cabeças”, porque as dicas espaciais de localização acústica são erradas ou inexistentes. A reprodução biauricular mitiga este efeito, mas talvez a síntese biauricular não preserve todas as dicas de localização acústica de uma fonte sonora virtual. Para o benefício deste experimento portanto, é prudente não insistir nesta investigação direcional

Figura 5.39 – Ajustes de nível de áudio digital de fontes reais e fantasmas divididas entre a sala de escuta crítica real e sua versão virtual via auralização.



Nota – As diferenças entre as médias são maiores entre os sinais biauriculares.

Fonte: Elaborada pelo autor.

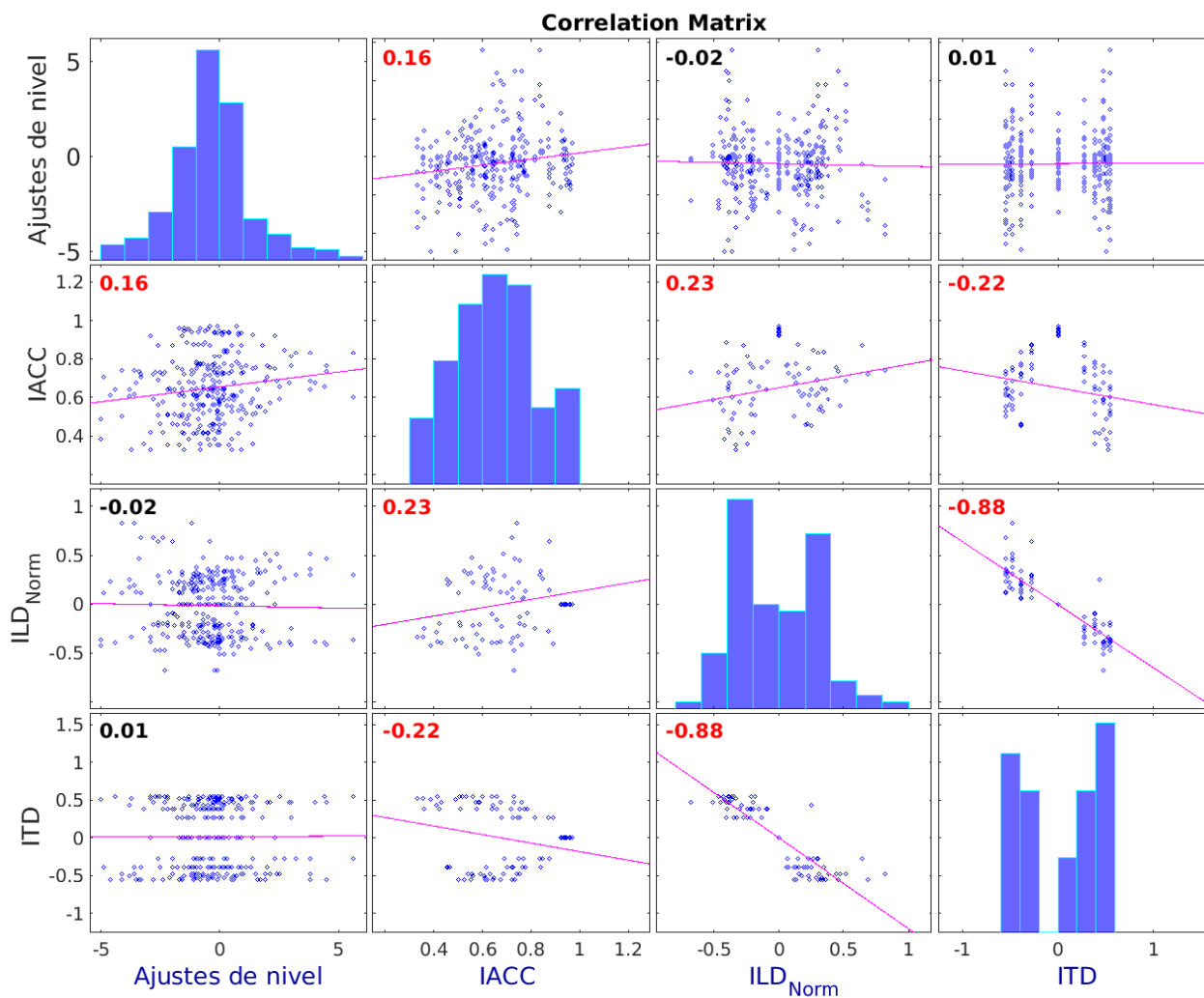
em salas virtuais.

Correlações

Relações entre a variável de resposta e as métricas IACC, ILD e ITD foram medidas usando coeficientes de correlação. A matriz de correlação é disposta na [Figura 5.40](#). Histogramas das variáveis são ilustrados ao longo da diagonal da matriz. Enquanto os ajustes de nível de áudio digital e os valores de IACC estão próximos da distribuição normal, as métricas ILD e ITD normalizadas parecem ser de distribuição bivariada. Logo, é razoável presumir que os ajustes do participante e os valores das métricas ILD e ITD são descorrelacionados.

Pode haver alguma correlação entre ajustes de nível de áudio digital e valores de IACC – ou entre DLSs e impressões espaciais – num conjunto de dados completo. O fator de correlação r de Pearson pode crescer com mais dados

Figura 5.40 – Matriz de correlação da variável de resposta com as métricas espaciais. Os coeficientes de correlação em vermelho indicam correlações significativamente diferentes de zero.



Nota – Histogramas das variáveis são dispostos ao longo da diagonal da matriz de correlação.

Fonte: Elaborada pelo autor.

à disposição ou com médias das respostas de todos os participantes, de modo a reduzir a dispersão dos dados.

Ganhos biauriculares

Com a intenção de obter mais dados para esse passo em particular, contou-se com ajustes de nível e medidas de SPL não somente feitas na sala de escuta crítica, como também nas salas virtuais do experimento de reverberação. Os ganhos biauriculares foram estimados pelo valor mínimo de um problema especí-

ficado para $i = 35$ direções e $j = 3$ repetições. A função objetivo da [Equação 5.30](#) é reescrita na forma abaixo:

$$SSE = \sum_{i=1}^{35} \sum_{j=1}^3 \left\{ DLS_{i,j} - \left[L_{mon,comp_i}(g) - L_{mon,ref}(g) \right] \right\}^2, \quad (5.38)$$

onde os SPLs de comparação e referência são computados pela [Equação 5.31](#) e pela [Equação 5.32](#) com base nas medidas de SPL feitas com o HATS.

A [Equação 5.31](#) foi usada como uma condição de contorno tal que $L_{mon,comp_i}(g)$ não fosse maior que o nível de referência somado à condição de somatório biauricular perfeito de 10 dB (ver [subseção 2.5.1](#)) para todas as i -ésimas direções. A solução foi limitada entre o somatório perfeito e nenhum somatório ($0 \leq g \leq 10$), e o ganho biauricular de 6 dB da fórmula de Robinson ([Equação 5.29](#)) foi usado como valor inicial ($g_0 = 6$).

Cálculos individuais por sala levaram a mínimos locais próximos das condições de contorno devido aos poucos pontos de dados. Quando mais de uma sala é considerada (sala de escuta crítica real e virtual), ou todas as salas com suas direções em comum ($(\pm 30,0)$, $(\pm 60,0)$ e $(\pm 90,0)$ em graus), os mínimos locais não são encontrados tão próximos assim dos limitantes inferior e superior (ver [Tabela 5.14](#)). Ganhos biauriculares comparáveis à ponderação direcional do algoritmo BS.1770 do [ITU-R \(2015b\)](#) poderiam então ser estimados a partir de um conjunto de dados completo.

Tabela 5.14 – Ganhos de somatório biauricular de *loudness* calculados e níveis apresentados na contribuição da NHK para o grupo relator de *loudness* do [ITU-R \(2014c\)](#).

Ângulo de azimute (θ)	0°	±30°	±60°	±90°	±135°	180°
Ganhos (todas as salas) (dB)	0,00	0,30	0,24	0,32	1,93	9,98
Ganhos (Salas BS.1116 espacial e biauricular) (dB)	0,00	8,93	8,96	7,61	10,00	9,98
Níveis apresentados em (ITU-R, 2014c) (dB)	0,00	1,36	4,47	5,22	0,84	8,25

As amplas diferenças nos ganhos direcionais estimados nas salas virtuais auxiliaram na construção e no funcionamento das rotinas de otimização, e somente nisso. Com as salas virtuais trabalhou-se somente com informações de azimute, portanto não foi possível avançar na questão da elevação, justamente o conjunto de ganhos ausente do modelo ITU-R.

Por fim, este teste piloto e a verificação preliminar da subseção anterior deixaram três possibilidades para estimação de ganhos direcionais baseada nos dados de coleta do experimento principal a saber:

1. Derivar uma curva de ganho baseada nas diferenças significativas entre as médias – se observadas – ao se tratar o efeito da direção na variável de resposta;
2. Obter um conjunto de ganhos direcionais pela minimização da soma de quadrados dos erros de sensibilidade de *loudness*, erros esses caracterizados pelas diferenças entre as sensibilidades observadas e as sensibilidades calculadas; e
3. Definir um conjunto de preditores a partir das métricas altamente correlacionadas com a variável de resposta – caso existam – e abordar a estimação de ganhos como um problema de regressão.

5.5.5 Experimento principal

Medidas de SPL com o HATS foram feitas no estágio preliminar, e uma GUI para controle dos estímulos e coleta das respostas dos participantes na sala de escuta crítica foi elaborada para o teste piloto. No primeiro, as fontes balanceadas por VBAP foram geradas por trios de alto-falantes próximos. Já no segundo, as fontes fantasmas foram geradas pelos alto-falantes de menor correlação interauricular. Porém, desta vez, foi preciso uma melhor sistematização da geração de fontes fantasmas em termos de dados interessantes para coleta.

Fontes balanceadas por VBAP

Para uma fonte real localizada na posição X , seja $g(X)$ o ajuste de nível que deve ser aplicado para que o nível de *loudness* da fonte esteja casado com o nível de *loudness* de uma fonte real localizada à posição de referência (diretamente à frente). Ao considerar-se criar uma fonte fantasma na posição P , entre dois alto-falantes A e B , as formas mais interessantes de se fazer isso do ponto de vista deste experimento, seriam:

- (i) $g(A)$ é muito diferente de $g(B)$; ou
- (ii) $g(P)$ é muito diferente de $g(A)$ e $g(B)$

O caso (i) é interessante pois é sabido que para se balancear uma fonte sonora dinamicamente de A para B , o *loudness* deve ser mantido constante

enquanto os ajustes de nível dos alto-falantes varia. Porém não é sabido se a variação resultante seria linear, logarítmica, etc. Ter dados para casos como este no qual o ajuste de nível é amplo, pode lançar alguma luz sobre esta incógnita.

Já o caso (ii) é interessante porque não é sabido que o ajuste de ganho necessário para a fonte fantasma seria $g(P)$, $g(A)$ ou $g(B)$, ou qualquer outro. Ter dados para casos como este onde há uma ampla diferença entre as opções possíveis pode também lançar alguma luz neste aspecto.

Idealmente, seria desejável coletar dados também das fontes reais nas posições A e B . Para o caso (ii), é importante coletar dados também de uma fonte real situada em P . Tomou-se então o método de predição usado pela NHK para balizar a escolha dos pares de alto-falantes (A, B) para todos os casos e para as posições P correspondentes às localizações das fontes reais. O trecho a seguir foi retirado da contribuição da emissora para o grupo relator do [ITU-R \(2014c\)](#):

O *loudness* direcional pode ser estimado usando funções de transferência relativas à cabeça (HRTFs) ([SIVONEN; ELLERMEIER, 2006](#)). O *loudness* percebido pode ser estimado a partir dos níveis de pressão sonora de ambos os ouvidos ao aplicar uma regra de somatório de 3 dB / 6 dB ([SIVONEN; ELLERMEIER, 2006](#); [SIVONEN; ELLERMEIER, 2008](#)). O Japão calculou o *loudness* tridimensional estimado usando as HRTFs de um manequim ([ITU-R, 2014c](#), p. 4, tradução minha).

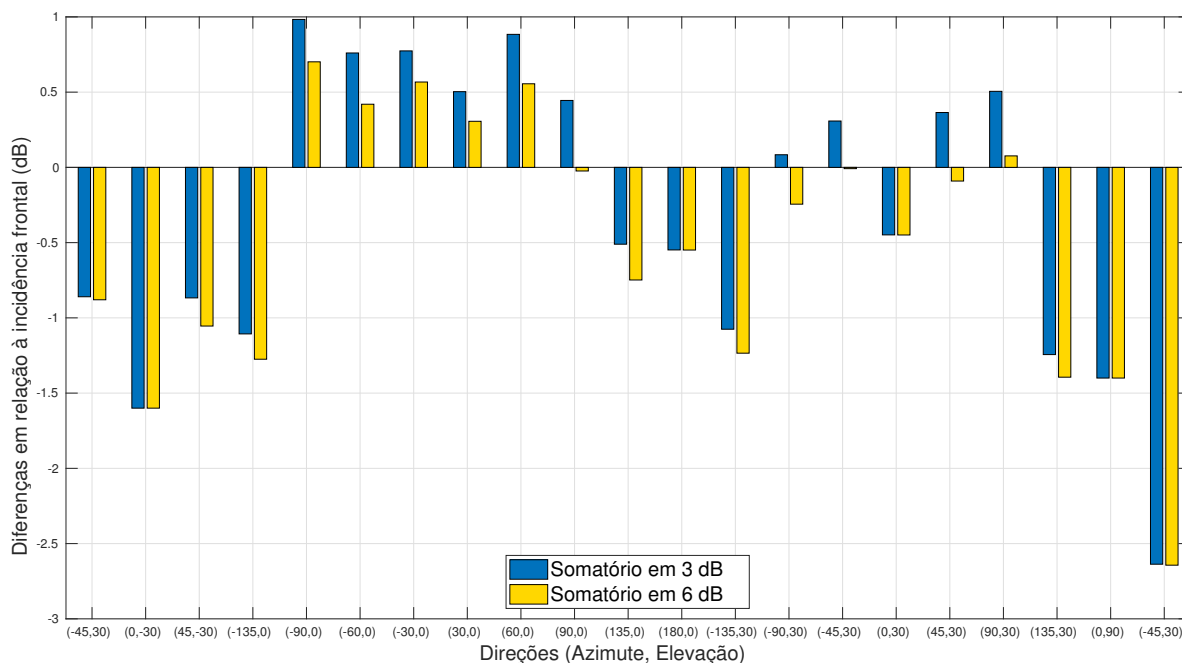
Por “aplicar uma regra de somatório de 3 dB / 6 dB”, entendeu-se que foi aplicada a fórmula de Robinson na sua forma geral:

$$L_{mon} = g \times \log_2 \left(2^{\frac{L_{esquerdo}}{g}} + 2^{\frac{L_{direito}}{g}} \right) \quad (5.39)$$

com $g = 3$ dB (ou 6 dB), onde $(L_{esquerdo}, L_{direito})$ são os SPLs intra-auriculares do HATS.

Os autores “aplicaram a regra de somatório” para 19 itens de programação em convolução com as HRTFs de uma sala de escuta, enquanto que nesta reprodução, a [Equação 5.39](#) é calculada com os SPLs medidos com o HATS ao se reproduzir o estímulo deste experimento por diferentes direções. Os resultados foram organizados na forma de diferenças em relação à incidência frontal $(L_{mon}(\theta, \phi) - L_{mon}(0^\circ, 0^\circ))$ e são dispostos na [Figura 5.41](#). Os pares de alto-falantes geradores de fontes fantasmas de mesma direção que fontes reais, para os casos (i) e (ii), estão relacionados na [Tabela 5.15](#).

Figura 5.41 – Gráficos cartesianos das diferenças de somatório biauricular em relação à incidência frontal com ganhos biauriculares de 3 dB and 6 dB.



Fonte: Elaborada pelo autor.

Coleta de dados

Doze participantes, em duas sessões cada, executaram tarefas de casamento de *loudness* pelo método de ajuste no qual a variável de resposta se dá no formato de sensibilidades direcionais de *loudness* (DLS), a diferença de níveis de áudio digital entre as posições frontal, de referência, e não-frontal, de teste. O item de programação foi o mesmo ruído rosa entrecortado e limitado em faixa usado no teste piloto e nos experimentos anteriores. Um total de 58 estímulos (22 fontes reais + 18 fontes fantasmas do caso (i) + 18 fontes fantasmas do caso (ii)) foram aleatoriamente apresentados 58 vezes, uma por direção da fonte.

Assumindo que uma tarefa de casamento de *loudness* seja completada entre 20 e 25 segundos, uma sessão com 58 apresentações e uma repetição cada, levaria dois rounds de 20 a 25 minutos, separados por um intervalo de 10 minutos, com duração total de aproximadamente uma hora. Embora esta previsão tenha parecido razoável com base no experimento piloto, o tempo de conclusão das tarefas foi subestimado. Além disso, os participantes das duas primeiras sessões se queixaram de cansaço após 116 tarefas. Então da terceira sessão em diante,

Tabela 5.15 – Pares de alto-falantes para os casos de balanceamento VBAP 1 e 2, identificados por suas posições (azimute e elevação em graus).

Posição P (θ, ϕ)	Par (A,B) para o caso 1	Par (A,B) para o caso 2
(-60, 0)	(0, 0) / (-90, 0)	(30, 0) / (-135, 0)
(60, 0)	(0, 0) / (90, 0)	(-30, 0) / (135, 0)
(0, 0)	(-30, 0) / (60, 0)	(-60, 0) / (60, 0)
(-135, 0)	(-90, 0) / (135, 0)	(-60, 0) / (135, 0)
(135, 0)	(90, 0) / (-135, 0)	(60, 0) / (-135, 0)
(-30, 0)	(0, 0) / (-90, 0)	(30, 0) / (-90, 0)
(30, 0)	(0, 0) / (90, 0)	(-30, 0) / (90, 0)
(180, 0)	(-90, 0) / (135, 0)	(-135, 0) / (135, 0)
(-90, 0)	(0, 0) / (-135, 0)	(-30, 0) / (-135, 0)
(90, 0)	(0, 0) / (135, 0)	(30, 0) / (135, 0)
(-45, 30)	(0, 30) / (-90, 30)	(45, 30) / (-90, 30)
(45, 30)	(0, 30) / (90, 30)	(-45, 30) / (90, 30)
(0, 30)	(-45, 30) / (90, 30)	(-45, 30) / (45, 30)
(-135, 30)	(-90, 30) / (180, 30)	(-90, 30) / (135, 30)
(135, 30)	(90, 30) / (180, 30)	(90, 30) / (-135, 30)
(-90, 30)	(-45, 30) / (180, 30)	(-45, 30) / (135, 30)
(90, 30)	(45, 30) / (180, 30)	(45, 30) / (135, 30)
(180, 30)	(90, 30) / (-135, 30)	(-135, 30) / (135, 30)

manteve-se um número de 58 tarefas por sessão, durando aproximadamente 40 minutos cada.

A tela da interface de usuário foi uma versão em língua inglesa da tela na [Figura 5.27](#), e as mudanças deram-se internamente no controle da interface de áudio para reprodução de estímulos e no roteamento dos 22 canais. Já a folha de informações e o formulário de consentimento foram versões em inglês dos utilizados no experimento anterior, alterando-se somente a logomarca da universidade no timbrado dos papéis. Os documentos estão disponíveis para consulta no [Apêndice B](#).

Desempenho dos participantes

Médias e intervalos de confiança dos participantes são exibidos na [Figura 5.42a](#) e na [Figura 5.42b](#). As médias dos participantes no geral oscilaram entre -0,5 dB e 2 dB em relação à referência frontal. Houve menos variabilidade quando comparadas com as médias dos ouvintes inexperientes no experimento

de reverberação, e mais variabilidade em comparação com o desempenho do conjunto de ouvintes do experimento de distância, formado somente por *ton-meisters*. A variabilidade das respostas no grupo de fontes balanceadas foi maior do que a do grupo de fontes reais para a maioria dos participantes.

As diferenças nas médias entre fontes reais e fantasmas são dispostas na [Figura 5.43](#), na qual as anotações nas setas indicam o par de azimutes dos alto-falantes que geraram esta ou aquela fonte imagem em particular, de mesma elevação. Foram observadas diferenças mais amplas entre fontes reais e fantasmas para as incidências frontais e posteriores, nas quais as dicas de localização dos alto-falantes geradores do balanceamento interferem diretamente na formação de uma imagem sonora nestas direções.

Análise de dados

Diagramas de caixa das sensibilidades direcionais de *loudness* por fonte sonora, agrupada por alto-falantes e pelas fontes balanceadas por VBAP, são ilustrados na [Figura 5.44](#). Para as fontes sonoras reais, note que as maiores sensibilidades provocadas correspondem às incidências ipsilaterais. O mesmo não ficou claro para as fontes fantasmas dos casos (i) e (ii).

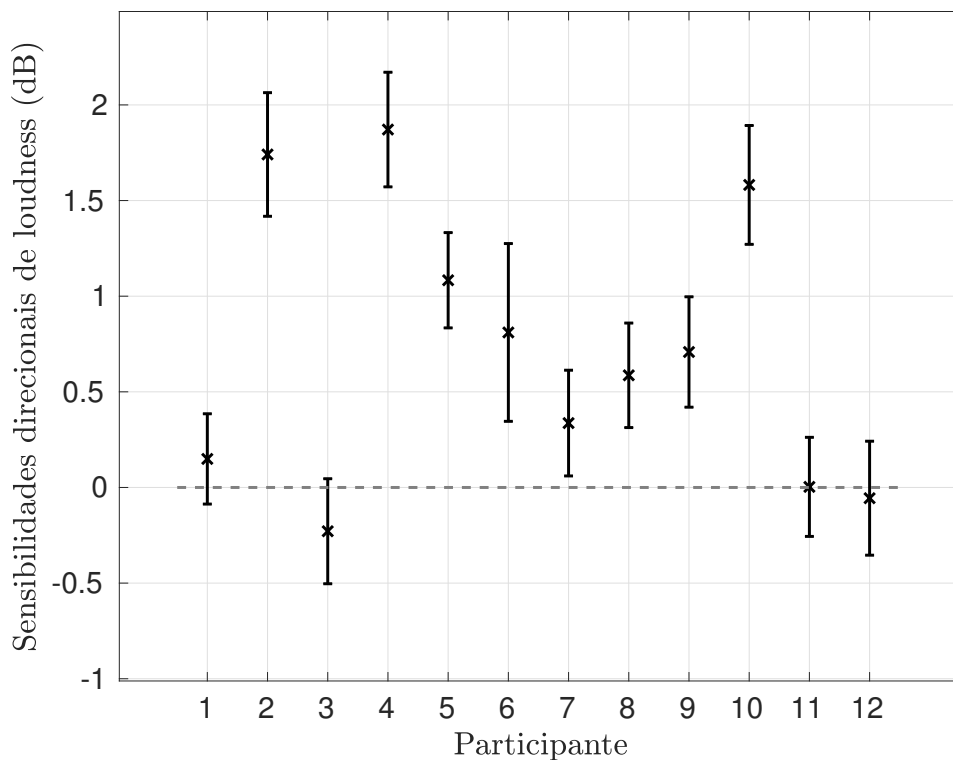
Histogramas e gráficos quantil-quantil das fontes reais e fantasmas dispostos na [Figura 5.45](#) sugerem que as distribuições estão próximas da distribuição normal em todos os casos.

Os dados dos participantes foram ajustados a um modelo linear composto pelas DLSs como variável de resposta, e os pelos fatores experimentais “direção”, “fonte sonora” e “participante”. Um gráfico dos valores ajustados contra os resíduos do modelo é ilustrado na [Figura 5.46](#). A concentração dos valores no entorno da origem sugere problemas de heteroscedasticidade.

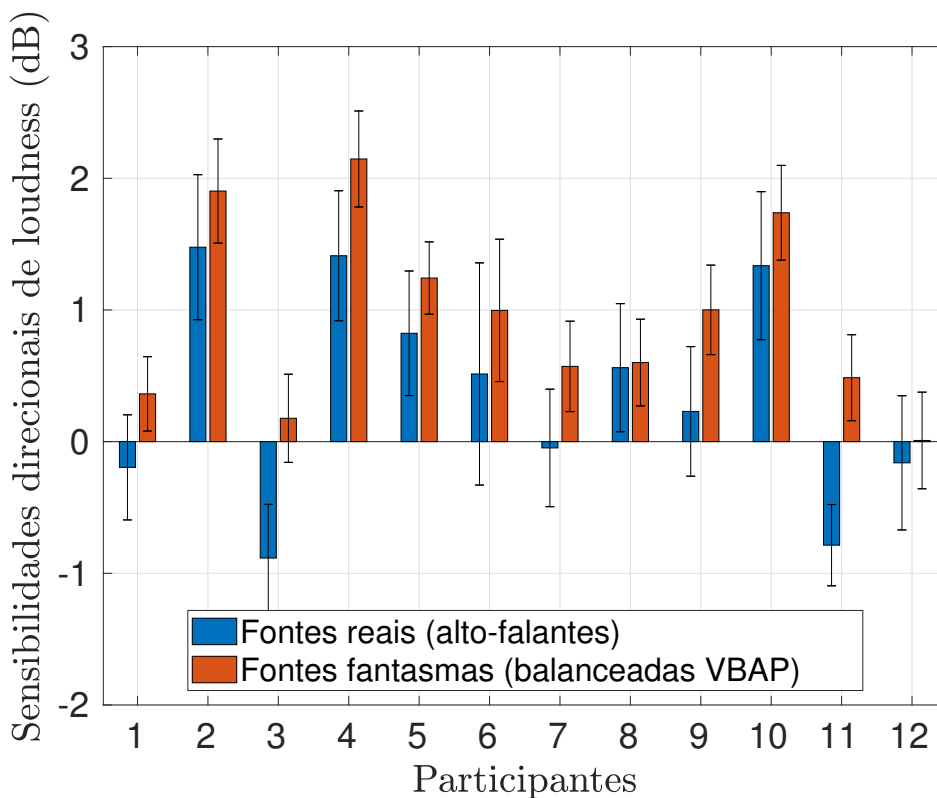
Estatística descritiva

De modo que a distribuição dos dados se aproximasse mais da distribuição normal, as respostas dos participantes foram submetidas a um tratamento de *outliers*. Valores superiores em módulo a $\pm 2.5 \times \sigma$ foram considerados extremos e então removidos (*trimmed*), e os *ouliers* remanescentes foram recodificados

Figura 5.42 – Avaliação do desempenho dos participantes.

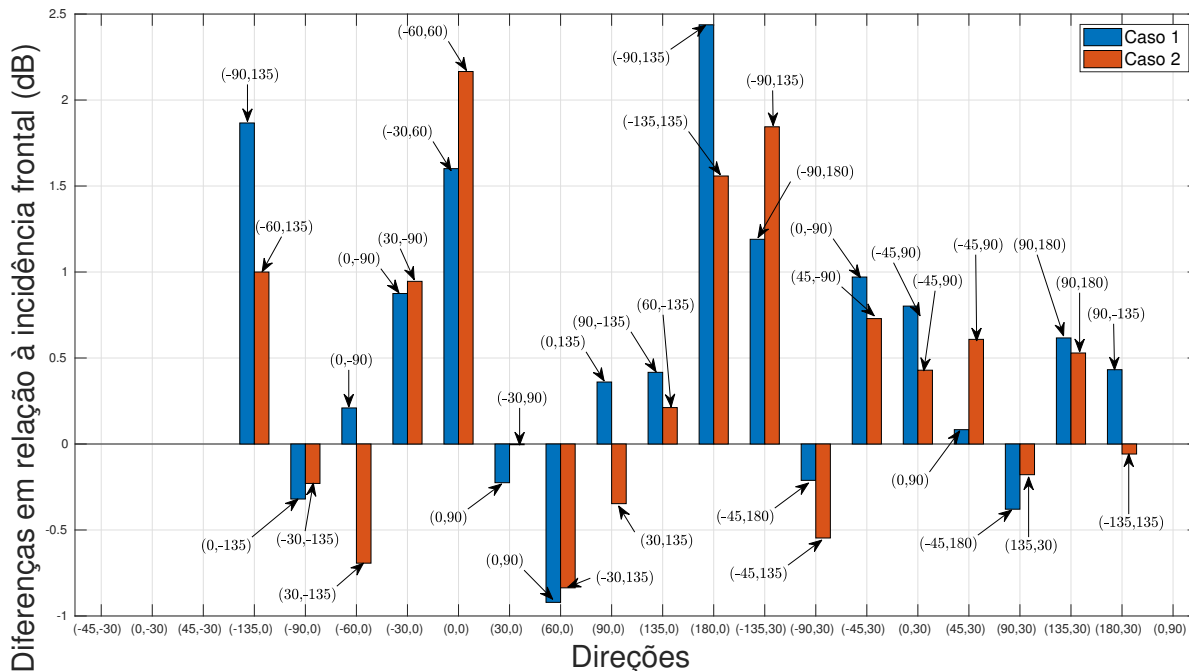


(a) Médias e intervalos de confiança de todas as respostas.



(b) Médias e intervalos de confiança quebrados entre fontes reais e fantasmas.

Figura 5.43 – Diferenças nas médias dos participantes entre fontes reais e fantasmas. Anotações com setas indicam o par de azimutes dos alto-falantes que geraram esta ou aquela fonte imagem em particular, de mesma elevação.



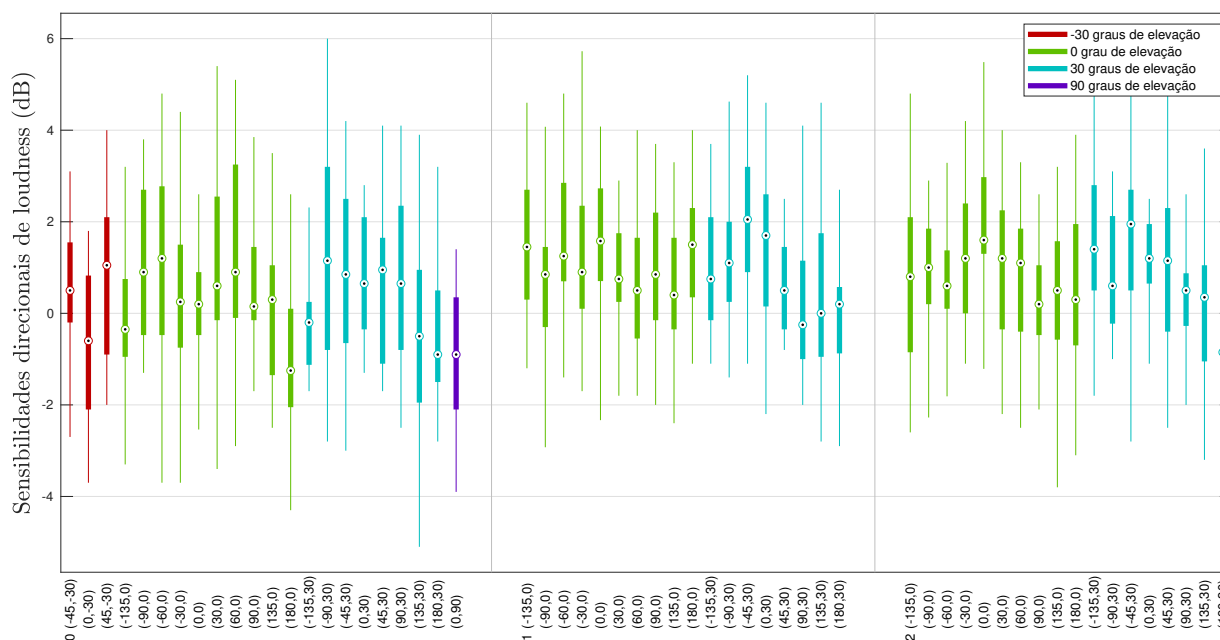
Nota – Foram observadas diferenças mais amplas entre fontes reais e fantasmas para as incidências de frente e costas, nas quais as dicas de localização dos alto-falantes geradores do balanceamento interferem diretamente na formação de uma imagem sonora nestas direções.

Fonte: Elaborada pelo autor.

pelos valores mais altos não-outliers (*winsorized*).

Testes de Kolmogorov-Smirnov por alto-falante de referência a um intervalo de confiança de 95% rejeitaram a hipótese nula de distribuições normais para todos os grupos de fonte sonora. Porém, pelos gráficos da Figura 5.45, testes de normalidade importam pouco, porque o tamanho da amostra neste experimento é grande o suficiente para que as distribuições amostrais sejam normais, independentemente de como os dados da amostra se apresentem. A assimetria dos dados totais é de 0,11 e, quando quebrados em grupos por fonte sonora, o grupo das fontes reais possui a maior assimetria (0,23) e o grupo das fontes do caso de balanceamento (i) possui o maior valor de curtose (0,34). Com esses valores é possível dizer que as distribuições da Figura 5.45 são aproximadamente simétricas (assimetria < 0,5 em módulo) e com formato de sino normal (curtose < 1,0 em módulo).

Figura 5.44 – Diagramas de caixa de Sensibilidades Direcionais de Loudness (DLS) por posição da fonte sonora, divididos em grupos de fontes reais e fantasmas (casos (i) e (ii)).

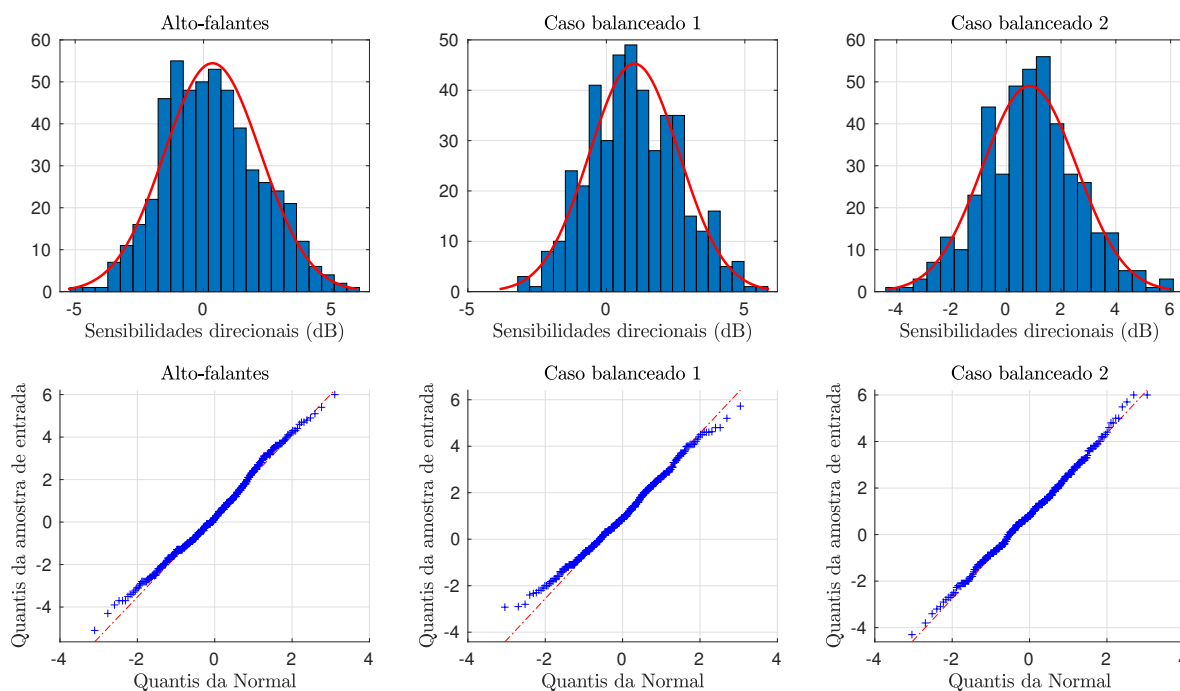


Nota – Para as fontes sonoras reais, note que as maiores sensibilidades provocadas correspondem às incidências ipsilaterais. O mesmo não ficou claro para as fontes fantasmas dos casos (i) e (ii).

Fonte: Elaborada pelo autor.

Já com relação às variâncias, usar o teste de Levene para avaliar sua homogeneidade entre cada variável independente foi importante para apontar os caminhos por onde a análise inferencial deveria seguir. O teste entre os grupos de fontes sonoras revelou variâncias significativamente desiguais [$F_{(2,1277)} = 5,30, p = 0,005$]. Embora a ANOVA padrão não possa ser usada neste caso em virtude da violação da premissa de homoscedasticidade, é possível que as diferenças entre os grupos possam ser evidenciadas via ANOVA de Friedman, usada para comparação de grupos quando as respostas vêm dos mesmos participantes. Já o teste de Levene entre direções não rejeitou a hipótese nula de homoscedasticidade [$F_{(17,1262)} = 1,23, p = 0,233$] e, portanto, análises de variância de uma via por grupo de fontes sonoras poderiam ser feitas. E o mesmo teste entre participantes revelou uma desigualdade de variâncias muito significativa [$F_{(11,1268)} = 11,55, p < 0,001$], confirmando o que já se suspeitava por inspeção visual da Figura 5.42a (gráfico à esquerda). Logo, para se examinar todos os

Figura 5.45 – Histogramas e gráficos Q-Q dos dados dos participantes, divididos em grupos de fontes reais e fantasmas (casos (i) e (ii)).



Nota – As distribuições estão próximas da distribuição normal em todos os casos.

Fonte: Elaborada pelo autor.

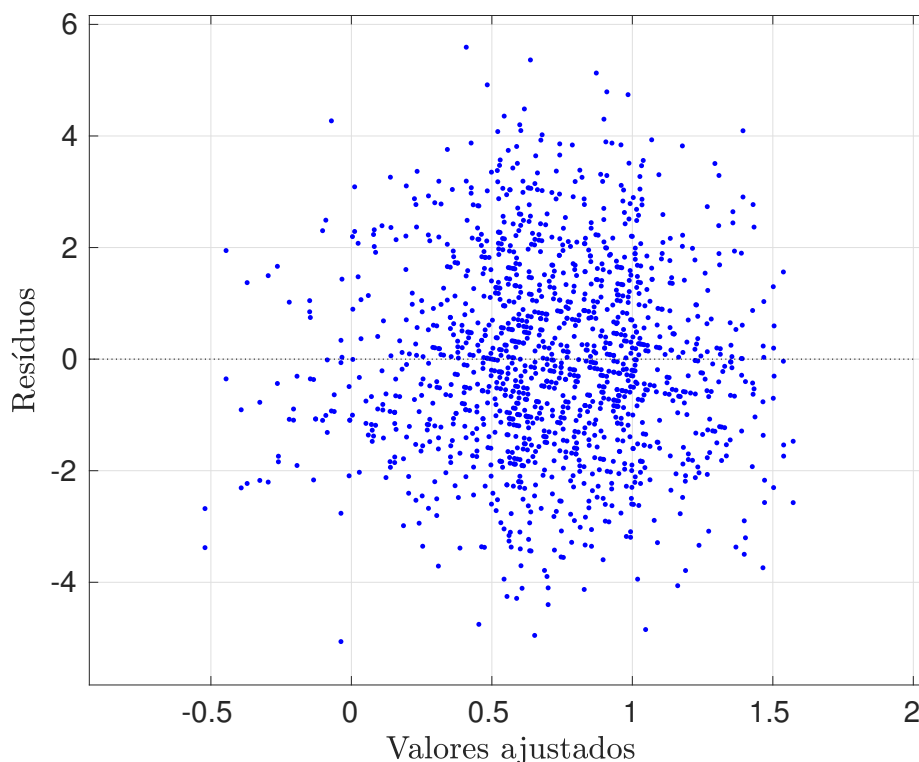
efeitos e suas interações, o fator experimental “participante” deve ser incluído na análise de variâncias como uma variável aleatória, num modelo de efeitos mistos.

Estatística inferencial

Com heteroscedasticidade encontrada entre os grupos de fontes sonoras, uma ANOVA padrão não se presta a investigar o efeito “fonte sonora” individualmente. Apesar do uso da ANOVA de Friedman ter sido uma possibilidade, o teste de Mauchly indicou que a premissa de esfericidade foi violada ($p < 0,001$).

A alternativa aqui foi não mais fazer nenhum tratamento de *outliers*, e sim seguir com um projeto de experimento balanceado para então trabalhar-se com uma ANOVA de medidas repetidas, com graus de liberdade ajustados pela estimação de esfericidade de Greenhouse-Geisser ($\epsilon = 0,25452$), sob o risco de se obter *F-scores* mais conservadores e menos acurados (FIELD, 2013). Com efeito, diferenças muito significativas entre as médias foram encontradas para

Figura 5.46 – Predições vs. resíduos de um modelo linear ao qual os dados foram ajustados.



Nota – A concentração dos valores no entorno da origem sugere problemas de heteroscedasticidade.

Fonte: Elaborada pelo autor.

todas as variáveis independentes: fonte sonora [$F_{(1,23)} = 33,26, p_{GG} < 0,001$], direção [$F_{(17,391)} = 6,00, p_{GG} < 0,001$] e suas interações [$F_{(1,23)} = 5,01, p_{GG} < 0,001$].

Posteriormente à ANOVA de medidas repetidas, comparações par-a-par entre grupos de fontes sonoras pelo método das HSDs de Tuckey resultaram em diferenças muito significativas nas médias entre as fontes reais e do caso (i) ($p < 0,001$), diferenças significativas nas médias entre as fontes reais e do caso (ii) ($p = 0,002$), e diferenças não significativas nas médias entre as fontes dos casos balanceados por VBAP (i) e (ii) ($p = 0,341$). Muito embora estas estatísticas possam ser enviesadas por *outliers* e com *F-scores* de menor resolução, ganha-se a noção de que os mesmos participantes ajustaram os níveis de áudio de uma fonte fantasma de forma diferente dos níveis de áudio reproduzidos por um alto-falante real na mesma direção desta fonte fantasma, simulada por balanceamento de pares de alto-falantes.

Dado que a hipótese de homoscedasticidade da variável independente “direção” não foi rejeitada, ANOVAs padrão de uma via em cada grupo de fontes sonoras podem clarificar como as sensibilidades variaram por direção em cada um dos grupos. *F-scores* de dados pós-tratamento de *outliers* resultaram em diferenças muito significativas nas médias entre os níveis do fator experimental “direção” por grupo de fontes: com efeito de tamanho médio para o grupo de fontes reais [$F_{(21,1093)} = 8,26, p < 0,001, \omega^2 = 0,12$], e com efeitos de tamanho pequeno nos casos de balanceamento (i) [$F_{(17,900)} = 3,82, p < 0,001, \omega^2 = 0,05$] e (ii) [$F_{(17,895)} = 8,26, p < 0,001, \omega^2 = 0,03$].

Em seguida às ANOVAs padrão em cada grupo, comparações *post hoc* foram feitas com o procedimento *S* de Scheffé para tamanhos de amostra desiguais. Dentre as fontes sonoras reais representadas por suas coordenadas (Azimute, Elevação), diferenças muito significativas foram encontradas somente entre os pares de direções: (180, 0)/(-90, 0), (-90, 0)/(0, 90), e (-90, 0)/(180, 30) em graus, e diferenças significativas foram observadas entre os pares de direções (-30, 0)/(180, 0), (180, 0)/(-45, -30), (180, 0)/(45, -30), (-90, 0)/(0, -30), (0, 90)/(-45, 30) e (180, 30)/(-45, -30) em graus. Nenhuma diferença significativa entre pares de direções foi observada nos casos de balanceamento (i) e (ii). Apesar destes *F-scores* terem sido contaminados pela inclusão do efeito da variável “participante” no ruído experimental, pode-se esperar que as diferenças nas médias entre diferentes níveis do fator direcional numa análise geral não sejam muito pronunciadas em virtude dos efeitos observados serem de tamanho médio a pequeno.

Para a observação geral de todos os efeitos e suas interações, os dados foram ajustados a um Modelo Linear Generalizado de Efeitos Mistos (GLME), por ser uma extensão de um GLM para dados coletados e divididos por grupos, como também é uma generalização de um Modelo Linear de Efeitos Mistos (LME) para os dados cujas variáveis de resposta não são normalmente distribuídas. O modelo possui “direção” e “fonte sonora” como variáveis fixas, e “participante” como variável aleatória, formulado como:

$$Y_{i,j,s} = \underbrace{\mu + \tau_{1,i} + \tau_{2,j} + \tau_{i,j}}_{\mu_{i,j}} + \alpha_s + \beta_{j,s} + \gamma_{i,s} + \varepsilon_{i,j,s} \quad (5.40)$$

Tabela 5.16 – Estatísticas de ajuste do Modelo Linear Generalizado de Efeitos Mistos.

Estatísticas de qualidade de ajuste	GLME sem interações	GLME com interações
Critério de Informações de Akaike (AIC)	4828,60	4808,00
Critério de Informações de Bayesiano (BIC)	4952,30	5106,90
Critério de Informações de Desvio (DIC)	4780,60	4692,00

onde $Y_{i,j,s}$ são as observações (variável de resposta), μ é a média geral, $\tau_{1,i}$ é o tratamento do efeito da i -ésima direção na média, $\tau_{2,j}$ é o tratamento do efeito da j -ésima categoria de fonte sonora na média, $\tau_{i,j}$ é o efeito da interação da i -ésima direção com a j -ésima categoria de fonte sonora na média, α_s é o efeito do s -ésimo participante, $\beta_{j,s}$ é o efeito da interação da j -ésima categoria de fonte sonora com o s -ésimo participante, $\gamma_{i,s}$ é o efeito da interação da i -ésima direção com o s -ésimo participante, e $\varepsilon_{i,j,s}$ é o resíduo do modelo.

De início, pairava a preocupação quanto a se trabalhar com três variáveis independentes – duas fixas, uma aleatória – e suas interações resultaria em perda de resolução. Os dados foram então ajustados a GLMEs com e sem interações entre efeitos, porém as métricas de qualidade de ajuste Critério de Informações de Akaike (AIC), Critério de Informações de Bayesiano (BIC) e Critério de Informações de Desvio (DIC) foram aproximadamente as mesmas (ver [Tabela 5.16](#)); e por essa razão, optou-se por manter o modelo com interações entre efeitos tal como descrito pela [Equação 5.40](#).

O teste de determinação se todos os coeficientes de efeitos fixos são iguais a zero – o que equivaleria à hipótese nula de uma ANOVA se o modelo fosse um GLM comum – revelou que todos os fatores experimentais fixos foram significativamente diferentes entre seus níveis: fonte sonora [$F_{(2,1226)} = 15,21$, $p < 0,001$], direção [$F_{(17,1226)} = 6,38$, $p < 0,001$] e suas interações [$F_{(34,1226)} = 2,72$, $p < 0,001$]. Os coeficientes estimados dos efeitos fixos do modelo que foram muito significativamente diferentes de zero foram os correspondentes a todos os níveis do fator “fonte sonora”, as direções $(-45, 30)$, $(135, 30)$ e $(180,30)$ em graus, e as interações da direção $(180, 0)$ com cada categoria de fonte sonora. Os coeficientes significativamente diferentes de zero foram os das direções $(-45, 30)$, $(-90, 30)$, $(0, 0)$, $(0, 30)$, $(135, 0)$ e $(180, 0)$ em graus, interações das direções $(60, 0)$ e $(-135, 0)$ em graus com todas as categorias de fonte sonora, e interações da categoria fontes reais com as direções $(-135, 30)$, $(-90, 0)$, $(-90, 30)$ e $(0, 0)$ em graus. Apesar de as fontes sonoras concentrarem

as observações mais significativas dos seus efeitos, pôde-se observar também que as direções referentes ao plano de topo e às costas do ouvinte tiveram seus níveis os mais diferentemente ajustados em relação à incidência frontal.

Matrizes de contraste foram aplicadas para se testar diferenças entre categorias de fontes sonoras para direções específicas. Diferenças muito significativas foram observadas para a incidência frontal – real vs. caso de balanceamento (ii) [$F_{(1,1226)} = 10,76, p < 0,001$] – e a incidência traseira – real vs. caso de balanceamento (i) [$F_{(1,1226)} = 20,33, p < 0,001$]. Diferenças significativas foram encontradas entre fontes reais e fontes fantasmas nas direções $(-135, 0)$, $(-135, 30)$, $(\pm 60, 0)$ e $(\pm 90, 30)$. Ao que tudo indica, as maiores diferenças de sensibilidade no casamento de *loudness* entre direções de fontes reais e direções de fontes fantasmas estão mais associadas com as fontes sonoras frontais e traseiras, do que propriamente com as incidências ipsilaterais.

Comentários

Quando se olha para a análise inferencial do experimento de reverberação, o tratamento do efeito do azimute na variável de resposta foi significativo, mas não tanto quanto o efeito da reverberação em si. Esta diferença na significância dos efeitos foi refletida na superfície de ganho derivada das médias dos participantes, na qual os tempos de reverberação das salas virtuais contribuíram muito mais para o seu formato do que os azimutes das fontes sonoras – a ponto de a implementação de múltiplas curvas de ganho por canal não justificar o esforço, se comparada a ter uma única curva de ganho anterior à filtragem K , tal como na modificação do algoritmo feita no experimento de distância.

Para esta montagem experimental, o efeito da localização também foi “ofuscado” por outro efeito bem mais significativo (categoria de fonte sonora). Na análise geral, somente as direções originárias no plano de topo e as interações com as incidências traseiras corresponderam aos coeficientes do modelo linear significativamente diferentes de zero. Ademais, quando os dados são divididos entre fontes reais e os casos balanceados, o efeito observado foi de tamanho médio para as fontes sonoras reais e de tamanho pequeno para as fontes fantasmas. Estas constatações depositam pouca confiança na obtenção de elementos de ganho puramente baseados no relacionamento entre localizações tratadas e

respostas dos participantes.

Estes resultados abriram uma difícil escolha na procura por modos alternativos de estimação de ganho: resolver um problema de minimização com uma função objetivo composta pela soma dos quadrados dos erros entre as respostas dos participantes e sensibilidades estimadas, ou resolver um problema de regressão usando métricas correlacionadas com a variável de resposta como preditores do modelo. Ambas as abordagens considerarão somente os dados das fontes reais, dado o efeito observado ser de maior tamanho. Os dados das respostas dos participantes a partir de estímulos produzidos por fontes balanceadas VBAP serão então guardados e explorados em pesquisa futura.

5.5.6 Estimação de ganhos: problema de otimização

A função objetivo descrita na Equação 5.30 e na Equação 5.38, uma soma de quadrados de erros entre respostas dos participantes e sensibilidades estimadas, é agora reescrita a se somar a SSE, para cada conjunto de coordenadas, ao longo de 12 participantes e 2 repetições para obtenção de um vetor \vec{g} de ganhos direcionais estimados.

O problema de otimização é definido como Minimização Restrita. É o problema de se encontrar um vetor \vec{g} que represente um mínimo local para uma função escalar $f(\vec{g})$ sujeita às restrições :

$$\min_{\vec{g}} f(\vec{g}) \text{ tal que } \begin{cases} c(\vec{g}) \leq 0 \\ ceq(\vec{g}) = 0 \\ A \cdot \vec{g} \leq b \\ Aeq \cdot \vec{g} = beq \\ lb \leq \vec{g} \leq ub, \end{cases} \quad (5.41)$$

onde \vec{b} e \vec{beq} são vetores, A e Aeq são matrizes e $c(\vec{g})$ e $ceq(\vec{g})$ são funções que retornam vetores. As funções $f(\vec{g})$, $c(\vec{g})$ e $ceq(\vec{g})$ podem ser não lineares, e \vec{g} , lb e ub são argumentos na forma de vetores ou matrizes.

O problema deste experimento não está sujeito a desigualdades lineares,

então a [Equação 5.41](#) é reescrita da forma:

$$\min_{\vec{g}} f(\vec{g}) \text{ tal que } \begin{cases} c(\vec{g}) \leq 0 \\ lb \leq \vec{g} \leq ub, \end{cases} \quad (5.42)$$

onde a restrição não linear é definida tal que cada SPL monauricular estimado não poderia ser superior a um nível prático de segurança de 80 dBSPL:

$$c(\vec{g}) = \vec{g} \times \log_2 \left(2^{\left(\frac{L_{\text{esquerdo, comp}}}{\vec{g}}\right)} + 2^{\left(\frac{L_{\text{direito, comp}}}{\vec{g}}\right)} \right) \leq 80 \text{ dBSPL}. \quad (5.43)$$

O limitante superior é o somatório biauricular perfeito e o limitante inferior é um piso seguro: $-1,0 \leq \vec{g}_i \leq 10,0$ dB. Os valores iniciais em cada caso são próximos de nenhum ganho: $\vec{g}_{0i} = 0,1$ dB \forall_i .

O algoritmo da Minimização Restrita usou o método de Programação Sequencial Quadrática de [Schittkowski \(1986\)](#) e os valores das médias das respostas dos participantes e das médias das estimações, calculadas ao longo do número de participantes e repetições, foram passadas à rotina de minimização. A incidência traseira atingiu o limitante superior enquanto as direções originárias do plano inferior, incidência frontal, azimutes 0° e 180° no plano de topo, e o alto-falante de topo (0° , 90°) atingiram um mínimo global de 0,69. Resultados pós-equalização de ganhos simétricos e normalização das incidências laterais ($\pm 90^\circ$, 0°) a 1,5 dB são listados na [Tabela 5.17](#).

Com exceção da direção (180° , 0°), que atingiu o limitante superior, todas as direções localizadas no plano inferior e no plano médio vertical atingiram um mínimo global de 0,69 (normalizado a zero na [Tabela 5.17](#)). Não obstante os valores estimados parecerem razoáveis se comparados com os encontrados em ([ITU-R, 2014c](#)) (dispostos na [Tabela 5.9](#)), ainda não forneceram nenhuma perspectiva sobre efeitos específicos de elevação, o que torna estes resultados de pouca serventia, não muito diferentes do próprio conjunto de ganhos direcionais G_N da Recomendação BS.1770 do [ITU-R \(2015b\)](#). Contudo, isso não significa que este método seja inadequado para se estimar um ganho biauricular geral, como assim fizeram seus proponentes. Reposicionando o problema de minimização restrita com uma função objetivo que some os quadrados dos erros ao longo das direções e das repetições, estima-se então um conjunto de ganhos biauriculares individuais por participante, e conseqüentemente um ganho geral resultante

Tabela 5.17 – Ganhos direcionais estimados por solução de um problema de minimização restrita.

Azimute θ ($^{\circ}$)	Elevação ϕ ($^{\circ}$)	Ganho g (dB)
-45	-30	0,00
0	-30	0,00
45	-30	0,00
-135	0	0,44
-90	0	1,50
-60	0	1,10
-30	0	0,65
0	0	0,00
30	0	0,65
60	0	1,10
90	0	1,50
135	0	0,44
180	0	0,00
-135	30	0,28
-90	30	0,93
-45	30	1,06
0	30	0,00
45	30	1,06
90	30	0,93
135	30	0,28
180	30	0,00
0	90	0,00

da média dos valores estimados. Os ganhos biauriculares por participante são dispostos na [Tabela 5.18](#).

Tabela 5.18 – Ganhos biauriculares estimados por participante.

1	2	3	4	5	6	7	8	9	10	11	12
4,13	1,18	6,89	1,50	1,66	3,43	4,16	2,94	3,67	0,94	7,27	4,76

O cálculo da média dos valores na [Tabela 5.18](#) resulta num ganho de somatório biauricular de *loudness* estimado de $g = 3,54$. Este seria o ganho biauricular geral deste experimento direcional.

5.5.7 Estimação de ganhos: problema de regressão

Ao se tratar a estimação de ganhos como um problema de regressão, objetiva-se treinar modelos de regressão para prever a variável de resposta. O conjunto de preditores é escolhido considerando sua correlação com as respostas dos participantes. Daí se segue com o treinamento e a validação cruzada de um modelo de regressão adequado.

Correlações

Um dos retornos do teste piloto foi o de se poder escolher não correlacionar a variável de resposta com as diferenças interauriculares de tempo e nível de pressão sonora, pois estas têm distribuição bivariada, estimada com base nas medidas de SPL feitas pelo HATS para fontes sonoras reais (alto-falantes) nas direções testadas. A eliminação destas métricas foi contrabalançada por um modelo de inibição biauricular encontrado em literatura recente.

As sensibilidades de *loudness* estimadas por [Sivonen e Ellermeier \(2006\)](#) foram calculadas pela fórmula de somatório biauricular de [Robinson e Whittle \(1960\)](#). Mais recentemente, uma extensão do modelo de *loudness* de [Glasberg e Moore \(2002\)](#) incorporando inibição biauricular foi publicada por [Moore et al. \(2016\)](#). Sendo $INH_{IPSI}(i)$ o fator pelo qual o *loudness* de curta duração do sinal ipsilateral é reduzido pelo sinal contralateral, o modelo de inibição biauricular foi proposto da forma:

$$INH_{IPSI}(i) = \frac{2}{\left[1 + \left\{ \operatorname{sech} \left(\frac{N'_{CONTRA}(i)_{suavizado}}{N'_{IPSI}(i)_{suavizado}} \right) \right\}^\gamma \right]} \quad (5.44)$$

onde N'_{CONTRA} e N'_{IPSI} são vetores compostos por níveis de *loudness* de curta duração para os ouvidos contra e ipsilaterais, e $\gamma = 1,598$. Os valores de *loudness* de curta duração do modelo de Moore da [Equação 3.34](#) $N'_L(i)$, para o ouvido esquerdo, e $N'_R(i)$, para o ouvido direito, são então divididos por $INH_L(i)$ e $INH_R(i)$, respectivamente. O valor de γ é arbitrado de tal maneira que, para sons dióticos, $[\operatorname{sech}(1)]^{1,598} = 0,5$, e um som dicótico é estimado uma vez e meia mais perceptivamente intenso do que seu equivalente monoauricular ([MOORE et al., 2016, p. 5](#)).

Valores de *loudness* de curta duração ITU-R BS.1770 foram calculados a partir dos sinais biauriculares gravados pelo HATS na sala de escuta crítica – onde se deu o experimento – com o objetivo de se obter um valor geral de *loudness* de curta duração após o estágio de inibição. Assim como na fórmula de predição de Robinson e na métrica de coeficiente de correlação interauricular, o cálculo dos valores de inibição é orientado à direção da fonte. Gráficos de espalhamento contendo as correlações dos valores experimentais de DLS com as diferenças estimadas pelo somatório biauricular de ganhos 3 dB e 6 dB, com a impressão espacial ($1 - \text{IACC}$) e com o modelo de inibição biauricular de Moore, são exibidos na [Figura 5.47](#).

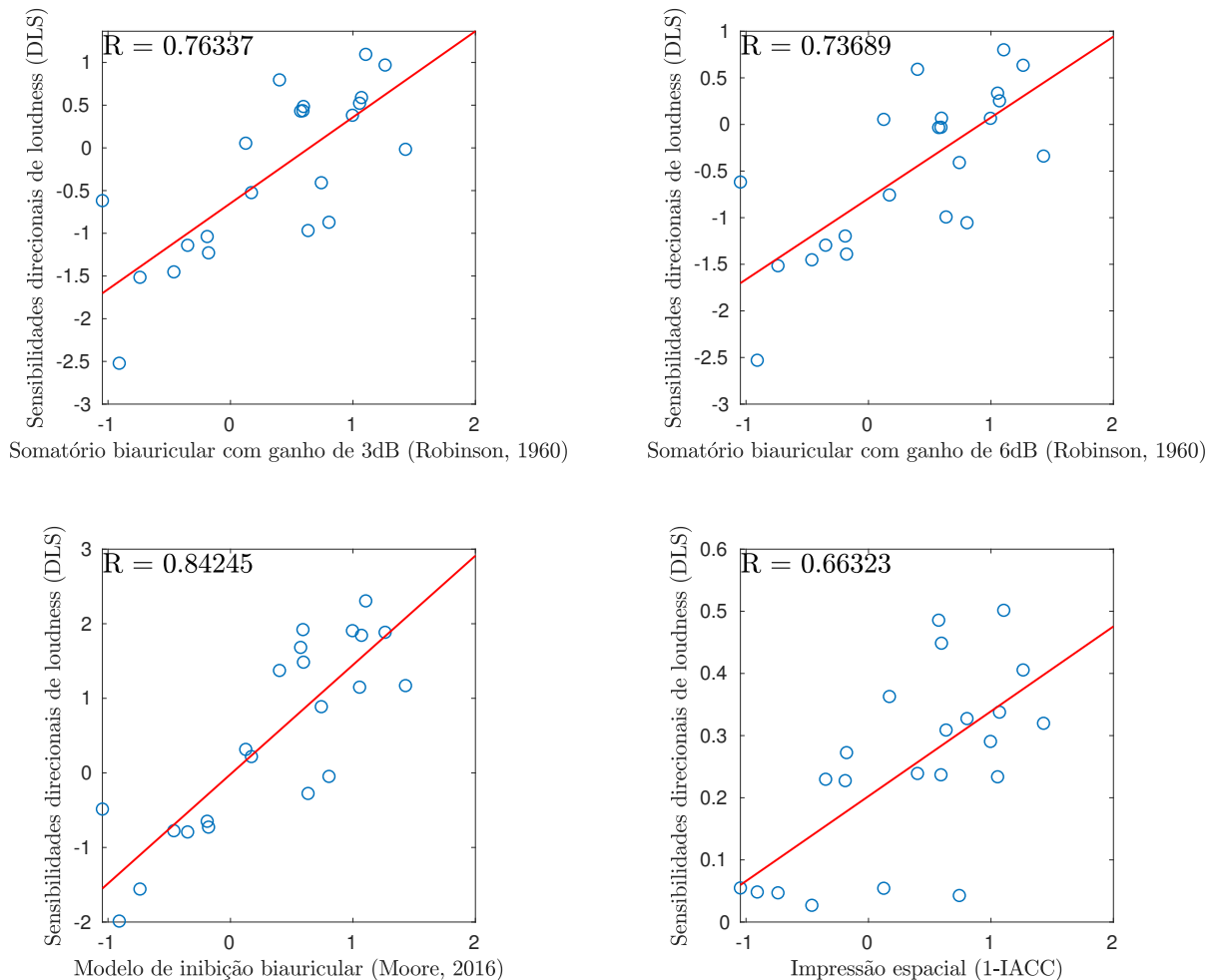
Desta vez, ao se tomar as respostas dos participantes pelas suas médias ao longo do total de participantes e repetições e correlacioná-las com as métricas de áudio espacial derivadas das medidas biauriculares de SPL, as correlações são fortes o suficiente para seguir confortavelmente na escolha destas métricas como preditores do modelo de regressão. Todavia, note que os gráficos de espalhamento das duas variações da fórmula somatório biauricular de Robinson com ganhos de 3 dB e 6 dB são muito semelhantes, sugerindo que ambos seriam preditores redundantes. Destarte, em vez de se usar ambas as versões da fórmula como preditores, optou-se por usar a fórmula somente uma vez, porém com o ganho experimental de 3,54 dB estimado na [subseção 5.5.6](#).

Treinamento

A etapa de treinamento deu-se num esquema de validação cruzada em k -dobras, no qual os dados são particionados em k conjuntos disjuntos ou dobras. Para cada dobra, o modelo é treinado com as observações fora da dobra, e o desempenho é avaliado com os dados dentro da dobra. Então o erro de teste médio é calculado ao longo de todas as dobras. Cinco foi o número de dobras em todas as sessões de treinamento.

Uma série de tipos conhecidos de modelos de regressão – com algumas variações – foram treinados: regressão linear, árvores de regressão, Máquinas de Vetor de Suporte (SVMs) (ver [subseção 4.2.1](#)), Modelo de Regressão de Processo Gaussiano (GPR), e conjuntos de árvores de regressão. Subtipos de regressão linear incluíram modelos com e sem interações, com função objetivo

Figura 5.47 – Correlações com o somatório biauricular de Robinson com ganhos de 3 dB e 6 dB, com a impressão espacial e com o modelo de inibição biauricular de Moore.



Nota – Os gráficos de espalhamento das duas variações da fórmula de Robinson são muito parecidos, o que os caracterizaria como preditores redundantes se levados em consideração num modelo de regressão.

Fonte: Elaborada pelo autor.

robusta para fazer com que o modelo fosse menos sensível a *outliers*, e com remoções passo a passo dos termos de um modelo completo. Subtipos de árvores de regressão variaram no tamanho da folha (4, 12 e 36 saltos). Subtipos SVM diferenciaram-se nas funções de *kernel*: linear, quadrático, cúbico e gaussiano, com variâncias do *kernel* de 0,5, 2 e 8. Os subtipos GPR também se diferenciaram nas funções de *kernel*: racional quadrático, exponencial ao quadrado e exponencial. Já os conjuntos de árvores de regressão podem também ser considerados como variações das árvores de regressão, com impulso de mínimos

quadráticos (*boosted*) ou agregação com recursos próprios (*bagged*).

Em termos das características escolhidas, além de se atacar uma possível redundância com a eliminação de uma das versões da fórmula de Robinson (e considerando uma versão única com o ganho biauricular estimado para este experimento), foi considerada a transformação das características com a mesma Análise de Componentes Principais (PCA) usada na [subseção 4.2.1](#) para reduzir a dimensionalidade do espaço de predição. PCA foi configurada para manter um número de componentes principais suficiente para explicar 95% da variância total.

O desempenho dos modelos foi avaliado pela conferência da Raiz Média Quadrática dos Erros (RMSE) nos conjuntos de validação cruzada. O RMSE é o de todas as observações, contando cada observação quando esta estava fora da dobra. O conjunto completo de modelos testados com e sem transformação de características (PCA) e seus valores de RMSE estão listados na [Tabela 5.19](#). Com base nas sessões de treinamento, o modelo de regressão linear padrão foi escolhido por sua simplicidade e pelo menor erro apresentado na etapa de treinamento (RMSE = 1,7872).

Regressão linear

O modelo resultante pode ser escrito da forma:

$$y_i = \alpha + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \varepsilon_i \quad (5.45)$$

onde y_i são as predições da variável de resposta, α é o termo constante de interceptação, $x_{1,i}$ é o preditor oriundo da métrica de somatório biauricular para a i -ésima direção, $x_{2,i}$ é o preditor oriundo da métrica de inibição biauricular para a i -ésima direção, $x_{3,i}$ é o preditor oriundo da métrica de impressão espacial para a i -ésima direção, e ε_i é o resíduo do modelo. O termo de interceptação ($\alpha = -0,302$) e os betas ($\beta = [-0,262; 0,597; 0,980]$) foram estimados a partir de 523 observações, e o coeficiente β_2 foi estimado de modo muito significativo a um nível de confiança de 95% conforme [Figura 5.48](#).

Um gráfico de resposta, que ilustra os dados de treinamento em conjunto com as predições do modelo, é mostrado na [Figura 5.49](#). A aparência de escadas irregulares do gráfico de resposta está relacionada a como os dados foram

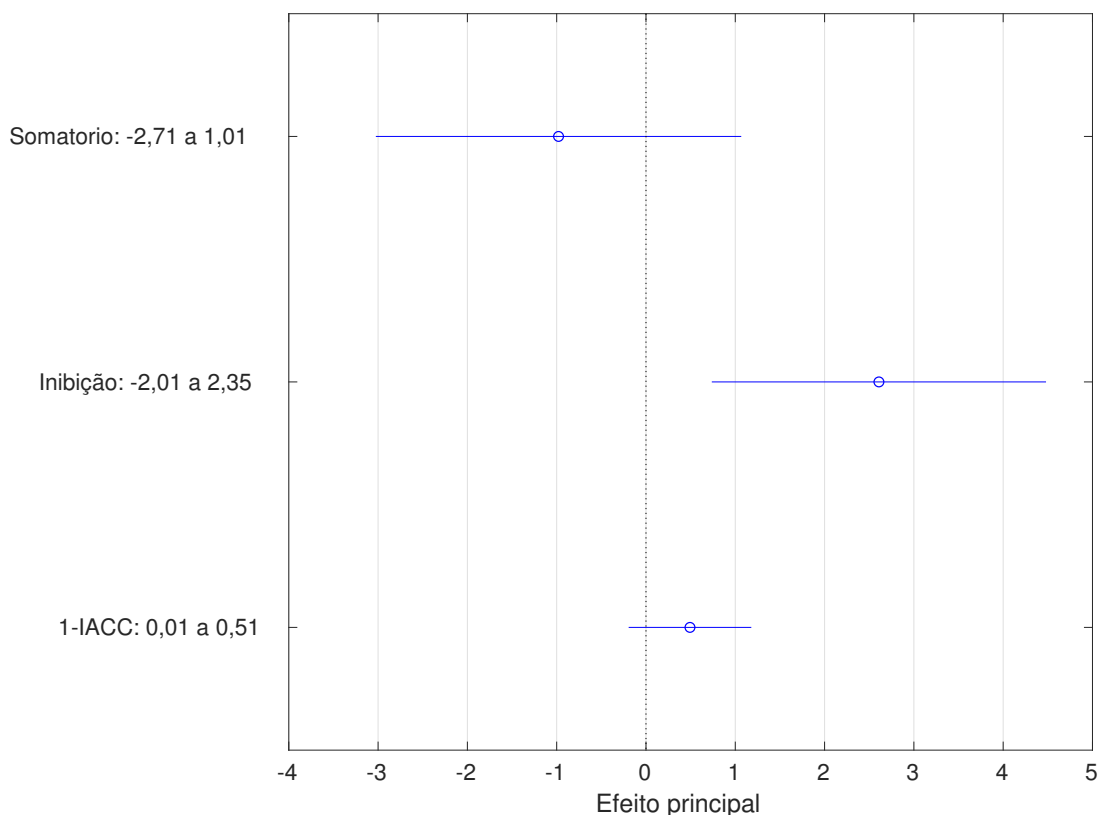
Tabela 5.19 – Desempenho de modelos de regressão treinados. valores de RMSE inferiores a 1,79 estão em negrito.

Modelos de regressão	Valores RMSE	
	Dados diretos	com PCA
Regressão linear	1,7872	1,7907
Regressão linear com interações	1,7876	1,7907
Regressão linear robusta	1,7880	1,7913
Regressão linear passo a passo	1,7918	1,7907
Árvore de regressão fina	2,1651	2,0299
Árvore de regressão média	2,0112	1,8556
Árvore de regressão grosseira	1,8613	1,8306
SVM kernel linear	1,7943	1,8003
SVM kernel quadrático	1,8124	1,7978
SVM kernel cúbico	1,8577	6,3012
SVM kernel gaussiano fino ($\sigma = 0.5$)	1,8613	1,8311
SVM kernel gaussiano médio ($\sigma = 2$)	1,8020	1,7991
SVM kernel gaussiano grosseiro ($\sigma = 8$)	1,7925	1,7976
Conjunto de árvores (<i>boosted</i>)	1,8435	1,7923
Conjunto de árvores (<i>bagged</i>)	1,8488	1,8339
GPR exponencial quadrática	1,7912	1,7931
GPR exponencial	1,7890	1,7979
GPR racional quadrática	1,7917	1,7931

organizados. Cada “degrau” de predição corresponde a uma volta completa de participantes e repetições para um único par azimute/elevação no conjunto de dados de treinamento, de tal forma que as estimativas numa mesma faixa de observações não são muito distintas umas das outras. As médias dos valores em cada passo resultou nos ganhos estimados da [Tabela 5.20](#).

Quando se compara as estimações da regressão linear na [Tabela 5.20](#) com as estimações da minimização restrita da [Tabela 5.17](#), a conclusão a que se chega é que as novas estimações são melhores que as anteriores. Desta vez, somente uma pequena correção do zero foi feita e nenhuma normalização foi necessária. Além do mais, todos os ganhos estão na faixa de $\pm 1,5$ dB e os efeitos de elevação nos ganhos direcionais estimados estão mais claramente definidos, o que era o objetivo da análise de dados desde o princípio. Logo, estes ganhos direcionais irão substituir a ponderação direcional do algoritmo ITU-R BS.1770 no modelo proposto de *loudness* direcional.

Figura 5.48 – Gráfico de efeitos da regressão linear.



Nota – A variação do valor do somatório biauricular de $-2,71$ a $2,35$ resulta num decremento estimado do efeito do preditor na variável de resposta da ordem de -1 , situado no intervalo entre -3 e 1 com 95% de confiança. Leia-se da mesma forma para os preditores “inibição biauricular” e “impressão espacial” (1-IACC).

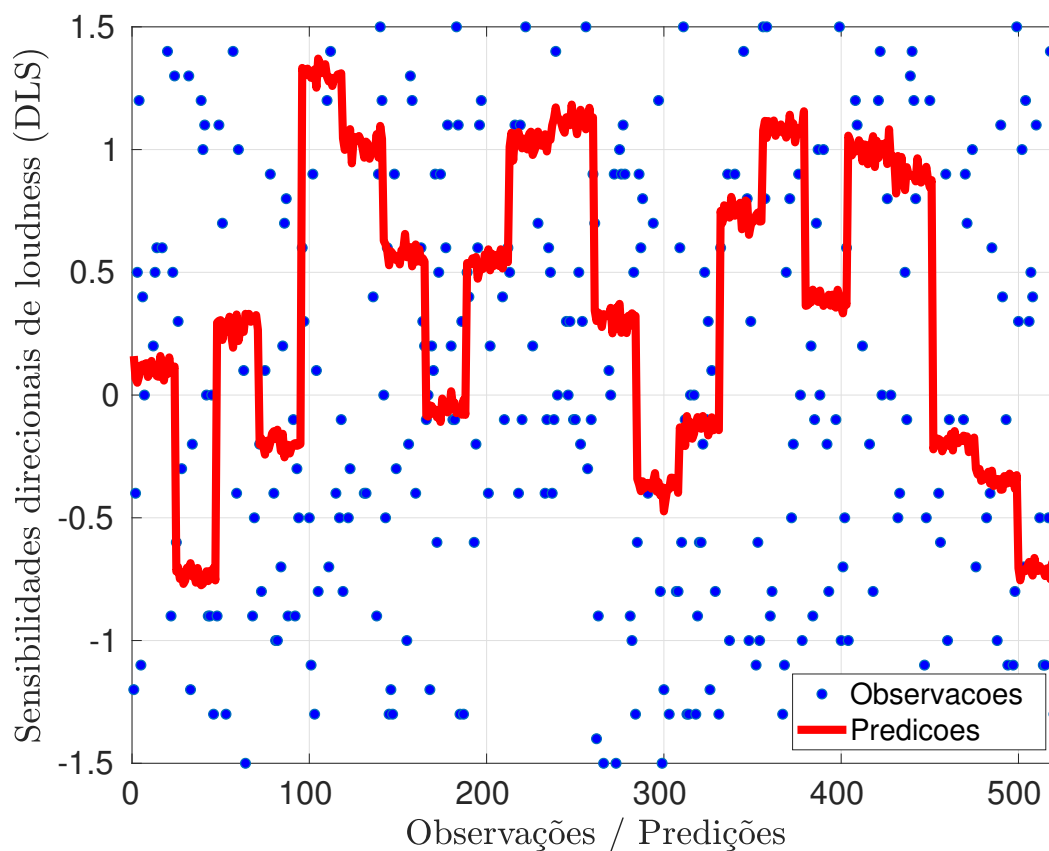
Fonte: Elaborada pelo autor.

5.5.8 Modelo de loudness ITU-R como função da direção

Após substituição dos pesos direcionais, os modelos de *loudness* foram comparados. As diferenças nas medidas de *loudness* entre o modelo ajustado e o algoritmo BS.1770 são ilustradas na Figura 5.50. Note que a linha tracejada de referência representa as medidas BS.1770 e as duas irregularidades na linha tracejada correspondem aos ganhos de $+1,5$ dB nas incidências ipsilaterais.

A comparação é feita ao se observar quando as previsões dos modelos se situam no interior dos intervalos de confiança dos participantes para cada direção. Os acertos ficaram empatados no plano inferior: 1 acerto em 3 direções para ambos os modelos. No plano horizontal, o modelo proposto teve melhor desempenho: 9/10 acertos contra 7/10 para o modelo original. No plano de topo

Figura 5.49 – Gráfico de resposta do modelo. Cada “degrau” de predição corresponde a um conjunto total de participantes e repetições para uma única direção no conjunto de dados de treinamento.



Fonte: Elaborada pelo autor.

porém, foi onde o modelo ajustado fez diferença: 8/9 acertos contra 3/9 do modelo ITU-R.

Por outro lado, o modelo ajustado perde no plano médio vertical: 1/4 acertos contra 4/4 do modelo original. Contudo, era de se esperar que o modelo ajustado desempenhasse mal no plano no qual as dicas de localização são menos discrimináveis. Conseqüentemente, os participantes apresentaram maior variabilidade neste plano durante o teste de escuta. Logo, é até confortável o fato de o desempenho do modelo BS.1770 ser melhor que o do modelo ajustado neste plano em particular.

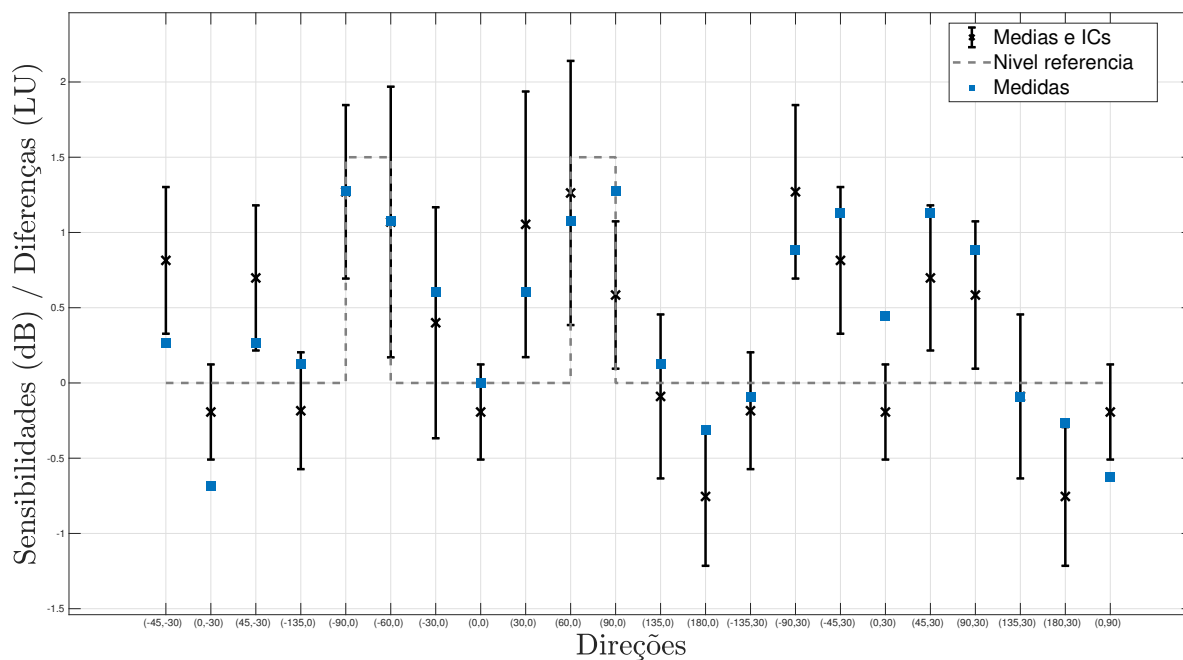
Tabela 5.20 – Ganhos direcionais estimados pela solução de um problema de regressão.

Azimute θ ($^{\circ}$)	Elevação ϕ ($^{\circ}$)	Ganho g (dB)
-45	-30	0,26
0	-30	-0,68
45	-30	0,26
-135	0	0,12
-90	0	1,28
-60	0	1,08
-30	0	0,60
0	0	0,00
30	0	0,60
60	0	1,08
90	0	1,28
135	0	0,12
180	0	-0,31
-135	30	-0,10
-90	30	0,88
-45	30	1,13
0	30	0,45
45	30	1,13
90	30	0,88
135	30	-0,10
180	30	-0,27
0	90	-0,63

5.6 Medição de Loudness para Áudio Espacial (revisitada)

Nas últimas três seções, buscou-se um conjunto de parâmetros para se perseguir com a estratégia de aprimoramento do modelo ITU-R proposta na [subseção 5.2.1](#). A ideia da última seção deste capítulo é lançar luz sobre como um modelo de *loudness* orientado a objetos de áudio funcionaria, propor um modelo consolidado e comparar seu desempenho não somente com a fortuna crítica de modelos de *loudness*, como também com a primeira proposta de modelo para áudio imersivo feita na [seção 4.3](#).

Figura 5.50 – Diferenças entre medidas objetivas plotadas contra as respostas dos participantes.



Nota – Saltos nos intervalos de azimute $[[60^\circ, 120^\circ]]$ se devem aos ganhos de +1,5 dB no modelo ITU-R BS.1770.

Fonte: Elaborada pelo autor.

5.6.1 Áudio baseado em objetos

A lista de recomendações adicionais do documento ITU-R BS.1770-4 possui dois itens categóricos (ITU-R, 2015b, p. 2):

1. Deve-se considerar possível necessidade de atualização desta Recomendação para o caso do surgimento de novos algoritmos de *loudness* cujo desempenho seja significativamente melhor que o do algoritmo especificado no Anexo 1 e no Anexo 3;
2. Esta Recomendação deverá ser atualizada quando novos algoritmos habilitarem a medida de *loudness* em programas de áudio baseado em objetos e de áudio baseado em cenas.

Estes são sinais claros de que o modelo de *loudness* do ITU-R foi concebido para operação estrita em áudio multicanal e deverá ser repensado num contexto de áudio imersivo. O Relatório BS.2266 do ITU-R (2014b) mapeou

três representações para sistemas avançados de reprodução de áudio: baseados em canais, baseados em objetos e baseados em cenas.

Na representação baseada em canais, os sinais oriundos dos microfones são mixados para um número pré-definido de canais e cada canal está associado a um alto-falante específico. Tanto o trabalho de produção, quanto as redes de radiodifusão e os sistemas de reprodução são definidos por um conjunto de alto-falantes e suas posições. Exemplos: estéreo, 5.1, 7.1, 9.1 ou 22.2.

Já no conceito baseado em objetos, a cena é representada por sinais que, individualmente ou combinados, configuram objetos de áudio. Estes são acompanhados de metadados que possibilitam a renderização dos objetos de áudio da forma mais adequada ao sistema e ao ambiente de reprodução. Dentre os algoritmos de renderização de objetos de áudio para diferentes configurações de alto-falantes estão o Sistema Vetorial de Panorama por Amplitude (VBAP) – bastante trabalhado na [seção 5.5](#) – e a Síntese de Campo de Onda (WFS).

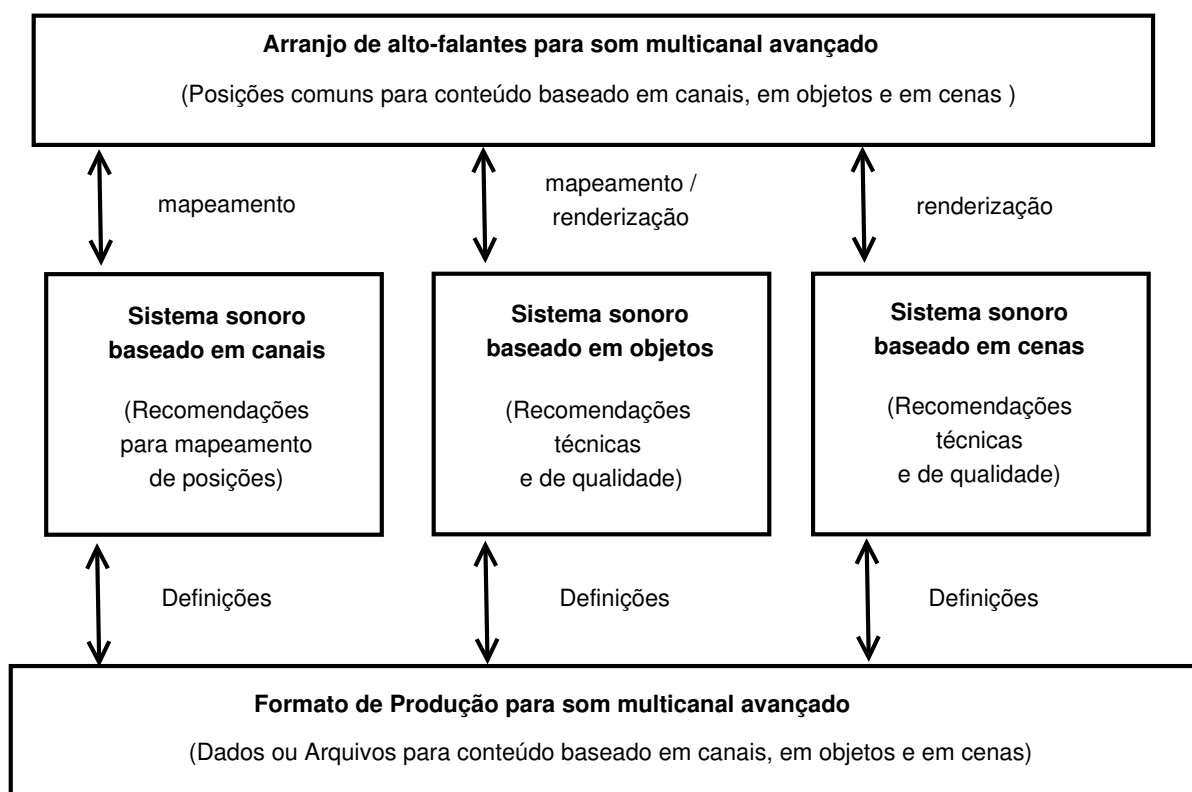
O terceiro e último conceito é o baseado em cenas. A cena sonora é representada por um conjunto de sinais que caracterizam o campo sonoro capturado e que são independentes das posições dos alto-falantes. Exemplos de codificação são o Formato-B e o *Ambisonics* de Alta Ordem (HOA). A relação de desenvolvimento das representações é ilustrada na [Figura 5.51](#).

Todavia, como discutido na [subseção 5.2.2](#), a configuração de uma cama de canais com alguns objetos de destaque, como diálogo controlável separadamente e sons específicos em movimento, é a que mais se aproxima de uma mixagem sonora para UHD TV. O foco dos radiodifusores num modelo híbrido entre o áudio multicanal e o baseado em objetos, é refletido de certa maneira nas próprias questões do grupo de estudo nos termos de referência em [ITU-R \(2014e\)](#) e na reformulação das questões em [ITU-R \(2016b\)](#). Por esse motivo, esta pesquisa se dirigiu rumo a este hibridismo na representação de sistemas avançados de áudio.

Parâmetros posicionais de áudio baseado em objetos

No Áudio Baseado em Objetos (OBA), o requisito principal para se permitir que diferentes tipos de áudio sejam distribuídos, por arquivos ou *streaming*, é de que qualquer que seja o formato do áudio utilizado, deve haver coexistência

Figura 5.51 – Arcabouço dos sistemas futuros de radiodifusão



Fonte: Adaptada de [ITU-R \(2014b, p. 3\)](#).

com metadados que o descrevam. Cada faixa individual num arquivo ou numa transmissão deve estar apta a ser corretamente renderizada, processada ou distribuída conforme os metadados que a acompanhem. O Modelo de Definição de Áudio (ADM) do [ITU-R \(2017\)](#), ao contrário do que acontece nas codificações MPEG-H 3D *Audio* e Dolby AC-4, é um padrão aberto de metadados para garantir compatibilidade entre sistemas, e por isso é a referência aqui.

Um elemento de metadados deste padrão é de particular interesse: *audioBlockFormat*. Ele representa uma única sequência de amostras com parâmetros fixos de formato de canal – inclusive posição – dentro de um intervalo de tempo específico. Uma sequência de elementos *audioBlockFormat* descreve como o formato do canal de áudio varia com o tempo. Para um áudio de localização estática, um único elemento *audioBlockFormat* é suficiente para descrever sua posição ([ITU-R, 2017](#)).

Quando o formato é do tipo OBA (*audioChannelFormat.typeDefinition == Object*), os sub-elementos posicionais do elemento *audioBlockFormat* contêm

atributos de coordenadas polares: azimute, elevação e distância⁹, como disposto na Tabela 5.21. Um exemplo de código XML descrevendo um objeto com coordenadas polares (r, θ, ϕ) onde $r = 0,9$ vezes a distância absoluta fonte-ouvinte, $\theta = 22,5$ graus à esquerda $\phi = 5,0$ graus acima seria da forma:

```
<audioBlockFormat ...>
<position coordinate="azimuth">-22.5</position>
<position coordinate="elevation">5.0</position>
<position coordinate="distance">0.9</position>
</audioBlockFormat>
```

Tabela 5.21 – Sub-elementos posicionais do elemento *audioBlockFormat* para OBA.

Sub-elemento	Atributo	Descrição	Unidades	Exemplo	Quantidade	Default
position	coordinate=azimuth	azimute θ da localização sonora	Graus ($-180 \leq \theta \leq 180$)	-22.5	1	
position	coordinate=elevation	elevação ϕ da localização sonora	Graus ($-90 \leq \phi \leq 90$)	5.0	1	
position	coordinate=distance	distância r da origem	$abs(r)$	0.9	0 or 1	1.0

Fonte: Adaptada de ITU-R (2017, Tabela 14).

Todos os testes de escuta realizados nesta pesquisa para investigar os efeitos de azimute, elevação e distância¹⁰ foram executados com o único propósito de se adquirir dados subjetivos para melhor preparar o modelo ITU-R para medir *loudness* de objetos sonoros com localização específica. A ideia geral está em tirar vantagem dos metadados do objeto para se computar medidas de *loudness* mais próximas dos julgamentos dos ouvintes.

5.6.2 Proposta de modelo para objetos sonoros

Um diagrama em blocos do modelo de *loudness* proposto está na Figura 5.52. É uma versão alternativa do modelo ITU-R BS.1770 com curvas

⁹ A medida de distância é normalizada, porque a informação das distâncias dos alto-falantes à origem é raramente usada (ITU-R, 2017), mas o valor absoluto de distância está disponível no elemento *audioPackFormat*.

¹⁰ O efeito da reverberação foi investigado em razão de como ele interage com o relacionamento entre *loudness* e distância, e não porque exista uma correspondência direta com os metadados ADM. Ainda assim, o tempo de reverberação é um atributo da sala de reprodução e se provou útil no cômputo de medidas de *loudness*, quando conhecido.

de ganho como funções de distância e reverberação, obtidas a partir de dados experimentais, e a ponderação direcional é substituída por novas estimações de ganho que levam em consideração o efeito da elevação.

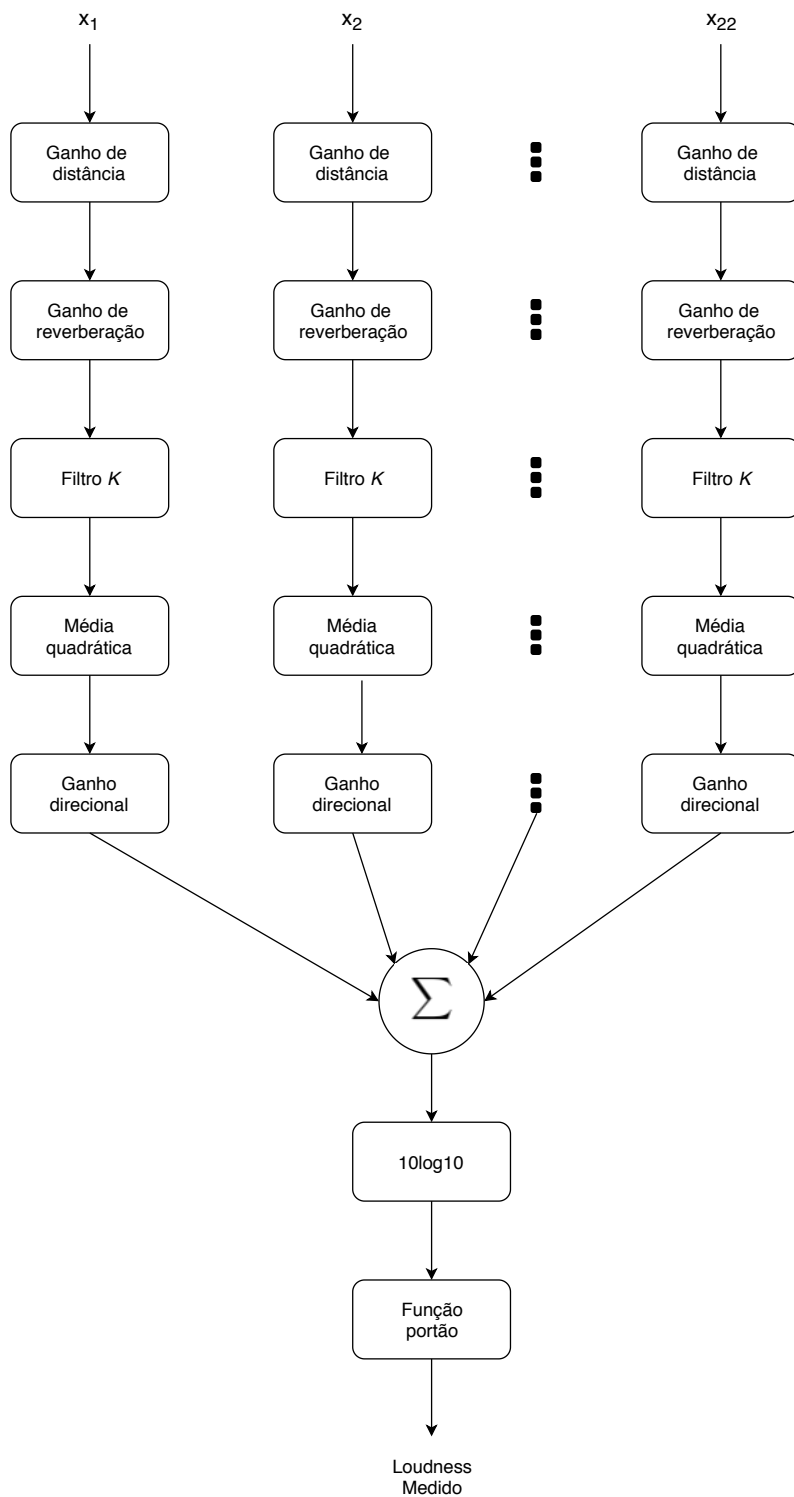
Cada elemento de modificação foi testado contra conjuntos particulares de estímulos que variavam entre si somente no atributo sob teste: o mesmo item de programação foi reproduzido a diferentes distâncias fonte-ouvinte, em diferentes salas e coordenadas. Agora o modelo completo deve ser testado com uma programação multicanal real contra a fortuna crítica de modelos de *loudness*, e isto é feito revisitando-se os dados apresentados na [seção 4.3](#). As respostas do teste de escuta e o material de programação são os mesmos obtidos por [Francombe *et al.* \(2015a\)](#) e gentilmente cedidos pelos autores.

5.6.3 Avaliação do modelo

Assim como nos experimentos aqui relatados, alguns sinais idênticos aos de referência foram incluídos nos testes de escuta feitos por [Francombe *et al.* \(2015a\)](#). Os intervalos de confiança de 95% das respostas dos participantes se situaram aproximadamente em intervalos de $\pm 0,5$ dB a $\pm 1,0$ dB, o que é maior que a diferença no limite do observável conhecida para o *loudness*: (0,5-1,0 dB) para o ruído de faixa larga ([MOORE, 2012](#), p. 144). Considerando que a comparação do sinal de referência com ele mesmo é tido como o melhor caso, há de se reconhecer que a tarefa de casamento de *loudness* é razoavelmente difícil de ser executada.

Neste mesmo caso de comparação, nenhum participante produziu uma média que fosse significativamente diferente de zero, indicando que os participantes estiveram similarmente aptos a executar a tarefa. Logo, os resultados de todos os participantes foram incluídos na análise de dados feita pelos autores. Contudo, ao se considerar todos os dados, alguns intervalos de confiança são amplos (até aproximadamente $\pm 2,5$ dB), indicando a presença de alguns erros consideráveis. Levando em conta a própria experiência adquirida com os testes de escuta feitos para esta pesquisa, esta constatação só reforça o entendimento do parágrafo anterior de que o casamento de *loudness* é sim uma tarefa difícil, até no mais simples dos cenários (ver [Figura 4.9](#)).

Figura 5.52 – Diagrama em blocos do modelo de *loudness* proposto: uma versão modificada do modelo ITU-R BS.1770 com correções de ganho como funções da distância e da reverberação, além de nova ponderação direcional levando em conta o efeito da elevação.



Fonte: Elaborada pelo autor.

Material de programação

Os itens de programação a seguir serão os utilizados na avaliação do modelo consolidado (mais detalhes sobre a criação dos itens de programação foram relatados por [Francombe et al. \(2015b\)](#)). Cada trecho possui duração de 20 segundos:

- Dois itens (um quinteto de sopros e um quinteto de *jazz*) gravados ao vivo no Estúdio 1 da Universidade de Surrey com técnicas de captura para diferentes métodos de reprodução;
- Uma gravação pop multi-faixa mixada para diferentes métodos de reprodução;
- *Big band* (gravação multi-microfone feita no Royal Albert Hall mixada para diferentes métodos de reprodução);
- Música experimental (Renderização por formato-B de *Rotating psychoacoustic tuning curves* por Florian Hecker, decodificada para diferentes métodos de reprodução); e
- Trecho cinematográfico (Clipe 5.1 do filme *007 – Operação Skyfall* manualmente remixada para menos (*downmixed*) e para mais canais (*upmixed*), e diferentes formatos de reprodução, acrescido de ruído de chuva no formato-B decodificado para diferentes arranjos de alto-falantes).

O material original também incluía a transmissão de uma partida de futebol mixada para diferentes métodos de reprodução, incluindo a decodificação *Ambisonics* de um microfone *Soundfield* nas arquibancadas, mas este item de programação em particular não pôde ser compartilhado devido aos direitos de reprodução.

Todos os programas foram reproduzidos por uma esfera de 1,9 m de raio na mesma sala de escuta crítica onde os experimentos desta pesquisa foram realizados ($RT_{60} = 0,22$ s). Estes valores, juntamente com as posições dos alto-falantes em cada um dos sistemas de reprodução, alimentaram o modelo de *loudness* proposto nas suas medições.

Fortuna crítica comparada

Assim como na [seção 4.3](#), os resultados do teste perceptivo anteriormente descrito foram utilizados para testar um conjunto de modelos de *loudness* variando em complexidade de medidas puramente físicas a modelos de maior carga psicoacústica. A relação dos modelos para comparação com a proposta desta seção, estão relacionados abaixo:

- Nível RMS da soma canal a canal de gravações biauriculares;
- dB LAeq(20s) medido com um *NTI AL1 Acoustilyser*;
- ITU-R BS.1770-4, na sua última versão para um número irrestrito de canais ([ITU-R, 2015b](#));
- ITU-R BS.1770-3 modificado para entrada biauricular substitutiva do filtro de sombreamento de cabeça elaborado por [Pike e Melchior \(2013\)](#);
- Modelo de *loudness* para sons variantes no tempo de [Glasberg e Moore \(2002\)](#), mais precisamente o descritor de nível máximo de *loudness* de longa duração (LTL);
- Modelo de *loudness* para sons variantes no tempo de [Zwicker e Fastl \(1999\)](#), mais precisamente o descritor de nível máximo de loudness instantâneo; e
- ITU-R BS.1770-3 modificado para um número irrestrito de canais com base num modelo de fonte-imagem de uma sala de reprodução de referência por [Pires et al. \(2017\)](#).

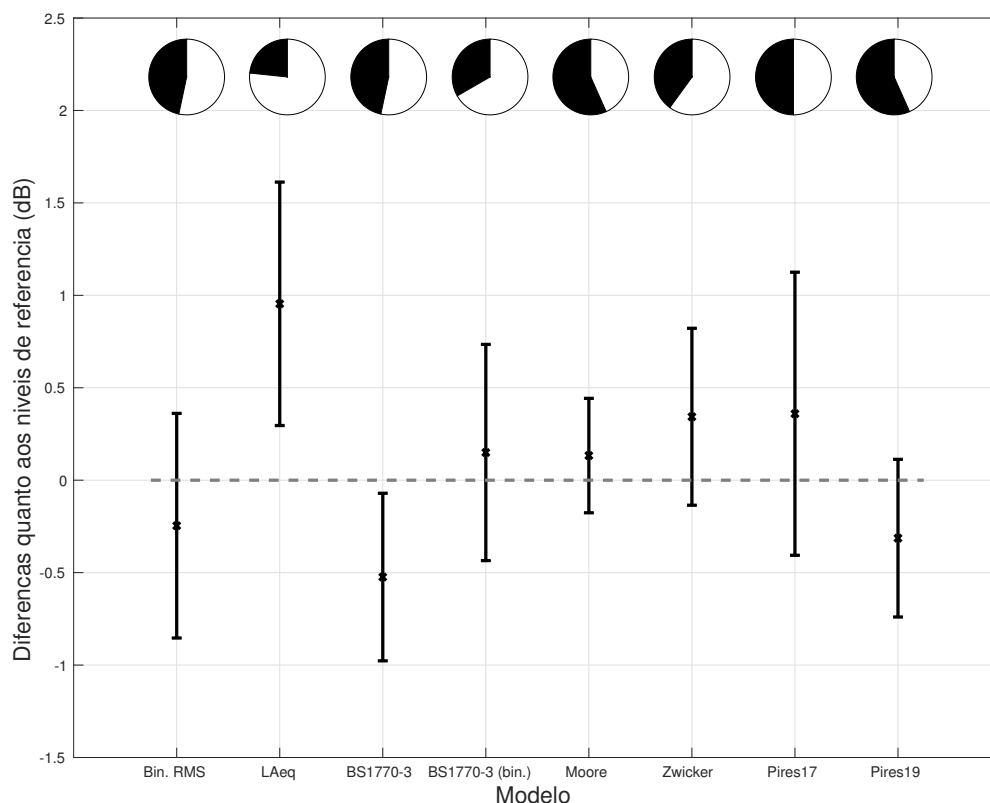
Dentre os modelos listados na comparação realizada na [seção 4.3](#), havia também um cálculo de nível RMS médio por todos os canais. Este modelo de detecção de energia foi excluído desta rodada de comparações devido aos seus intervalos de confiança demasiadamente amplos e, conseqüentemente, seu potencial de diminuir a resolução da variável de resposta nos gráficos a seguir.

5.6.4 Resultados e discussões

A [Figura 5.53](#) ilustra o desvio médio dos modelos em relação ao sistema de referência (5.1) por todos os itens de programação, com intervalos de confiança de 95% calculados com o uso da distribuição normal, para cada modelo de

loudness. Um valor positivo indica que o modelo estimou que a reprodução foi mais intensa que a referência. É possível notar que o desvio das médias em relação à referência não é superior a ± 1 dB em nenhum dos casos. Os modelos são consistentes no geral, com o modelo proposto (Pires19) figurando entre o estado da arte dos seus pares (BS.1770, Moore e Zwicker) e exibindo um intervalo de confiança de 95% de aproximadamente $\pm 0,5$ dB, dentro da JND documentada para o *loudness*. O modelo de Moore e a implementação biauricular do algoritmo BS.1770-3 foram os que tiveram as médias mais próximas do zero de referência, sugerindo um melhor desempenho geral a princípio. Contudo, os modelos BS.1770-4, Zwicker e Pires19 obtiveram intervalos de confiança menores e médias não significativamente diferentes do valor de referência.

Figura 5.53 – Diferenças entre as médias dos resultados dos modelos para o sistema de reprodução de referência (5.1) e outros métodos de reprodução.



Nota – Os segmentos escuros dos gráficos em pizza estão relacionados com o percentual dos resultados dos modelos de previsão que se situaram no interior dos intervalos de confiança dos dados subjetivos de [Francombe et al. \(2015a\)](#).

Fonte: Elaborada pelo autor.

Há também na [Figura 5.53](#), gráficos pizza cujos segmentos escuros estão relacionados com o percentual das respostas dos modelos situadas no interior dos intervalos de confiança dos resultados do teste de escuta, ou seja, casos nos quais pode-se dizer que o modelo teve um desempenho tão bom quanto o dos ouvintes. Os modelos Pires19 e Moore desempenharam melhor neste aspecto, com 57% de predições alinhadas com a percepção dos ouvintes, seguidos pelo modelo Pires17 com 50%. Os modelos binauriculares BS.1770-3 e RMS também tiveram um desempenho próximo, com alinhamento superior a 45%. Os baixos percentuais de alinhamento com os ouvintes como um todo indicam que, embora os modelos sejam razoavelmente precisos, isto é, tenham aparentemente um baixo erro médio, ainda há uma boa diferença de desempenho se comparados a um painel de ouvintes treinados.

As estatísticas gerais do desempenho de cada modelo são dadas na [Tabela 5.22](#). A Raiz Média Quadrática do Erro (RMSE) é uma medida de qualidade de ajuste entre os dados do teste de escuta e as estimações do modelo, usada outras vezes ao longo deste trabalho. Mas desta vez, é usada também uma variante mais adequada denominada Raiz Média Quadrática do Erro insensível a ϵ (RMSE*), especificada na Recomendação P.1401 do [ITU-T \(2012\)](#), que dispõe sobre métodos, métricas e procedimentos estatísticos de avaliação, qualificação e comparação de qualidade de modelos objetivos de predição, e recomendada para comparações como esta: avaliação estatística no contexto de incerteza subjetiva. Em sua definição, “epsilon” é o próprio intervalo de confiança de 95% em torno da média das respostas dos participantes. A métrica é calculada da mesma forma que o RMSE padrão (ver [Equação 5.25](#)), porém considerando somente as diferenças em relação à média situadas fora do intervalo de confiança no entorno desta média (fora do epsilon). Por fim, a última métrica de comparação é o coeficiente de correlação de Pearson, também usado algumas vezes aqui, que descreve a linearidade da relação entre respostas dos participantes e predições dos modelos.

O modelo de Moore exibiu o melhor desempenho segundo todas as métricas. Pires19, Zwicker e BS.1770-4 tiveram estatísticas muito similares, com $RMSE^* \leq 0,7$ dB, enquanto que o modelo Pires17 teve o pior desempenho em todos os aspectos.

Tabela 5.22 – Estatísticas dos modelos de *loudness* ordenadas por RMSE*.

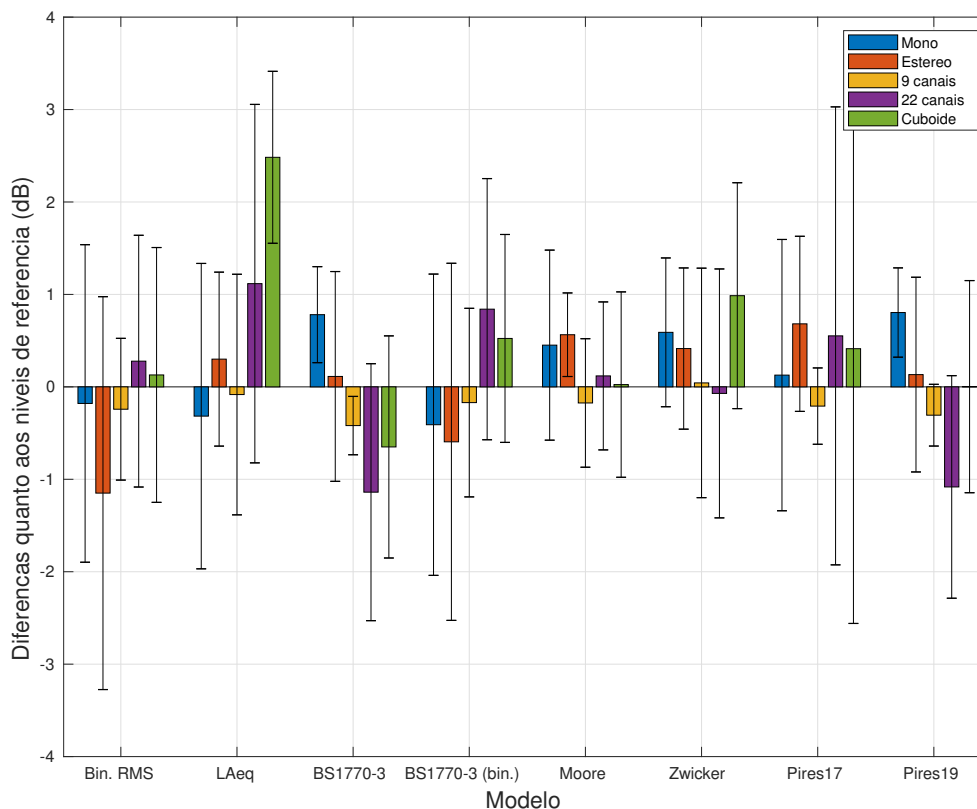
Modelo	RMSE (dB)	RMSE* (dB)	<i>r</i> de Pearson	Tempo médio (s)
Glasberg e Moore (2002)	0,80	0,39	0,9596	≈ 280
Pires (2019)	1,02	0,60	0,9248	< 1
Zwicker e Fastl (1999)	1,13	0,65	0,9201	≈ 10
BS.1770-4 (2015)	1,14	0,70	0,9162	< 1
BS.1770-3 biauricular (2013)	1,41	0,87	0,9252	< 1
RMS biauricular (N/A)	1,46	1,00	0,9262	< 1
LAeq (1979)	1,76	1,24	0,8549	< 1
Pires <i>et al.</i> (2017)	1,74	1,36	0,8348	≈ 19

No que diz respeito aos tempos de execução¹¹, o modelo de Moore foi o mais demorado no cálculo do *loudness* dos extratos de 20 segundos de programação (um tempo de execução médio de 280 segundos por extrato), seguido pelo modelo Pires17 (tempo de execução médio de 19 segundos por extrato) e pelo modelo de Zwicker (tempo de execução médio de 10 segundos por extrato). Todos os demais demoraram menos de um segundo para estimar o *loudness* dos mesmos trechos, com exceção do modelo *LAeq* cujas medições dão-se em tempo real.

Os resultados são subdivididos na [Figura 5.54](#), que mostra os desvios da média em relação ao sistema de referência 5.1 por modelo de *loudness* e sistema de reprodução: mono, estéreo, 9 canais, 22 canais e cuboide *Ambisonics*. Desta vez, os intervalos de confiança de 95% foram calculados com a distribuição *t* de *student*, em razão de um menor número de amostras por subgrupo. Valores significativamente diferentes do zero de referência aparecem na reprodução estereofônica para o BS.1770-3 biauricular, na reprodução em 22 canais para o LAeq, BS.1770-4 e Pires19, e no cuboide *Ambisonics* para o LAeq. O modelo de Moore é o mais consistente em todos os sistemas de reprodução (porém ao custo de um tempo de execução demasiadamente longo, como visto anteriormente). Pela figura, não é possível estabelecer uma relação óbvia entre o sistema de reprodução e os erros de predição, mas tudo indica que o desempenho dos modelos é normalmente inferior para sistemas com mais alto-falantes.

¹¹ Os tempos de execução foram avaliados no mesmo *MacBook Pro* utilizado nos experimentos e com os mesmos códigos MATLAB[®] escritos para esta pesquisa.

Figura 5.54 – Diferenças entre as médias dos resultados dos modelos para o sistema de reprodução de referência (5.1) e outros métodos de reprodução, agrupados por método de reprodução.



Fonte: Elaborada pelo autor.

Comentários

Pelas medidas realizadas e métricas de desempenho apresentadas, foi possível notar a evolução de um modelo de *loudness* puramente baseado em processamento de sinais (Pires17) para um modelo mais orientado à percepção (Pires19). A partir da [Figura 5.53](#), houve melhoramentos no desvio da média em relação à referência de 5 canais e no intervalo de confiança de 95%. O modelo proposto nesta seção é mais preciso, tem menor desvio e é mais próximo das respostas dos participantes se comparado ao modelo proposto na [seção 4.3](#).

A [Tabela 5.22](#) apresenta estes melhoramentos de modo mais compreensível. Os valores de RMSE* – calculados somente com as previsões que caíram fora dos intervalos de confiança dos participantes – mostram que o modelo consolidado desta seção possui a melhor relação de compromisso entre tamanhos de erro e tempos de execução. O erro médio e a correlação com as respostas dos

participantes são similares aos do modelo de Zwicker. As mesmas estatísticas, quando calculadas para o modelo de Pires *et al.* (2017), colocam o modelo como inferior às medidas acústicas feitas com a curva A (LAeq), sem mencionar seu tempo de execução, superior ao modelo de Zwicker.

Este modelo de *loudness* para objetos sonoros, quando comparado com o seu *benchmark* (o modelo ITU-R), também coleciona notas positivas. Levar em consideração a distância, a reverberação e a elevação, traduziu-se num menor desvio da média em relação à referência, num menor erro fora dos intervalos de confiança das respostas dos participantes, e numa melhor correlação com a mesma variável de resposta. Mas é na Figura 5.54 que a melhora é aparente. O modelo proposto teve desempenho um pouco melhor que o ITU-R BS.1770-4 na maioria dos sistemas de reprodução, com exceção do cuboide, no qual o desvio do novo modelo em relação à referência foi de aproximadamente 0 dB, e isso se deve a uma melhor modelagem do fator elevação. Quanto mais elementos variem em elevação no arranjo de alto-falantes, pior é o desempenho dos medidores acústicos (LAeq). E apesar do menor número de canais no cuboide *Ambisonics* em relação ao sistema de 22 canais, todos os alto-falantes no cuboide têm $\phi \neq 0$. Sendo um atributo não levado em conta no modelo ITU-R e considerado no modelo proposto, a elevação fez diferença neste sistema, em particular com relação à consistência das medidas com o sistema de reprodução de referência de 5 canais, e em relação à quantidade de medidas situadas no interior dos intervalos de confiança das respostas dos participantes.

CONCLUSÃO

O advento do algoritmo de *loudness* para a radiodifusão do Setor de Radiocomunicação da União Internacional de Telecomunicações (ITU-R) possibilitou ir além do monitoramento de volume e de picos de programação para um controle psicoacústico da sensação de intensidade, normalizando itens de programação e mitigando abusos de compressão de faixa dinâmica em vinhetas e peças comerciais. As regulações nacionais que se basearam na Recomendação BS.1770 atacam um problema que afeta a Qualidade da Experiência dos consumidores de serviços audiovisuais em todo o mundo, enquanto o próprio ITU-R estuda adaptar a solução para os formatos de áudio de nova geração para a UHD TV.

Esta pesquisa, de caráter regulatório, foi motivada pelas duas perguntas que dão nome à [subseção 5.2.1](#) e [subseção 5.2.2](#): “*Sob que aspectos a norma brasileira de loudness pode ser revisada?*” e “*Como o modelo de loudness do ITU-R pode ser aprimorado para áudio imersivo?*”. Suas investigações, portanto, concentraram-se em duas frentes: i) fazer uma leitura crítica da regulação brasileira e elaborar estratégias para aprimorá-la à luz da experiência internacional e ii) buscar contribuir com as questões de estudo em andamento no ITU-R com uma proposta de melhoramento do algoritmo BS.1770, adaptando-o para os novos formatos de áudio digital para consumo.

Para tanto, fez-se um levantamento bibliográfico detalhado no qual é apresentado um mapa de desenvolvimento de uma medida objetiva de *loudness* desde sua concepção no campo da psicoacústica até sua adoção pela indústria.

Em seguida, foram abertas linhas de investigação para ambas as questões, cujos seguimentos se deram na forma de estratégias de controle e proposições de modelos de medida: sendo um deles puramente baseado em processamento de sinais e o outro mais perceptivamente motivado.

O [Capítulo 2](#) trouxe um mapa de desenvolvimento de uma medida objetiva de *loudness* para fala e música da psicoacústica à radiodifusão. Neste, além de se definir o *loudness* com algum rigor, procurou-se explicitar os efeitos de frequência, de duração e de espacialidade dos quais é dependente, destacando-se a acumulação espectral de *loudness*, a integração temporal de *loudness* e o somatório biauricular de *loudness*. Já o [Capítulo 3](#) passou pela fortuna crítica dos principais modelos de *loudness*, indo desde o modelo clássico de detecção de energia da psicofísica, passando pelos modelos multi-faixa para sinais estacionários e variantes no tempo, pelos modelos de faixa única e suas diferentes ponderações em frequência, até chegar ao padrão ITU-R para conteúdo de radiodifusão, descrevendo os métodos de medidas e descritores auxiliares presentes em outras recomendações.

Primeiramente, o [Capítulo 4](#) teve por objetivo apresentar o histórico recente do tratamento de *loudness* na radiodifusão, os últimos desenvolvimentos sobre o tema acompanhados de uma noção dos sentidos para onde a pesquisa de *loudness* tem progredido, e quais linhas de investigação estão abertas. Em seguida, discorreu sobre as primeiras tentativas experimentais de se perseguir cada um dos objetivos. Uma leitura crítica da norma brasileira de *loudness* motivou o primeiro experimento, que propôs um controle de *loudness* em conteúdo de formato curto baseado em descritores não constantes da norma. Já o segundo experimento perseguiu uma ideia para adaptação do modelo ITU de *loudness* considerando medidas em sistemas imersivos multicanal.

Por fim, o [Capítulo 5](#) fechou algumas das linhas de investigação referentes ao primeiro objetivo e, com base nas conclusões do experimento de controle de *loudness* e nos valores de referência praticados internacionalmente, propôs uma estratégia de atualização da norma brasileira de *loudness* para reforçar sua efetividade. Motivada pelas perspectivas futuras do áudio na radiodifusão e nas questões de estudo no ITU-R, a investigação do segundo objetivo focou em ajustes do modelo BS.1770 baseados em parâmetros posicionais. Os ajustes

foram obtidos a partir das respostas de participantes em testes de escuta, e juntos compuseram um modelo consolidado de *loudness* para objetos sonoros, testado em conteúdo multicanal e comparado com o estado da arte dos modelos apresentados no [Capítulo 3](#).

6.1 Principais Contribuições

Um exame cuidadoso da norma brasileira de *loudness* para a radiodifusão revelou algumas inconsistências, das quais destacam-se duas: a ausência de descritores curtos de *loudness* e o uso do descritor “faixa de *loudness*” na caracterização de uma peça de áudio como regular ou ofensora, ambas inadequadas para avaliação de conteúdo de formato curto (comerciais, vinhetas e inserções). O experimento de controle automático de *loudness* para este formato funcionou como uma espécie de “prova de bancada” para a norma, pois o sucesso da proposta dependeu diretamente do uso de descritores “*loudness* de curta duração”, “*loudness* momentâneo” e “pico verdadeiro”, ausentes da redação da Portaria nº 354/2012 do antigo [MC \(2012\)](#). Já a experiência internacional forneceu *insights* regulatórios importantes com relação aos valores de referência praticados para os descritores de *loudness* no geral, bem como sobre quais descritores são mais adequados para a radiodifusão e para a produção de mídia digital. Isso resultou na elaboração de uma estratégia para averiguação do cumprimento de obrigações das emissoras, que consiste em correções de pontos específicos da Portaria supra mencionada referentes a tolerâncias da medida, unidades, definições, substituição de alguns descritores e inclusão de novos.

A proposta de redução da faixa de tolerância do valor de referência (-23 LKFS) de ± 2 LU independentemente do conteúdo para $\pm 0,5$ LU em conteúdo de estúdio e ± 1 LU em conteúdo ao vivo, tem por objetivo limitar os saltos de intensidade às diferenças no limite do observável documentadas para o *loudness* para sons de faixa larga ([MOORE, 2012](#), p.144), sem riscos de comprometimento de resolução da medida, considerando a tolerância de $\pm 0,1$ LU para medidores em conformidade com as recomendações EBU e ITU-R correspondentes. Desta forma, garante-se que o consumidor não precise ajustar o volume de reprodução entre diferentes itens de programação, entre programação

e intervalos comerciais, ou entre canais de distribuição, pois as inconsistências de *loudness* nesses casos não seriam percebidas pela maioria do público.

Exigir-se limites de faixa de *loudness* somente para itens de programação tem por objetivo mitigar o desconforto causado por faixas dinâmicas largas demais para os níveis de reprodução desejados para o ambiente. Dessa forma, não são demandados recursos de normalização – e fiscalização – para peças publicitárias de formato curto nas quais o problema não é percebido, tampouco medido adequadamente. O parâmetro “Faixa de *Loudness*” não é aplicável a conteúdos de formato curto por ser baseado numa análise estatística de valores de “*Loudness* de Curta Duração”, calculado em janelas de 3 segundos sobrepostas a cada segundo. Para comerciais e vinhetas, isso resulta num conjunto de valores pequeno demais para se obter um resultado significativo. Supondo uma peça publicitária de 30 segundos, é razoável presumir que um vetor de 30 amostras resultaria num teste estatístico de baixa potência, com alta sensibilidade a erros do tipo II, ou seja, de não se rejeitar a hipótese nula (faixa de *loudness* inferior ao limite estabelecido) quando esta for falsa.

As alterações nas duas principais métricas presentes na regulamentação brasileira, somadas às propostas de controle de ceifamento por medidas de pico verdadeiro, e de discussão quanto ao controle da dinâmica de peças publicitárias por meio de descritores curtos de *loudness*, é um material com potencial de subsidiar futura revisão e/ou expansão da norma vigente, vislumbrando a inclusão de outros serviços de entrega de conteúdo de áudio, a exemplo do SeAC e dos serviços de *streaming* no país.

Já no que diz respeito aos formatos avançados de áudio digital, a perspectiva de profissionais da área é de que um modelo híbrido de mixagem composto por uma cama sonora multicanal, com diálogo e alguns outros objetos – estáticos ou em movimento – sobre ela, é o que se avizinha como um paradigma de formato de áudio para o UHD TV, e as questões de estudo do grupo relator de *loudness* no ITU-R refletem isso de certa maneira. Logo, entendeu-se que o aprimoramento do modelo ITU-R para os novos formatos de áudio passa por manter a arquitetura multicanal e imprimir correções quanto aos aspectos posicionais dos objetos sonoros. Mais precisamente distância, azimute e elevação, parâmetros especificados nos metadados do Modelo de Definição de Áudio (ADM) do

ITU-R (2017), bem como nos metadados dos esquemas de codificação MPEG-H 3D Audio e Dolby AC-4. A primeira proposta deste trabalho procurou aumentar a carga psicoacústica do método ITU-R no que diz respeito ao aprimoramento da modelagem dos efeitos espaciais na percepção de intensidade, ao se considerar o agnosticismo em relação ao número e ao posicionamento dos alto-falantes no sistema do usuário, e ao se explorar a incorporação dos efeitos acústicos da sala de reprodução, seja uma sala de referência, seja a própria sala do consumidor virtualmente construída por meio de sua resposta espacial ao impulso. Esta teve um bom desempenho em comparação com outros modelos, porém era puramente uma solução de processamento de sinais e suas leituras não se assemelhavam tanto aos resultados de participantes de testes de escuta. Os potenciais benefícios de um modelo mais perceptivamente motivado, levaram à condução de experimentos de casamento de *loudness* para avaliação de parâmetros posicionais, elaborados durante meu estágio doutoral no Instituto de Gravação de Som da Universidade de Surrey, no Reino Unido, cujos resultados serviram de base para a obtenção de curvas de correção de ganho e nova ponderação direcional para o modelo ITU-R.

No experimento desenvolvido para se investigar a relação do *loudness* com a distância, foi possível observar um efeito muito significativo das distâncias das fontes sonoras nas respostas dos participantes, aliado a um efeito significativo das duas salas de reprodução. Estas observações foram consistentes com a revisão bibliográfica e fundamentaram a noção de que uma correção do modelo baseada nas distâncias das fontes sonoras não poderia ser desacompanhada de uma correção baseada em reverberação, que motivou o experimento seguinte. Neste, feito em seis salas virtuais, o efeito da reverberação foi observado de modo muito significativo, juntamente com um efeito do azimute da fonte sonora também significativo, porém de tamanho bem menor. Esta segunda rodada de observações possibilitou implementar uma correção conjunta do modelo ITU-R baseada em distância/reverberação, como também motivou o experimento seguinte, dado o efeito direcional significativo na variável de resposta.

Para se investigar a relação do *loudness* com a direção da fonte sonora, usou-se de alguma rastreabilidade dos trabalhos progressos que levaram à definição dos pesos direcionais empregados no modelo BS.1770. Na oportunidade,

além das diferenças nos ganhos biauriculares estimados, observou-se que nem as fontes balanceadas (fantasmas), nem as elevações dos alto-falantes, tiveram seus efeitos contemplados nestas estimações. Contudo, assim como no experimento anterior, o efeito da localização das fontes também foi menor frente a outro efeito bem mais significativo (fontes sonoras: reais e balanceadas). Quando os dados foram divididos entre fontes reais e os casos balanceados, o efeito direcional observado foi de tamanho médio para as fontes sonoras reais e de tamanho pequeno para as fontes balanceadas. Isto fez com que a estimação dos ganhos direcionais se desse somente a partir dos dados da primeira categoria de fontes sonoras.

A estimação dos ganhos direcionais contemplando os efeitos de elevação não teve sucesso quando se tentou reproduzir o método de otimização original, mas somente quando se abordou o cálculo dos ganhos direcionais como um problema de regressão, a partir do uso de métricas conhecidas de áudio espacial como preditores do modelo linear escolhido. Esta nova ponderação direcional, combinada com as correções de ganho baseadas tanto em distância da fonte sonora quanto nos tempos de reverberação dos ambientes de reprodução, consolidaram a modificação do modelo ITU-R proposta nesta pesquisa. Em testes com conteúdo imersivo multicanal, o desempenho do novo modelo consolidado para objetos sonoros foi superior tanto em relação à primeira proposta quanto em relação ao *benchmark* da indústria no que tange à similaridade de suas predições com as respostas dos ouvintes. Embora careça de testes com um maior número de formatos de áudio para que seja alçado à categoria de solução padrão, este material se mostrou importante como contribuição para as discussões sobre *loudness* no ITU-R.

6.2 Trabalhos Futuros

Fez parte do planejamento desta pesquisa um esforço para que os experimentos principais dialogassem com os preliminares, como se fossem evoluções naturais destes. Esse objetivo foi atingido, em alguma medida, com a comparação entre o modelo consolidado e a proposta preliminar nas mesmas condições. O mesmo não pode ser dito sobre o controlador de *loudness*, pois não houve

tempo hábil para implementar o modelo consolidado para operação em *streaming*, calculando um vetor de *loudness* momentâneo (ou de curta duração) a partir dos tempos e das posições indicadas pelo elemento *audioBlockFormat* nos metadados que acompanham o áudio. Fornecer medidas de *loudness* em tempo de execução da peça de áudio seria uma boa característica do modelo a ser testada em objetos sonoros de localização dinâmica, dado que o resultado final seria a integração de um vetor cujos valores foram calculados a partir de vários elementos *audioBlockFormat* durante a movimentação espacial do objeto. Testar o modelo proposto sob essas condições, sujeito às regras de controle sugeridas no primeiro experimento preliminar e aperfeiçoadas na estratégia de modificação da norma seria um fechamento adequado para este trabalho, uma vez que a comprovação de seu funcionamento no ADM seria como uma chancela de funcionamento também nos esquemas de codificação MPEG e Dolby. Deixa-se esta ideia portanto como forte sugestão de desenvolvimento futuro.

Pesquisar sobre como medir níveis de *loudness* de objetos sonoros em movimento seria a evolução natural deste trabalho. Até porque uma medida fidedigna pode ser mais complexa do que simplesmente sobrepor e somar valores instantâneos calculados como “fotografias” de um objeto em movimento, isto é, a cada fração de tempo na qual o objeto é estático, definido por um bloco de metadados. Experimentos a serem planejados para observação de efeitos posicionais não mais estáticos, mas sim dinâmicos, não se furtariam de testar áudio acompanhado por vídeo, por exemplo. Pois, intuitivamente, tem-se a impressão de que as dicas visuais de localização se sobreponham às dicas auditivas em algum grau. Serão necessários experimentos perceptivos audiovisuais para observação de efeitos de imagem na contaminação das dicas auditivas, algo na linha do que foi feito neste trabalho quanto à contaminação da percepção das distâncias auditivas pelos tempos de reverberação das salas de reprodução. Entende-se que esta é uma direção na qual pode ser possível formular perguntas de pesquisa relevantes e testar hipóteses ainda mais interessantes.

REFERÊNCIAS

ABE, S. **Support Vector Machines for Pattern Classification**. Springer London, 2006. (Advances in Computer Vision and Pattern Recognition). ISBN 9781846282195. Disponível em: <https://books.google.com.br/books?id=3DswtC_ZYrwC>. Citado 3 vezes nas páginas 139, 144 e 145.

AGÊNCIA NACIONAL DE TELECOMUNICAÇÕES. **Resolução nº 284 de 2001**: Regulamento técnico para a prestação dos serviços de radiodifusão de sons e imagens e de retransmissão de televisão. Brasília, 2001. 77 p. Citado na página 135.

_____. **Resolução nº 67 de 1998, alterada pela Resolução nº 546 de 2010**: Regulamento técnico para emissoras de radiodifusão sonora em frequência modulada. Brasília, 2010. 100 p. Citado na página 135.

_____. **Portaria nº 559, de 22 de julho de 2014**: Aprova o procedimento de fiscalização para medição da intensidade subjetiva de áudio (Loudness) no serviço de radiodifusão de sons e imagens (TV) com tecnologia digital. Brasília, 2014. 10 p. Citado 3 vezes nas páginas 36, 127 e 173.

ALGAZI, V. R.; DUDA, R. O.; THOMPSON, D. M.; AVENDANO, C. The CIPIC HRTF database. In: **Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)**. [S.l.: s.n.], 2001. p. 99–102. Citado 3 vezes nas páginas 74, 75 e 159.

ALPAYDIN, E. **Introduction to Machine Learning**. MIT Press, 2014. (Adaptive computation and machine learning). ISBN 9780262028189. Disponível em: <<https://books.google.com.br/books?id=NP5bBAAAQBAJ>>. Citado na página 146.

AUDIO ENGINEERING SOCIETY. **AES17-1998: AES standard method for digital audio engineering - Measurement of digital audio equipment**. Geneva, 1998. AES17-1998. Citado na página 186.

_____. **Technical Document AES TD1004.1.15-10**: Recommendation for loudness of audio streaming and network file playback. New York, 2015. 10 p. Citado 2 vezes nas páginas 120 e 127.

BARKHAUSEN, H. Ein neuer schallmesser für die praxis. **A. f. Techn. Physik**, v. 599, 1926. Citado 2 vezes nas páginas 50 e 56.

BAUER, B.; TORICK, E. Researches in loudness measurement. **IEEE Transactions on Audio and Electroacoustics**, v. 14, n. 3, p. 141–151, Sep 1966. ISSN 0018-9278. Citado na página 87.

BAUER, B.; TORICK, E.; ROSENHECK, A.; ALLEN, R. A loudness-level monitor for broadcasting. **IEEE Transactions on Audio and Electroacoustics**, v. 15, n. 4, p. 177–182, December 1967. ISSN 0018-9278. Citado na página 87.

BECH, S.; ZACHAROV, N. **Perceptual audio evaluation - Theory, method and application**. [S.l.]: John Wiley & Sons, 2007. Citado 6 vezes nas páginas 199, 202, 203, 204, 205 e 208.

BÉKÉSY, G. V. **Zur Theorie des Hörens: Über die eben merkbare Amplituden-und Frequenzänderung eines Tones**. [S.l.]: éditeur inconnu, 1929. Citado na página 63.

BÉKÉSY, G. V.; WEVER, E. G. **Experiments in hearing**. [S.l.]: McGraw-Hill New York, 1960. v. 8. Citado na página 247.

BERANEK, L. **Acoustics**. [s.n.], 1954. Disponível em: <<https://books.google.com.br/books?id=73nsjwEACAAJ>>. Citado 2 vezes nas páginas 48 e 68.

BERANEK, L.; MELLOW, T. **Acoustics: Sound Fields and Transducers**. Academic Press, 2012. (Academic Press). ISBN 9780123914217. Disponível em: <<https://books.google.com.br/books?id=VYvS7MyaEE8C>>. Citado 3 vezes nas páginas 49, 50 e 67.

BHARITKAR, S.; KYRIAKAKIS, C. **Immersive Audio Signal Processing**. Springer New York, 2008. (Information Technology: Transmission, Processing and Storage). ISBN 9780387285030. Disponível em: <<https://books.google.com.br/books?id=P6ndOChbFZEC>>. Citado na página 52.

BLAUERT, J. **Spatial Hearing: The Psychophysics of Human Sound Localization**. [S.l.]: MIT press, 1997. Citado 2 vezes nas páginas 201 e 257.

BLEIDT, R. L.; SEN, D.; NIEDERMEIER, A.; CZELHAN, B.; FÜG, S.; DISCH, S.; HERRE, J.; HILPERT, J.; NEUENDORF, M.; FUCHS, H.; ISSING, J.; MURTAZA, A.; KUNTZ, A.; KRATSCHMER, M.; KÜCH, F.; FÜG, R.; SCHUBERT, B.; DICK, S.; FUCHS, G.; SCHUH, F.; BURDIEL, E.; PETERS, N.; KIM, M. Development of the MPEG-H TV audio system for ATSC 3.0. **IEEE Transactions on Broadcasting**, v. 63, n. 1, p. 202–236, March 2017. ISSN 0018-9316. Citado na página 180.

BLESSER, B. Audio dynamic range compression for minimum perceived distortion. **IEEE Transactions on Audio and Electroacoustics**, v. 17, n. 1, p. 22–32, Mar 1969. ISSN 0018-9278. Citado na página 31.

BOLEY, J.; DANNER, C.; LESTER, M. Measuring dynamics: Comparing and contrasting algorithms for the computation of dynamic range. In: **Audio Engineering Society Convention 129**. [s.n.], 2010. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=15601>>. Citado na página 124.

BOULLET, I. **The loudness of impulsive sounds : perception, measures and models**. Tese (Theses) — Université de la Méditerranée - Aix-Marseille II, jun. 2005. Encadrant industriel : Patrick Boussard (GENESIS). Disponível em: <<https://tel.archives-ouvertes.fr/tel-00009870>>. Citado na página 101.

BRAGA, A. d. P. **Redes neurais artificiais: teoria e aplicações**. LTC Editora, 2007. ISBN 9788521615644. Disponível em: <<https://books.google.com.br/books?id=R-p1GwAACAAJ>>. Citado na página 145.

BRIXEN, E. **Audio Metering: Measurements, Standards and Practice**. Focal Press, 2011. (Focal Press). ISBN 9780240814674. Disponível em: <<https://books.google.com.br/books?id=CAmNxDZV6jEC>>. Citado 4 vezes nas páginas 34, 101, 104 e 107.

BROWN, C. P.; DUDA, R. O. A structural model for binaural sound synthesis. **IEEE Transactions on Speech and Audio Processing**, v. 6, n. 5, p. 476–488, Sep 1998. ISSN 1063-6676. Citado 5 vezes nas páginas 71, 72, 73, 74 e 214.

BUUS, S.; FLORENTINE, M.; POULSEN, T. Temporal integration of loudness, loudness discrimination, and the form of the loudness function. **The Journal of the Acoustical Society of America**, v. 101, n. 2, p. 669–680, 1997. Disponível em: <<http://dx.doi.org/10.1121/1.417959>>. Citado 2 vezes nas páginas 60 e 61.

CABRERA, D.; MIRANDA, L.; DASH, I. Directional loudness measurements for a multichannel system. **The Journal of the Acoustical Society of America**, v. 123, n. 5, p. 3725–3725, 2008. Disponível em: <<http://dx.doi.org/10.1121/1.2935203>>. Citado na página 123.

CAMERER, F. *et al.* Loudness normalization: The future of file-based playback. **Paper for the audio industry**, p. 12, 2012. Citado na página 127.

CAMPBELL, M.; GREATED, C. **The Musician's Guide to Acoustics**. OUP Oxford, 1994. ISBN 9780191591679. Disponível em: <<https://books.google.com.br/books?id=iiCZwwFG0x0C>>. Citado 3 vezes nas páginas 42, 43 e 55.

CANADIAN RADIO-TELEVISION AND TELECOMMUNICATIONS COMMISSION. **Broadcasting Regulatory Policy CRTC 2011-584**: Measures to control the loudness of commercial messages. Ottawa, 2011. 2 p. Citado na página 126.

CBS, B. I. **ITU-R WP6C Document 6G/52-E: 2008**: Measurement and control of television program loudness. Results of a recent CBS test. Genebra, 2008. 2 p. Citado na página 123.

CHALUPPER, J.; FASTL, H. Dynamic Loudness Model (DLM) for normal and hearing-impaired listeners. **Acta Acustica united with Acustica**, v. 88, n. 3, 2002. Citado 6 vezes nas páginas 38, 97, 98, 100, 114 e 148.

CHANAUD, R. Effects of geometry on the resonance frequency of Helmholtz resonators. **Journal of Sound and Vibration**, v. 178, n. 3, p. 337 – 348, 1994. ISSN 0022-460X. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0022460X84714908>>. Citado na página 42.

CHENG, C. I.; WAKEFIELD, G. H. Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in time, frequency, and space. In: **Audio Engineering Society Convention 107**. [s.n.], 1999. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=8154>>. Citado 2 vezes nas páginas 71 e 76.

COMISSÃO INTERNACIONAL DE ELETROTÉCNICA. **IEC 60651**: Sound level meters. Genebra, 1979. 86 p. Citado 2 vezes nas páginas 35 e 102.

CONGRESSO NACIONAL. **Parecer nº 8 de 2013**: da comissão mista, sobre a medida provisória no. 589, de 13 de novembro de 2012, que dispõe sobre o parcelamento de débitos junto à fazenda nacional relativos às contribuições previdenciárias. Brasília, 2013. 30 p. Citado na página 135.

COUSEIL SUPÉRIEUR DE LAUDIOVISUEL. **Délibération n 2011-29 du 19 juillet 2011**: relative aux caractéristiques techniques de l'intensité sonore des programmes et des messages publicitaires de télévision. Paris, 2011. 2 p. Citado na página 127.

DASH, I. True peak metering—a tutorial review. In: **Audio Engineering Society Convention 136**. [s.n.], 2014. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=17188>>. Citado 3 vezes nas páginas 35, 112 e 126.

DASH, I.; BASSETT, M.; CABRERA, D. Relative importance of speech and non-speech components in program loudness assessment. In: **Audio Engineering**

Society Convention 128. [s.n.], 2010. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=15340>>. Citado 2 vezes nas páginas 124 e 126.

DASH, I. M. Octave-band analysis on ITU-R listening test data. In: **Audio Engineering Society Convention 126**. [s.n.], 2009. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=15007>>. Citado 2 vezes nas páginas 123 e 130.

DIETTERICH, T. G. Approximate statistical tests for comparing supervised classification learning algorithms. **Neural computation**, MIT Press, v. 10, n. 7, p. 1895–1923, 1998. Citado na página 152.

DIGITAL PRODUCTION PARTNERSHIP. **Technical specification for the delivery of television programmes as AS-11 files to the BBC (BBC file v5.0.5)**. Londres, 2018. 30 p. Disponível em: <<http://dpp-assets.s3.amazonaws.com/wp-content/uploads/specs/bbc/TechnicalDeliveryStandardsBBCFile.pdf>>. Citado 4 vezes nas páginas 173, 174, 178 e 179.

_____. **Technical Standard for Delivery of HD Commercial and Sponsorship copy**. Londres, 2018. 20 p. Disponível em: <<http://dpp-assets.s3.amazonaws.com/wp-content/uploads/specs/bbc/TechnicalDeliveryStandardsBBCFile.pdf>>. Citado 3 vezes nas páginas 173, 175 e 179.

DOLBY LABORATORIES INCORPORATED. **Model 737: Soundtrack loudness meter - Leq(m)**. São Francisco, EUA, 1988. 25 p. Disponível em: <<http://www.film-tech.com/warehouse/manuals/DOLBYMODEL737.pdf>>. Citado 2 vezes nas páginas 35 e 105.

_____. **Dolby LM100: Broadcast loudness meter user's manual**. São Francisco, EUA, 2011. 62 p. Disponível em: <<https://www.dolby.com/us/en/professional/broadcast/products/dolby-broadcast-loudness-meter-lm100-manual.pdf>>. Citado na página 102.

DUDA, R.; HART, P.; STORK, D. **Pattern Classification**. Wiley, 2012. ISBN 9781118586006. Disponível em: <<https://books.google.com.br/books?id=Br33IRC3PkQC>>. Citado 4 vezes nas páginas 71, 139, 143 e 161.

EBU. **Tech 3253 Sound Quality Assessment Material recordings for subjective tests**. 3. ed. Geneva, 2008. EBU Tech 3253. Citado na página 200.

EKMAN, G.; BERGLUND, B. Loudness as a function of the duration of auditory stimulation. **Scandinavian Journal of Psychology**, Blackwell Publishing Ltd, v. 7, n. 1, p. 201–208, 1966. ISSN 1467-9450. Disponível em: <<http://dx.doi.org/10.1111/j.1467-9450.1966.tb01354.x>>. Citado na página 63.

EPSTEIN, M.; FLORENTINE, M. Binaural loudness summation for speech and tones presented via earphones and loudspeakers. **Ear and hearing**, LWW, v. 30, n. 2, p. 234–237, 2009. Citado 2 vezes nas páginas 65 e 66.

FARINA, A. Advancements in impulse response measurements by sine sweeps. In: **Audio Engineering Society Convention 122**. [s.n.], 2007. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=14106>>. Citado na página 218.

FASTL, H. Alternatives to a-weighting: Psychoacoustic background. **The Journal of the Acoustical Society of America**, v. 128, n. 4, p. 2469–2469, 2010. Disponível em: <<http://asa.scitation.org/doi/abs/10.1121/1.3508846>>. Citado na página 50.

FASTL, H.; ZWICKER, E. **Psychoacoustics: Facts and Models**. Springer Berlin Heidelberg, 2007. (Springer series in information sciences). ISBN 9783540688884. Disponível em: <<https://books.google.com.br/books?id=eGcfn9ddRhcC>>. Citado na página 84.

FECHNER, G. In sachen a' er psychophysik. **Leipzig: Breitkopf & Härtel**, 1877. Citado 2 vezes nas páginas 45 e 46.

_____. **Elemente der psychophysik**. Breitkopf & Härtel, 1907. ISBN 9785872620525. Disponível em: <<https://books.google.com.br/books?id=NpMLAwAAQBAJ>>. Citado 2 vezes nas páginas 46 e 47.

FEDERAL COMMUNICATIONS COMMISSION. **Second Report and Order**: In the matter of implementation of the Commercial Advertisement Loudness Mitigation (CALM) Act. Washington, 2014. 15 p. Citado na página 126.

FEILAT, E. A. Detection of voltage envelope using Prony analysis - Hilbert transform method. **IEEE Transactions on Power Delivery**, v. 21, n. 4, p. 2091–2093, Oct 2006. ISSN 0885-8977. Citado na página 149.

FIELD, A. **Discovering Statistics Using IBM SPSS Statistics**. [S.l.]: Sage, 2013. Citado 4 vezes nas páginas 206, 238, 239 e 272.

FLANAGAN, J. **Speech Analysis Synthesis and Perception**. Springer Berlin Heidelberg, 2013. (Communication and Cybernetics). ISBN 9783662015629. Disponível em: <<https://books.google.com.br/books?id=b1X-CAAAQBAJ>>. Citado na página 42.

FLETCHER, H. Auditory patterns. **Rev. Mod. Phys.**, American Physical Society, v. 12, p. 47–65, Jan 1940. Disponível em: <<http://link.aps.org/doi/10.1103/RevModPhys.12.47>>. Citado 5 vezes nas páginas 54, 56, 59, 79 e 82.

FLETCHER, H.; MUNSON, W. A. Loudness, its definition, measurement and calculation. **Bell System Technical Journal**, Blackwell Publishing Ltd, v. 12, n. 4, p. 377–430, 1933. ISSN 1538-7305. Disponível em: <<http://dx.doi.org/10.1002/j.1538-7305.1933.tb00403.x>>. Citado 10 vezes nas páginas 33, 34, 50, 59, 60, 65, 68, 70, 85 e 101.

FLORENTINE, M.; POPPER, A.; FAY, R. **Loudness**. Springer New York, 2010. (Springer Handbook of Auditory Research). ISBN 9781441967121. Disponível em: <<https://books.google.com.br/books?id=78L77CopexQC>>. Citado na página 41.

FOLLADOR, A.; SILVA, A. P.; YEHIA, H. C. **Base de Dados Corpus CEFALA-1**. 2017. Disponível em: <<https://www.cefala.org/>>. Citado na página 200.

FRANCOMBE, J. **IoSR Listening Room Multichannel BRIR dataset**. 2015. Disponível em: <<https://doi.org/10.15126/surreydata.00813511>>. Citado na página 193.

FRANCOMBE, J.; BROOKES, T.; MASON, R.; MELCHIOR, F. Loudness matching multichannel audio program material with listeners and predictive models. In: **Audio Engineering Society Convention 139**. [s.n.], 2015. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=18020>>. Citado 8 vezes nas páginas 129, 130, 161, 162, 183, 205, 292 e 296.

FRANCOMBE, J.; BROOKES, T.; MASON, R.; FLINDT, R.; COLEMAN, P.; LIU, Q.; JACKSON, P. Production and reproduction of program material for a variety of spatial audio formats. In: AUDIO ENGINEERING SOCIETY. **Audio Engineering Society Convention 138**. [S.l.], 2015. Citado na página 294.

FRANK, M.; ZOTTER, F.; SONTACCHI, A. Producing 3D audio in ambisonics. In: **Audio Engineering Society Conference: 57th International Conference: The Future of Audio Entertainment Technology—Cinema, Television and the Internet**. [s.n.], 2015. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=17605>>. Citado 2 vezes nas páginas 129 e 132.

FRAUNHOFER, I. **ITU-R WP6C Document 6C/48-E: 2016**: Report on listening tests for development of loudness measurement algorithm for the advanced sound system. Genebra, 2016. 2 p. Citado 2 vezes nas páginas 176 e 177.

GAMPER, H. Head-Related Transfer Function interpolation in azimuth, elevation, and distance. **The Journal of the Acoustical Society of America, ASA**, v. 134, n. 6, p. EL547–EL553, 2013. Citado na página 257.

GARNER, W. R. The loudness and loudness matching of short tones. **The Journal of the Acoustical Society of America**, v. 21, n. 4, p. 398–403, 1949. Disponível em: <<http://dx.doi.org/10.1121/1.1906526>>. Citado na página 63.

GENELEC. **Monitors in room performance**. [S.l.], 2018. Rev. 1. Citado 2 vezes nas páginas 187 e 190.

GERZON, M. A. Ambisonics in multichannel broadcasting and video. **J. Audio Eng. Soc.**, v. 33, n. 11, p. 859–871, 1985. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=4419>>. Citado na página 334.

GIANNAKOPOULOS, T.; PIKRAKIS, A. **Introduction to Audio Analysis: A MATLAB® Approach**. Elsevier Science, 2014. ISBN 9780080993898. Disponível em: <<https://books.google.com.br/books?id=zbHVAQAAQBAJ>>. Citado na página 141.

GIANNOULIS, D.; MASSBERG, M.; REISS, J. D. Digital dynamic range compressor design—a tutorial and analysis. **J. Audio Eng. Soc.**, v. 60, n. 6, p. 399–408, 2012. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=16354>>. Citado na página 148.

GLASBERG, B. R.; MOORE, B. C. Derivation of auditory filter shapes from notched-noise data. **Hearing Research**, v. 47, n. 1–2, p. 103 – 138, 1990. ISSN 0378-5955. Disponível em: <<http://www.sciencedirect.com/science/article/pii/037859559090170T>>. Citado 6 vezes nas páginas 8, 57, 58, 59, 83 e 88.

GLASBERG, B. R.; MOORE, B. C. J. A model of loudness applicable to time-varying sounds. **J. Audio Eng. Soc.**, v. 50, n. 5, p. 331–342, 2002. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=11081>>. Citado 15 vezes nas páginas 10, 34, 38, 96, 97, 99, 100, 114, 122, 129, 166, 167, 168, 280 e 295.

_____. The loudness of sounds whose spectra differ at the two ears. **The Journal of the Acoustical Society of America**, v. 127, n. 4, p. 2433–2440, 2010. Disponível em: <<http://dx.doi.org/10.1121/1.3336775>>. Citado na página 99.

GREEN, D.; SWETS, J. **Signal detection theory and psychophysics**. Wiley, 1966. Disponível em: <<https://books.google.com.br/books?id=Ykt9AAAAMAAJ>>. Citado 4 vezes nas páginas 34, 79, 81 e 83.

GREEN, D. M. Auditory detection of a noise signal. **The Journal of the Acoustical Society of America**, v. 32, n. 1, p. 121–131, 1960. Disponível em: <<http://dx.doi.org/10.1121/1.1907862>>. Citado 2 vezes nas páginas 79 e 81.

_____. Psychoacoustics and detection theory. **The Journal of the Acoustical Society of America**, v. 32, n. 10, p. 1189–1203, 1960. Disponível em: <<http://dx.doi.org/10.1121/1.1907882>>. Citado 2 vezes nas páginas 78 e 79.

GRIMM, E.; SKOVENBORG, E.; SPIKOFSKI, G. Determining an optimal gated loudness measurement for TV sound normalization. In: **Audio Engineering Society Convention 128**. [s.n.], 2010. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=15450>>. Citado 2 vezes nas páginas 110 e 124.

GRIMM, E. M.; EVERDINGEN, R. V.; SCHÖPPING, M. J. L. C. Toward a recommendation for a european standard of peak and LKFS loudness levels. **SMPTE Motion Imaging Journal**, v. 119, n. 3, p. 28–34, April 2010. ISSN 1545-0279. Citado na página 124.

GUYON, I.; GUNN, S.; NIKRAVESH, M.; ZADEH, L. **Feature Extraction: Foundations and Applications**. Springer Berlin Heidelberg, 2008. (Studies in Fuzziness and Soft Computing). ISBN 9783540354888. Disponível em: <<https://books.google.com.br/books?id=FOTzBwAAQBAJ>>. Citado na página 142.

HAVELOCK, D.; KUWANO, S.; VORLÄNDER, M. **Handbook of Signal Processing in Acoustics**. Springer New York, 2008. (Handbook of Signal Processing in Acoustics). ISBN 9780387304410. Disponível em: <<https://books.google.com.br/books?id=YaNCAAAAQBAJ>>. Citado 2 vezes nas páginas 44 e 45.

HELMHOLTZ, H. Theorie der luftschwingungen in röhren mit offenen enden. **Journal für die reine und angewandte Mathematik**, v. 57, p. 1–72, 1860. Disponível em: <<http://eudml.org/doc/147771>>. Citado na página 42.

HERRE, J.; HILPERT, J.; KUNTZ, A.; PLOGSTIES, J. MPEG-H Audio—The new standard for universal spatial/3D audio coding. **J. Audio Eng. Soc**, v. 62, n. 12, p. 821–830, 2015. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=17556>>. Citado 2 vezes nas páginas 128 e 131.

HEWITT, M. J.; MEDDIS, R. A computer model of amplitude-modulation sensitivity of single units in the inferior colliculus. **The Journal of the Acoustical Society of America**, v. 95, n. 4, p. 2145–2159, 1994. Disponível em: <<http://dx.doi.org/10.1121/1.408676>>. Citado na página 44.

HUMMERSON, C. **Binaural Room Impulse Response Measurements**. 2011. 6 p. Citado 4 vezes nas páginas 225, 226, 227 e 228.

INSTITUTO NACIONAL AMERICANO DE PADRÕES. **ANSI/ASA S3.4-2007 (R2012)**: American national standard procedure for the computation of loudness of steady sounds. Washington, D.C., 2012. Citado na página 34.

IOSR, I. of S. R. **S3A: Future Spatial Audio for an Immersive Listener Experience at Home**. 2013. Citado na página 172.

_____. **ITU-R BS.1116 critical listening room and audio laboratory**. 2017. Disponível em: <<http://iosr.uk/facilities/listeningroom.php>>. Citado na página 171.

ISO, E. 3382-1, 2009, “Acoustics—Measurement of Room Acoustic Parameters—Part 1: Performance Spaces,”. **International Organization for Standardization, Brussels, Belgium**, 2009. Citado na página 219.

JACKSON, P.; DEWHIRST, M.; CONETTA, R.; ZIELINSKI, S.; RUMSEY, F.; MEARES, D.; BECH, S.; GEORGE, S. Qestral (part 3): System and metrics for spatial quality prediction. In: AUDIO ENGINEERING SOCIETY. **Audio Engineering Society Convention 125**. [S.l.], 2008. Citado na página 257.

KEAN, J.; JOHNSON, E.; SHEFFIELD, E. Study of audio loudness range for consumers in various listening modes and ambient noise levels. **Consumer Electronics Association**, 2015. Citado na página 128.

KJÖRLING, K.; RÖDÉN, J.; WOLTERS, M.; RIEDMILLER, J.; BISWAS, A.; EKSTRAND, P.; GRÖSCHEL, A.; HEDELIN, P.; HIRVONEN, T.; HÖRICH, H.; KLEJSA, J.; KOPPENS, J.; KRAUSS, K.; LEHTONEN, H.-M.; LINZMEIER, K.; MUESCH, H.; MUNDT, H.; NORCROSS, S.; POPP, J.; PURNHAGEN, H.; SAMUELSSON, J.; SCHUG, M.; SEHLSTRÖM, L.; THESING, R.; VILLEMOES, L.; VINTON, M. AC-4 – The next generation audio codec. In: **Audio Engineering Society Convention 140**. [s.n.], 2016. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=18190>>. Citado na página 33.

KLEINER, M.; DALENBÄCK, B.-I.; SVENSSON, P. Auralization-an overview. **J. Audio Eng. Soc**, v. 41, n. 11, p. 861–875, 1993. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=6976>>. Citado na página 335.

KOLARIK, A. J.; MOORE, B. C.; ZAHORIK, P.; CIRSTEAN, S.; PARDHAN, S. Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. **Attention, Perception, & Psychophysics**, Springer, v. 78, n. 2, p. 373–395, 2016. Citado na página 201.

KOLLMEIER, B.; KOCH, R. Speech enhancement based on physiological and psychoacoustical models of modulation perception and binaural interaction. **The Journal of the Acoustical Society of America**, v. 95, n. 3, p. 1593–1602, 1994. Disponível em: <<http://dx.doi.org/10.1121/1.408546>>. Citado na página 45.

KOMORI, T.; OODE, S.; ONO, K.; IRIE, K.; SASAKI, Y.; HASEGAWA, T.; SAWAYA, I. Subjective loudness of 22.2 multichannel programs. In: **Audio Engineering Society Convention 138**. [s.n.], 2015. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=17643>>. Citado 9 vezes nas páginas 18, 33, 35, 113, 129, 130, 247, 248 e 249.

KROPUENSKE, G. **Setting AC-3 Dialnorm – Measuring AES dialog level Laeq**. 2007. Disponível em: <<http://www.theonlineengineer.org/DownloadDocs/MeasDialnormAeq.pdf>>. Citado na página 120.

KUECH, F.; KRATSCHMER, M.; NEUGEBAUER, B.; MEIER, M.; BAUMGARTE, F. Dynamic range and loudness control in MPEG-H 3D Audio. In: **Audio Engineering Society Convention 139**. [s.n.], 2015. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=18021>>. Citado 4 vezes nas páginas 33, 128, 131 e 181.

LARA, L. T.; PASQUAL, A. M. Síntese binauricular para produção de sinais sonoros espacializados. In: UNIVERSIDADE ESTADUAL DE CAMPINAS. **XXV Encontro da Sociedade Brasileira de Acústica, 2014, Campinas. Anais do XXV Encontro da SOBRAC**. Campinas. [S.l.], 2014. p. 442–449. Citado 2 vezes nas páginas 71 e 159.

LAVOIE, M. C.; SOULODRE, G. A. Development and evaluation of short-term loudness meters. In: **Audio Engineering Society Convention 121**. [s.n.], 2006. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=13723>>. Citado 2 vezes nas páginas 114 e 122.

LIANG, Y. C.; ZENG, Y.; PEH, E. C. Y.; HOANG, A. T. Sensing-throughput tradeoff for cognitive radio networks. **IEEE Transactions on Wireless Communications**, v. 7, n. 4, p. 1326–1337, April 2008. ISSN 1536-1276. Citado na página 82.

LONG, J. S.; ERVIN, L. H. Using heteroscedasticity consistent standard errors in the linear regression model. **The American Statistician**, Taylor Francis, v. 54, n. 3, p. 217–224, 2000. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/00031305.2000.10474549>>. Citado na página 242.

LUND, T. Stop counting samples. In: **Audio Engineering Society Convention 121**. [s.n.], 2006. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=13806>>. Citado 2 vezes nas páginas 111 e 125.

_____. Audio delivery specification. In: **Audio Engineering Society Convention 123**. [s.n.], 2007. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=14240>>. Citado na página 122.

_____. ITU-R BS. 1770 Revisited. In: **Proc. of the NAB-2011 Convention**. [S.l.: s.n.], 2011. Citado na página 125.

_____. Audio for mobile TV, iPad and iPod. In: **Proc. NAB BE Conference**. [S.l.: s.n.], 2013. Citado na página 127.

LYMAN, S.; SEEFELDT, A. A comparison of various multichannel loudness measurement techniques. In: **Audio Engineering Society Convention 121**. [s.n.], 2006. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=13752>>. Citado 2 vezes nas páginas 106 e 122.

MARKS, L. E. Binaural summation of the loudness of pure tones. **The Journal of the Acoustical Society of America**, v. 64, n. 1, p. 107–113, 1978. Disponível em: <<http://dx.doi.org/10.1121/1.381976>>. Citado 2 vezes nas páginas 65 e 66.

MASON, A. Use of the low frequency effects (LFE) channel in broadcasting. **BBC Research & Development White Paper WHP**, v. 203, 2011. Citado 3 vezes nas páginas 125, 130 e 176.

MERLEAU-PONTY, M. **Fenomenologia da Percepção**. WMF Martins Fontes, 1999. (Biblioteca do Pensamento Moderno). ISBN 85-336-1033-5. Disponível em: <<https://books.google.com.br/books?id=DHBptwAACAAJ>>. Citado na página 5.

METER, D. van; MIDDLETON, D. Modern statistical approaches to reception in communication theory. **Transactions of the IRE Professional Group on Information Theory**, v. 4, n. 4, p. 119–145, September 1954. ISSN 2168-2690. Citado na página 78.

MILLER, G. A.; TAYLOR, W. G. The perception of repeated bursts of noise. **The Journal of the Acoustical Society of America**, v. 20, n. 2, p. 171–182, 1948. Disponível em: <<http://dx.doi.org/10.1121/1.1906360>>. Citado na página 63.

MINISTÉRIO DAS COMUNICAÇÕES. **Portaria nº 354, de 11 de julho de 2012**: Regulamenta a padronização do volume de áudio nos intervalos comerciais da programação dos serviços de radiodifusão sonora e de sons e imagens nos termos da Lei nº 10.222, de 9 de maio de 2001. Brasília, 2012. 3 p. Citado 13 vezes nas páginas 31, 33, 36, 41, 107, 127, 136, 172, 173, 178, 200, 303 e 334.

MOHRMANN, K. Lautheitskonstanz im entfernungswechsel. (constancy of loudness with changes in distance). **Zeitschrift für angewandte Psychologie und Charakterkunde**, JA Barth Verlag, 1939. Citado na página 183.

MONTGOMERY, D. **Estatística Aplicada e Probabilidade para Engenheiros**. Livros Técnicos e Científicos, 2003. ISBN 9788521613602. Disponível em: <<https://books.google.com.br/books?id=hkt0AAAACAAJ>>. Citado na página 138.

MOORE, B. **An Introduction to the Psychology of Hearing**. Emerald, 2012. ISBN 9781780520384. Disponível em: <<https://books.google.com.br/books?id=LM9U8e28pLMC>>. Citado 9 vezes nas páginas 33, 47, 52, 53, 56, 62, 92, 292 e 303.

MOORE, B. C.; GLASBERG, B. R.; VARATHANATHAN, A.; SCHLITTEN-LACHER, J. A Loudness Model for Time-Varying Sounds Incorporating Binaural Inhibition. **Trends in Hearing**, v. 20, p. 1–16, 2016. ISSN 23312165. Citado na página 280.

MOORE, B. C. J.; GLASBERG, B. R. A revision of Zwicker's loudness model. **Acta Acustica united with Acustica**, v. 82, n. 2, 1996. Citado 4 vezes nas páginas 34, 88, 89 e 95.

_____. Modeling binaural loudness. **The Journal of the Acoustical Society of America**, v. 121, n. 3, p. 1604–1612, 2007. Disponível em: <<http://dx.doi.org/10.1121/1.2431331>>. Citado na página 99.

MOORE, B. C. J.; GLASBERG, B. R.; BAER, T. A model for the prediction of thresholds, loudness, and partial loudness. **J. Audio Eng. Soc**, v. 45, n. 4, p. 224–240, 1997. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=10272>>. Citado 11 vezes nas páginas 17, 34, 38, 39, 88, 89, 91, 94, 95, 96 e 159.

MUNSON, W. A. The growth of auditory sensation. **The Journal of the Acoustical Society of America**, v. 19, n. 4, p. 584–591, 1947. Disponível em: <<http://dx.doi.org/10.1121/1.1916525>>. Citado 3 vezes nas páginas 59, 60 e 63.

NIELSEN, S. H.; LUND, T. Overload in signal conversion. In: **Audio Engineering Society Conference: 23rd International Conference: Signal Processing in Audio Recording and Reproduction**. [s.n.], 2003. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=12326>>. Citado na página 125.

NIESE, H. Vorschlag für die definition und messung der deutlichkeit nach subjektiven grundlagen. **Hochfrequenztechnik und Elektroakustik**, v. 65, n. 1, p. 4, 1956. Citado na página 63.

_____. Die tragheit der lautstarkebildung in abhangigkeit vom schallpegel. **Hochfrequenztechnik und Elektroakustik**, v. 68, p. 143–152, 1959. Citado na página 63.

NOBRE, J. S.; SINGER, J. da M. Residual analysis for linear mixed models. **Biometrical Journal: Journal of Mathematical Methods in Biosciences**, Wiley Online Library, v. 49, n. 6, p. 863–875, 2007. Citado na página 211.

NORCROSS, S.; NANDA, S.; COHEN, Z. ITU-R BS. 1770 based loudness for immersive audio. In: **Audio Engineering Society Convention 140**. [s.n.], 2016. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=18199>>. Citado 2 vezes nas páginas 129 e 132.

NORCROSS, S.; POULIN, F.; LAVOIE, M. Evaluation of live meter ballistics for loudness control. In: **Audio Engineering Society Convention 130**. [s.n.], 2011. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=15809>>. Citado na página 125.

NORCROSS, S. G.; LAVOIE, M. C. Investigations on the inclusion of the LFE channel in the ITU-R BS.1770-1 loudness algorithm. In: **Audio Engineering Society Convention 127**. [s.n.], 2009. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=15025>>. Citado 3 vezes nas páginas 109, 123 e 130.

_____. The effect of downmixing on measured loudness. In: **Audio Engineering Society Convention 131**. [s.n.], 2011. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=16594>>. Citado 2 vezes nas páginas 125 e 130.

_____. Loudness normalization of wide-dynamic range broadcast material. In: **Audio Engineering Society Convention 132**. [s.n.], 2012. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=16244>>. Citado na página 126.

NORCROSS, S. G.; SOULODRE, G. A.; LAVOIE, M. C. The subjective loudness of typical program material. In: **Audio Engineering Society Convention 115**. [s.n.], 2003. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=12350>>. Citado na página 120.

OPPENHEIM, A.; SCHAFER, R. **Discrete-time Signal Processing**. Pearson, 2013. (Always learning). ISBN 9781292025728. Disponível em: <<https://books.google.com.br/books?id=LeQpnwEACAAJ>>. Citado 2 vezes nas páginas 112 e 149.

ORGANIZAÇÃO INTERNACIONAL PARA PADRONIZAÇÃO. **ISO 532:1975**: Acoustics – method for calculating loudness level. Genebra, 1975. 428 p. Citado 3 vezes nas páginas 34, 84 e 88.

_____. **ISO 131:1979**: Acoustics – expression of physical and subjective magnitudes of sound or noise in air. Genebra, 1996. Citado na página 53.

_____. **ISO 1999:2013**: Acoustics – estimation of noise-induced hearing loss. Genebra, 2013. 23 p. Citado na página 202.

_____. **ISO 226:2003**: Acoustics – normal equal-loudness-level contours. Genebra, 2014. 18 p. Citado 3 vezes nas páginas 50, 51 e 68.

_____. **ISO/IEC 23008-3:2015**: Information technology – High efficiency coding and media delivery in heterogeneous environments – part 3: 3D audio. Genebra, 2015. 428 p. Citado 2 vezes nas páginas 33 e 155.

PATTERSON, R. D. Auditory filter shapes derived with noise stimuli. **The Journal of the Acoustical Society of America**, v. 59, n. 3, p. 640–654, 1976. Disponível em: <<http://dx.doi.org/10.1121/1.380914>>. Citado 4 vezes nas páginas 8, 56, 57 e 58.

PAULUS, J. Perceptual loudness compensation in interactive object-based audio coding systems. In: **2015 23rd European Signal Processing Conference (EUSIPCO)**. [S.l.: s.n.], 2015. p. 579–583. Citado na página 132.

PEDERSEN, O.; LYREGAAR, P. Loudness of impulsive sounds. In: DECKER PERIODICALS INC 4 HUGHSON STREET SOUTH PO BOX 620, LCD 1, HAMILTON ON L8N 3K7, CANADA. **AUDIOLOGY**. [S.l.], 1972. v. 11, p. 86–87. Citado na página 63.

PETERS, N.; SEN, D.; KIM, M. Y.; WUEBBOLT, O.; WEISS, S. M. Scene-based audio implemented with higher order ambisonics (HOA). In: **SMPTE 2015 Annual Technical Conference and Exhibition**. [S.l.: s.n.], 2015. p. 1–13. Citado 4 vezes nas páginas 33, 131, 132 e 155.

PETERSON, W.; BIRDSALL, T.; FOX, W. The theory of signal detectability. **Transactions of the IRE Professional Group on Information Theory**, v. 4, n. 4, p. 171–212, September 1954. ISSN 2168-2690. Citado na página 78.

PIKE, C.; MELCHIOR, F. An assessment of virtual surround sound systems for headphone listening of 5.1 multichannel audio. In: **Audio Engineering Society Convention 134**. [s.n.], 2013. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=16720>>. Citado 2 vezes nas páginas 166 e 295.

PIRES, L. d. S. Medição de picos verdadeiros de sinais de áudio. In: **V Simpósio de Processamento de Sinais da UNICAMP**. [s.n.], 2014. Disponível em: <http://www.sps.fee.unicamp.br/anais/vol01/VSPS_a04_LPires.pdf>. Citado 2 vezes nas páginas 111 e 126.

PIRES, L. d. S.; VIEIRA, M. N.; YEHIA, H. C. Controle automático de *loudness* em conteúdo de formato curto para radiodifusão. In: **Anais do 14^o Congresso de Engenharia de Áudio da AES Brasil**. [s.n.], 2016. Disponível em: <http://aesbrasil.org/wp-content/uploads/2016/09/Anais_AESBR2016.pdf>. Citado 4 vezes nas páginas 127, 174, 178 e 179.

PIRES, L. da S.; VIEIRA, M. N.; YEHIA, H. C. Automatic loudness control in short-form content for broadcasting. **The Journal of the Acoustical Society of America**, v. 141, n. 3, p. EL287–EL292, 2017. Disponível em: <<http://dx.doi.org/10.1121/1.4978023>>. Citado 4 vezes nas páginas 153, 154, 178 e 179.

PIRES, L. S.; VIEIRA, M. N.; YEHIA, H. C.; PASQUAL, A. M. Medição de *loudness* em áudio imersivo para radiodifusão através de técnicas de auralização. In: **Anais do XXVII Encontro da Sociedade Brasileira de Acústica**. [S.l.: s.n.], 2017. Citado 4 vezes nas páginas 180, 183, 295 e 300.

PIRES, L. S.; VIEIRA, M. N.; YEHIA, H. C.; PASQUAL, A. M.; BROOKES, T. S.; MASON, R. D. Modelo de distância auditiva percebida para o algoritmo de *loudness* ITU-R BS.1770. In: **Anais do XXVIII Encontro da Sociedade Brasileira de Acústica**. [S.l.: s.n.], 2018. Citado na página 216.

POLLACK, I. Loudness of periodically interrupted white noise. **The Journal of the Acoustical Society of America**, v. 30, n. 3, p. 181–185, 1958. Disponível em: <<http://dx.doi.org/10.1121/1.1909531>>. Citado na página 63.

PORT, E. Über die lautstarke einzelner kurzer schallimpul. **Acta Acustica united with Acustica**, v. 13, n. 3, 1963. Citado na página 63.

POULTON, E. C. Models for biases in judging sensory magnitude. **Psychological bulletin**, v. 86, n. 4, p. 777–803, 1979. Citado na página 53.

PRESIDÊNCIA DA REPÚBLICA. **LEI Nº 10.222, DE 9 DE MAIO DE 2001.**: Padroniza o volume de áudio das transmissões de rádio e televisão nos espaços dedicados à propaganda e dá outras providências. Brasília, 2001. 8 p. Citado 4 vezes nas páginas 35, 127, 134 e 172.

_____. **LEI Nº 12.810, DE 15 DE MAIO DE 2013.**: Dispõe sobre o parcelamento de débitos com a fazenda nacional relativos às contribuições previdenciárias de responsabilidade dos estados, do distrito federal e dos municípios; altera as leis n^{os} 8.212, de 24 de julho de 1991, 9.715, de 25 de novembro de 1998, 11.828, de 20 de novembro de 2008, 10.522, de 19 de julho de 2002, 10.222, de 9 de maio de 2001, 12.249, de 11 de junho de 2010, 11.110, de 25 de abril de 2005, 5.869, de 11 de janeiro de 1973 - código de processo civil, 6.404, de 15 de

dezembro de 1976, 6.385, de 7 de dezembro de 1976, 6.015, de 31 de dezembro de 1973, e 9.514, de 20 de novembro de 1997; e revoga dispositivo da lei nº 12.703, de 7 de agosto de 2012. Brasília, 2013. 8 p. Citado 2 vezes nas páginas 36 e 134.

PULKKI, V. Virtual sound source positioning using vector base amplitude panning. **J. Audio Eng. Soc.**, v. 45, n. 6, p. 456–466, 1997. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=7853>>. Citado 2 vezes nas páginas 180 e 251.

PURIA, S.; PEAKE, W. T.; ROSOWSKI, J. J. Sound-pressure measurements in the cochlear vestibule of human-cadaver ears. **The Journal of the Acoustical Society of America**, v. 101, n. 5, p. 2754–2770, 1997. Disponível em: <<http://dx.doi.org/10.1121/1.418563>>. Citado na página 89.

RABINER, L. R.; SCHAFER, R. W. Introduction to digital speech processing. **Foundations and Trends® in Signal Processing**, v. 1, n. 1–2, p. 1–194, 2007. ISSN 1932-8346. Disponível em: <<http://dx.doi.org/10.1561/20000000001>>. Citado 2 vezes nas páginas 45 e 140.

RAFAELY, B. Spatial-temporal correlation of a diffuse sound field. **The Journal of the Acoustical Society of America**, v. 107, n. 6, p. 3254–3258, 2000. Disponível em: <<http://dx.doi.org/10.1121/1.429397>>. Citado na página 68.

REICHARDT, W.; NIESE, H. Choice of sound duration and silent intervals for test and comparison signals in the subjective measurement of loudness level. **The Journal of the Acoustical Society of America**, v. 47, n. 4B, p. 1083–1090, 1970. Disponível em: <<http://dx.doi.org/10.1121/1.1912009>>. Citado na página 63.

RG32, R. G. **ITU-R WP6C Document 6C/48-E: 201**: Progress report on loudness measurement algorithm for the advanced sound system. Genebra, 2017. 2 p. Citado 3 vezes nas páginas 177, 178 e 179.

RIEDMILLER, J.; KJÖRLING, K.; RÖDÉN, J.; WOLTERS, M.; BISWAS, A.; BOON, P.; CARROLL, T.; EKSTRAND, P.; GRÖSCHEL, A.; HEDELIN, P.; HIRVONEN, T.; HÖRICH, H.; KLEJSA, J.; KOPPENS, J.; KRAUSS, K.; LEHTONEN, H.; LINZMEIER, K.; MEHTA, S.; MUESCH, H.; MUNDT, H.; NORCROSS, S.; POPP, J.; PURNHAGEN, H.; RESCH, B.; SAMUELSSON, J.; SCHUG, M.; SEHLSTRÖM, L.; TSINGOS, N.; VILLEMOES, L.; VINTON, M. Delivering scalable audio experiences using AC-4. **IEEE Transactions on Broadcasting**, v. 63, n. 1, p. 179–201, March 2017. ISSN 0018-9316. Citado na página 180.

ROBINSON, D. **Replay Gain - A proposed standard**. 2001. Disponível em: <http://wiki.hydrogenaud.io/index.php?title=ReplayGain_specification>. Citado na página 119.

ROBINSON, D.; WHITTLE, L. The loudness of directional sound fields. **Acta Acustica united with Acustica**, v. 10, n. 2, 1960. Citado 4 vezes nas páginas 229, 247, 248 e 280.

ROBINSON, D. W.; DADSON, R. S. A re-determination of the equal-loudness relations for pure tones. **British Journal of Applied Physics**, v. 7, n. 5, p. 166, 1956. Disponível em: <<http://stacks.iop.org/0508-3443/7/i=5/a=302>>. Citado 2 vezes nas páginas 68 e 70.

ROBINSON, D. W.; WHITTLE, L. S.; BOWSHER, J. M. The loudness of diffuse sound fields. **Acta Acustica united with Acustica**, v. 11, n. 6, 1961. Citado na página 68.

ROEDERER, J. **The Physics and Psychophysics of Music: An Introduction**. Springer New York, 2008. ISBN 9780387094748. Disponível em: <<https://books.google.com.br/books?id=rYfqoc1dDmYC>>. Citado 2 vezes nas páginas 43 e 44.

RUMSEY, F. Loudness revisited. **J. Audio Eng. Soc**, v. 62, n. 12, p. 906–910, 2015. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=17561>>. Citado na página 127.

SCHARF, B. Dichotic summation of loudness. **The Journal of the Acoustical Society of America**, v. 45, n. 5, p. 1193–1205, 1969. Disponível em: <<http://dx.doi.org/10.1121/1.1911590>>. Citado na página 41.

_____. Loudness. In: CARTERETTE, E. (Ed.). **Handbook of Perception Vol.4: Hearing**. [S.l.]: Elsevier Science, 1978. cap. 6, p. 187–242. ISBN 9780323142755. Citado 4 vezes nas páginas 61, 62, 63 e 114.

SCHITTKOWSKI, K. NLPQL: A Fortran subroutine solving constrained non-linear programming problems. **Annals of Operations Research**, v. 5, n. 2, p. 485–500, Jun 1986. ISSN 1572-9338. Disponível em: <<https://doi.org/10.1007/BF02022087>>. Citado na página 278.

SCHLITTENLACHER, J.; ELLERMEIER, W.; HASHIMOTO, T. Spectral loudness summation: Shortcomings of current standards. **The Journal of the Acoustical Society of America**, v. 137, n. 1, p. EL26–EL31, 2015. Disponível em: <<http://dx.doi.org/10.1121/1.4902425>>. Citado na página 34.

SENATE AND HOUSE OF REPRESENTATIVES OF THE UNITED STATES OF AMERICA IN CONGRESS ASSEMBLED. **Public Law 111–311: Commercial Advertisement Loudness Mitigation Act**: To regulate the volume of audio on commercials. dec. 15, 2010. Washington, 2010. 2 p. Citado na página 126.

SHAO, Z.; MO, F.; MAO, D. The effect of stimulus bandwidth on binaural loudness summation. **The Journal of the Acoustical Society of America**, v. 138, n. 3, p. 1508–1514, 2015. Disponível em: <<https://doi.org/10.1121/1.4928955>>. Citado 2 vezes nas páginas 229 e 250.

SHAW, E. A. G. Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. **The Journal of the Acoustical Society of America**, v. 56, n. 6, p. 1848–1861, 1974. Disponível em: <<http://dx.doi.org/10.1121/1.1903522>>. Citado na página 89.

SIVONEN, V. P.; ELLERMEIER, W. Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation. **The Journal of the Acoustical Society of America**, v. 119, n. 5, p. 2965–2980, 2006. Disponível em: <<https://doi.org/10.1121/1.2184268>>. Citado 6 vezes nas páginas 248, 249, 254, 259, 265 e 280.

_____. Binaural loudness for artificial-head measurements in directional sound fields. **J. Audio Eng. Soc.**, v. 56, n. 6, p. 452–461, 2008. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=14394>>. Citado 8 vezes nas páginas 99, 113, 123, 248, 249, 256, 259 e 265.

SKOVENBORG, E. Loudness Range (LRA) - Design and Evaluation. **Audio Engineering Society Convention**, v. 132, p. 1 – 12, Abril 2012. Citado 3 vezes nas páginas 16, 32 e 115.

_____. Loudness range (lra) - design and evaluation. In: **Audio Engineering Society Convention 132**. [s.n.], 2012. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=16254>>. Citado 2 vezes nas páginas 126 e 128.

SKOVENBORG, E.; LUND, T. Loudness descriptors to characterize programs and music tracks. In: **Audio Engineering Society Convention 125**. [s.n.], 2008. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=14666>>. Citado na página 123.

_____. Loudness descriptors to characterize wide loudness-range material. In: **Audio Engineering Society Convention 127**. [s.n.], 2009. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=15142>>. Citado 2 vezes nas páginas 110 e 124.

_____. Level-normalization of feature films using loudness vs speech. In: **Audio Engineering Society Convention 135**. [s.n.], 2013. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=17031>>. Citado na página 120.

SKOVENBORG, E.; NIELSEN, S. Real-time Visualisations of Loudness Along Different Time Scales. **10th International Conference on Digital Audio Effects (DAFx)**, p. 1–6, Setembro 2007. Citado 3 vezes nas páginas 32, 122 e 138.

SMALL, A. M.; BRANDT, J. F.; COX, P. G. Loudness as a function of signal duration. **The Journal of the Acoustical Society of America**, v. 34, n. 4, p. 513–514, 1962. Disponível em: <<http://dx.doi.org/10.1121/1.1918157>>. Citado na página 63.

SOULODRE, G. A. Evaluation of objective loudness meters. In: **Audio Engineering Society Convention 116**. [s.n.], 2004. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=12790>>. Citado 5 vezes nas páginas 10, 31, 35, 105 e 106.

SOULODRE, G. A.; LAVOIE, M. C. Stereo and multichannel loudness perception and metering. In: **Audio Engineering Society Convention 119**. [s.n.], 2005. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=13330>>. Citado 2 vezes nas páginas 35 e 106.

STECKER, G. C.; HAFTER, E. R. An effect of temporal asymmetry on loudness. **The Journal of the Acoustical Society of America**, v. 107, n. 6, p. 3358–3368, 2000. Disponível em: <<http://dx.doi.org/10.1121/1.429407>>. Citado na página 184.

STEPHENS, S. Auditory temporal integration as a function of intensity. **Journal of Sound and Vibration**, v. 30, n. 1, p. 109 – 126, 1973. ISSN 0022-460X. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0022460X73800549>>. Citado na página 60.

STEVENS, J. C.; HALL, J. W. Brightness and loudness as functions of stimulus duration. **Perception & Psychophysics**, v. 1, n. 5, p. 319–327, 1966. ISSN 1532-5962. Disponível em: <<http://dx.doi.org/10.3758/BF03207399>>. Citado na página 63.

STEVENS, S. S. The measurement of loudness. **The Journal of the Acoustical Society of America**, v. 27, n. 5, p. 815–829, 1955. Disponível em: <<http://dx.doi.org/10.1121/1.1908048>>. Citado 2 vezes nas páginas 48 e 87.

_____. On the psychophysical law. **Psychological review**, American Psychological Association, v. 64, n. 3, p. 153–181, 1957. Citado 7 vezes nas páginas 33, 47, 48, 65, 86, 92 e 95.

_____. Procedure for calculating loudness: Mark vi. **The Journal of the Acoustical Society of America**, v. 33, n. 11, p. 1577–1585, 1961. Disponível em: <<http://dx.doi.org/10.1121/1.1908505>>. Citado 8 vezes nas páginas 9, 34, 38, 84, 85, 86, 87 e 88.

STEVENS, S. S.; GUIRAO, M. Loudness, reciprocity, and partition scales. **The Journal of the Acoustical Society of America**, ASA, v. 34, n. 9B, p. 1466–1471, 1962. Citado na página 183.

SUGIMOTO, T.; OODE, S.; NAKAYAMA, Y. Downmixing method for 22.2 multichannel sound signal in 8K super hi-vision broadcasting. **J. Audio Eng. Soc**, v. 63, n. 7/8, p. 590–599, 2015. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=17845>>. Citado na página 130.

SUPPER, B. **An Onset-Guided Spatial Analyser for Binaural Audio**. Tese (Doutorado) — University of Surrey, 2005. Citado na página 256.

TERVO, S.; PÄTYNEN, J.; KUUSINEN, A.; LOKKI, T. Spatial decomposition method for room impulse responses. **J. Audio Eng. Soc**, v. 61, n. 1/2, p. 17–28, 2013. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=16664>>. Citado 5 vezes nas páginas 156, 161, 163, 164 e 165.

TRAVAGLINI, A. Broadcast loudness: Mixing, monitoring and control. In: **Audio Engineering Society Convention 122**. [s.n.], 2007. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=14029>>. Citado na página 122.

TRAVAGLINI, A.; ALEMANNI, A.; LANTINI, F. Defining the listening comfort zone in broadcasting through the analysis of the maximum loudness levels. In: **Audio Engineering Society Convention 132**. [s.n.], 2012. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=16251>>. Citado na página 126.

TRAVAGLINI, A.; ALEMANNI, A.; UNCINI, A. HELM: High efficiency loudness model for broadcast content. In: **Audio Engineering Society Convention 132**. [s.n.], 2012. Disponível em: <<http://www.aes.org/e-lib/browse.cfm?elib=16250>>. Citado na página 125.

UNIÃO EUROPEIA DE RADIODIFUSÃO. **EBU R 128: 2014**: Loudness normalization and permitted maximum level of audio signals. Genebra, 2014. 5 p. Citado 9 vezes nas páginas 32, 35, 38, 113, 127, 172, 173, 178 e 179.

_____. **EBU R 128: s1-2016**: Loudness parameters for short-form content (advertisements, promos etc.). Genebra, 2016. 4 p. Citado 8 vezes nas páginas [32](#), [127](#), [138](#), [147](#), [174](#), [175](#), [178](#) e [179](#).

_____. **EBU Tech 3341-2016**: Loudness metering: ‘ebu mode’ metering to supplement loudness normalisation in accordance with ebu r 128. Genebra, 2016. 12 p. Citado 3 vezes nas páginas [113](#), [173](#) e [175](#).

_____. **EBU Tech 3342-2016**: Loudness range: A measure to supplement loudness normalisation in accordance with ebu r 128. Genebra, 2016. 9 p. Citado 5 vezes nas páginas [114](#), [116](#), [117](#), [173](#) e [174](#).

_____. **EBU Tech 3343-2016**: Guidelines for production of programmes in accordance with ebu r 128. Genebra, 2016. 9 p. Citado 4 vezes nas páginas [116](#), [173](#), [174](#) e [175](#).

UNIÃO INTERNACIONAL DE TELECOMUNICAÇÕES - SETOR DE PADRONIZAÇÃO. **ITU-T P 862: 2001**: Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. Genebra, 2001. 30 p. Citado na página [94](#).

_____. **ITU-T P 1401: 2012**: Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models. Genebra, 2012. 24 p. Citado na página [297](#).

UNIÃO INTERNACIONAL DE TELECOMUNICAÇÕES - SETOR DE RADIOCOMUNICAÇÃO. **ITU-R BS 468-4: 1986**: Measurement of audio-frequency noise voltage level in sound broadcasting. Genebra, 1986. 7 p. Citado 2 vezes nas páginas [104](#) e [105](#).

_____. **ITU-R BS 1387-1: 2001**: Method for objective measurements of perceived audio quality. Genebra, 2001. 100 p. Citado na página [94](#).

_____. **ITU-R WP6C Document 6C/490-E: 2011**: Chairman report. Genebra, 2011. 8 p. Citado na página [125](#).

_____. **ITU-R WP6C Document 6C/539-E: 2011**: Loudness metering algorithm: further analysis on 2003 listening test results. Genebra, 2011. 8 p. Citado na página [125](#).

_____. **ITU-R BS 1771-1: 2012**: Requirements for loudness and true-peak indicating meters. Genebra, 2012. 14 p. Citado 2 vezes nas páginas [114](#) e [123](#).

_____. **ITU-R BS 775-3: 2012**: Multichannel stereophonic sound system with and without accompanying picture. Genebra, 2012. 25 p. Citado na página [109](#).

_____. **ITU-R BS 2051-0: 2014**: Advanced sound system for programme production. Genebra, 2014. 14 p. Citado 3 vezes nas páginas [161](#), [252](#) e [253](#).

_____. **ITU-R BS 2266-1: 2014**: Framework of future audio broadcasting systems. Genebra, 2014. 10 p. Citado 3 vezes nas páginas [131](#), [288](#) e [290](#).

_____. **ITU-R Document 6C/353-E**: A work plan to revise recommendation itu-r bs.1770-3. studies of algorithm to measure loudness of advanced sound systems. Geneva, 2014. 6 p. Citado 7 vezes nas páginas [18](#), [248](#), [249](#), [260](#), [263](#), [265](#) e [278](#).

_____. **ITU-R WP6C Document 6C/353-E: A work plan to revise Recommendation ITU-R BS.1770-3**: Studies of algorithm to measure loudness of advanced sound systems. Genebra, 2014. 7 p. Citado 2 vezes nas páginas [17](#) e [128](#).

_____. **ITU-R WP6C Document 6C/380-E Annex 13: 2014**: Chairman report. Genebra, 2014. 8 p. Citado 2 vezes nas páginas [182](#) e [289](#).

_____. **ITU-R BS 1116-3: 2015**: Methods for the subjective assessment of small impairments in audio systems. Genebra, 2015. 32 p. Citado 8 vezes nas páginas [11](#), [13](#), [171](#), [186](#), [200](#), [201](#), [207](#) e [229](#).

_____. **ITU-R BS 1770-4: 2015**: Algorithms to measure audio programme loudness and true-peak audio level. Genebra, 2015. 24 p. Citado 29 vezes nas páginas [13](#), [17](#), [31](#), [32](#), [35](#), [38](#), [106](#), [107](#), [108](#), [109](#), [110](#), [111](#), [112](#), [113](#), [123](#), [126](#), [129](#), [166](#), [172](#), [173](#), [176](#), [177](#), [178](#), [215](#), [249](#), [263](#), [278](#), [288](#) e [295](#).

_____. **ITU-R BS 2217-2 Compliance material for Recommendation ITU-R BS.1770**. Geneva, 2016. Citado na página [193](#).

_____. **ITU-R WP6C Annex 19 to Document 6C/60-E**: Continuation of a rapporteur group (rg-32) on loudness measurement algorithm for the advanced sound system with extra terms of reference. Genebra, 2016. 3 p. Citado 3 vezes nas páginas [130](#), [182](#) e [289](#).

_____. **ITU-R BS 2076-1: Audio Definition Model**: Audio definition model. Geneva, 2017. 24 p. Citado 4 vezes nas páginas [217](#), [290](#), [291](#) e [305](#).

VICKERS, E. The loudness war: Background, speculation, and recommendations. **Audio Engineering Society Convention 129**, p. 1–27, Novembro 2010.

Disponível em: <<http://www.aes.org/e-lib/online/browse.cfm?elib=15598>>. Citado 3 vezes nas páginas 32, 150 e 334.

VIEMEISTER, N. F. Temporal modulation transfer functions based upon modulation thresholds. **The Journal of the Acoustical Society of America**, v. 66, n. 5, p. 1364–1380, 1979. Disponível em: <<http://dx.doi.org/10.1121/1.383531>>. Citado na página 44.

VORLÄNDER, M. **Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality**. Springer Berlin Heidelberg, 2007. (RWTHedition). ISBN 9783540488309. Disponível em: <<https://books.google.com.br/books?id=CuXF3JkTuhAC>>. Citado 3 vezes nas páginas 156, 157 e 166.

VYAS, A.; KANNAO, R.; BHARGAVA, V.; GUHA, P. Commercial block detection in broadcast news videos. In: **Proceedings of the 2014 Indian Conference on Computer Vision Graphics and Image Processing**. New York, NY, USA: ACM, 2014. (ICVGIP '14), p. 63:1–63:7. ISBN 978-1-4503-3061-9. Disponível em: <<http://doi.acm.org/10.1145/2683483.2683546>>. Citado 2 vezes nas páginas 140 e 154.

WARREN, R. M. Measurement of sensory intensity. **Behavioral and Brain Sciences**, Cambridge University Press, Cambridge, UK, v. 4, n. 2, p. 175–189, 006 1981. Disponível em: <<https://www.cambridge.org/core/article/div-class-title-measurement-of-sensory-intensity-div/747FD643AB33B08B2FACFE0628875F1F>>. Citado na página 53.

WARREN, R. M.; SERSEN, E. A.; PORES, E. B. A basis for loudness-judgments. **The American Journal of Psychology**, JSTOR, v. 71, n. 4, p. 700–709, 1958. Citado na página 183.

WEBER, E.; ROSS, H.; MURRAY, D. **E.H. Weber on the Tactile Senses**. Erlbaum (UK) Taylor & Francis, 1996. ISBN 9780863774218. Disponível em: <<https://books.google.com.br/books?id=xEd8JglYzFwC>>. Citado na página 46.

WELLS, W. **Measuring Advertising Effectiveness**. Taylor & Francis Group, 1997. (Advertising and consumer psychology). ISBN 9780805828122. Disponível em: <<https://books.google.com.br/books?id=-T7vtDatxHIC>>. Citado na página 32.

WENDT, F.; FRANK, M.; ZOTTER, F.; HÖLDRICH, R. Directivity patterns controlling the auditory source distance. In: **Proceedings of the International Conference on Digital Audio Effects**. [S.l.: s.n.], 2016. p. 295–303. Citado na página 184.

ZAHORIK, P. Direct-to-reverberant energy ratio sensitivity. **The Journal of the Acoustical Society of America**, v. 112, n. 5, p. 2110–2117, 2002. Disponível em: <<https://doi.org/10.1121/1.1506692>>. Citado na página 219.

ZAHORIK, P.; BRUNGART, D. S.; BRONKHORST, A. W. Auditory distance perception in humans: A summary of past and present research. **ACTA Acustica united with Acustica**, S. Hirzel Verlag, v. 91, n. 3, p. 409–420, 2005. Citado 2 vezes nas páginas 183 e 219.

ZAHORIK, P.; WIGHTMAN, F. L. Loudness constancy with varying sound source distance. **Nature neuroscience**, Nature Publishing Group, v. 4, n. 1, p. 78, 2001. Citado na página 183.

ZEMACK, M. **Implementing Methods for Equal Loudness in Radio Broadcasting**. [S.l.]: Skolan för datavetenskap och kommunikation, Kungliga Tekniska högskolan, 2007. Citado na página 41.

ZÖLZER, U. **DAFX: Digital Audio Effects**. Wiley, 2011. ISBN 9780470979679. Disponível em: <<https://books.google.com.br/books?id=DX-mRhkJL74C>>. Citado 4 vezes nas páginas 32, 72, 73 e 148.

ZWICKER, E. Über psychologische und methodische grundlagen der lautheit. **Acta Acustica united with Acustica**, S. Hirzel Verlag, v. 8, n. 4, p. 237–258, 1958. Citado 2 vezes nas páginas 34 e 87.

_____. Ein verfahren zur beredingung der lautstärke. **Acta Acustica united with Acustica**, v. 10, n. 4, 1960. Citado 3 vezes nas páginas 84, 87 e 88.

_____. Subdivision of the audible frequency range into critical bands (frequenzgruppen). **The Journal of the Acoustical Society of America**, v. 33, n. 2, p. 248–248, 1961. Disponível em: <<http://dx.doi.org/10.1121/1.1908630>>. Citado 5 vezes nas páginas 8, 56, 58, 59 e 83.

_____. Dependence of post-masking on masker duration and its relation to temporal effects in loudness. **The Journal of the Acoustical Society of America**, v. 75, n. 1, p. 219–223, 1984. Disponível em: <<http://dx.doi.org/10.1121/1.390398>>. Citado 2 vezes nas páginas 97 e 98.

ZWICKER, E.; FASTL, H. Loudness. In: _____. **Psychoacoustics: Facts and Models**. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999. p. 203–238. ISBN 978-3-662-09562-1. Disponível em: <http://dx.doi.org/10.1007/978-3-662-09562-1_8>. Citado 6 vezes nas páginas 34, 38, 97, 166, 167 e 295.

_____. **Psychoacoustics: Facts and Models**. Springer Berlin Heidelberg, 2013. (Springer Series in Information Sciences). ISBN 9783662095621. Disponível em: <<https://books.google.com.br/books?id=WLvtCAAQBAJ>>. Citado 19 vezes nas páginas 33, 51, 53, 61, 62, 63, 64, 67, 68, 69, 70, 88, 90, 92, 93, 94, 98, 110 e 159.

ZWICKER, E.; FASTL, H.; DALLMAYR, C. Basic-program for calculating the loudness of sounds from their 1/3-oct. band spectra according to iso 532 b. **Acustica**, v. 55, p. 63–67, 1984. Citado 2 vezes nas páginas 34 e 88.

ZWICKER, E.; FASTL, H.; WIDMANN, U.; KURAKATA, K.; KUWANO, S.; NAMBA, S. Program for calculating loudness according to din 45631 (iso 532b). **Journal of the Acoustical Society of Japan (E)**, v. 12, n. 1, p. 39–42, 1991. Citado 10 vezes nas páginas 17, 34, 38, 44, 88, 89, 94, 96, 100 e 122.

ZWICKER, E.; FLOTTORP, G.; STEVENS, S. S. Critical band width in loudness summation. **The Journal of the Acoustical Society of America**, v. 29, n. 5, p. 548–557, 1957. Disponível em: <<http://dx.doi.org/10.1121/1.1908963>>. Citado 3 vezes nas páginas 54, 82 e 87.

ZWICKER, E.; TERHARDT, E. Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. **The Journal of the Acoustical Society of America**, v. 68, n. 5, p. 1523–1525, 1980. Disponível em: <<http://dx.doi.org/10.1121/1.385079>>. Citado na página 56.

ZWISLOCKI, J. J. Loudness as a function of sound intensity and duration: An analysis. **The Journal of the Acoustical Society of America**, v. 39, n. 6, p. 1262–1262, 1966. Disponível em: <<http://dx.doi.org/10.1121/1.1942917>>. Citado na página 63.

GLOSSÁRIO

Ambisonics: Introduzida por [Gerzon \(1985\)](#), é uma técnica de decomposição do campo sonoro numa estrutura modal e reprodução destes modos na localização do ouvinte.

Gated loudness: *Loudness* entrecortado. Medida de *loudness* conjugada com uma função portão na saída do bloco integrador que segmenta os limites de integração em pequenos intervalos sobrepostos de fechamento, e que descarta segmentos com nível de *loudness* inferior a limiares pré-estabelecidos ou calculados em tempo de execução.

Loudness War: Guerra de Intensidade de Áudio. Termo aplicado ao crescente incremento de intensidade na música gravada, conforme músicos, engenheiros de masterização e gravadoras aplicam compressão dinâmica e limitação numa tentativa de tornar suas gravações soarem mais intensas que as de seus competidores ([VICKERS, 2010](#)).

Loudness de curta duração: Medida de *loudness* feita numa janela deslizante retangular de tamanho superior ao da janela de *loudness* momentâneo, projetada para representar os ajustes de volume da audiência em resposta a uma variação indesejável de *loudness*.

Loudness médio: Medida de *loudness* baseada na integração da energia ao longo de toda a duração do áudio.

Loudness: Percepção de intensidade do som ou dos sinais de áudio quando estes são reproduzidos acusticamente, tratando-se de uma função complexa, que pode ser medida objetivamente por meio de algoritmos definidos na Recomendação ITU-R BS.1770-2 e na Recomendação EBU R-128- 2011; ([MC, 2012](#)).

Stopping Power: Capacidade de o anúncio impactar/parar o público alvo, atrair sua atenção a ponto de fazer com que o anúncio seja lido.

Streaming: Distribuição de conteúdo multimídia através da Internet.

Sweet-spot: Termo que descreve o ponto focal de um arranjo de alto-falantes, no qual um indivíduo é plenamente capaz de ouvir uma peça de áudio espacializado da maneira pretendida pelo produtor.

Tonmeister: Profissional de gravação e produção musical, com treinamento formal em música e conhecimentos de gravação de som, mixagem e masterização.

de facto: Expressão em latim que significa “na prática”, tendo como expressão antônima a *de jure*, que significa “pela lei” ou “na teoria”.

de jure: Expressão em latim que significa “pela lei”, “pelo direito”, em contraste com *de facto*, que significa justamente “de fato”, ou seja, algo praticado.

Anecoico: “An-ecoico”, sem ecos ou livre de reflexões.

Auralização: Conceito introduzido por [Kleiner, Dalenbäck e Svensson \(1993\)](#), que significa “recriar a impressão aural das características acústicas de um espaço”.

Compressão de faixa dinâmica: Operação de processamento de sinais que atenua sons intensos e/ou amplifica sons suaves.

Conteúdo de formato curto: Conteúdo audiovisual de curta duração. Exemplos: clipes, vinhetas, chamadas e comerciais.

Contralateral: Que se encontra do lado oposto; heterolateral.

Decibel: Unidade logarítmica usada para expressar a razão entre dois valores de uma quantidade física, sendo um destes um valor de referência.

Dicótico: Representação do som cujos sinais são diferentes para cada ouvido.

Diótico: Sinal em ambos os ouvidos.

Faixa de loudness: Faixa média de variação da intensidade percebida de um programa de áudio, descontada das passagens 10% mais suaves e das passagens 5% mais intensas.

Faixa dinâmica: Razão entre os maiores e os menores valores que uma dada quantidade pode assumir.

Fator de crista: Diferença entre níveis de pico e valores eficazes.

Ipsilateral: Que se encontra do mesmo lado; homolateral.

Mascaramento: Efeito no qual o limiar de audibilidade de um som aumenta na presença de outro som.

Meato acústico: Canal que se estende desde a concha até a membrana do tímpano, com a função de transmitir os sons captados pela orelha para o tímpano, além de servir de câmara de ressonância para algumas frequências.

Monótico: Sinal em apenas um ouvido.

Outlier: Elemento destacadamente diferenciado dos demais elementos de um conjunto. “Ponto fora da curva”.

Pico verdadeiro: Nível de pico do sinal digital sobreamostrado em no mínimo quatro vezes de forma a simular a condição do sinal tal como será reproduzido após ser convertido num sinal analógico.

Psicofísica: Termo que descreve o estudo interdisciplinar sobre como os seres humanos percebem magnitudes físicas.

Quantização: Processo de atribuição de valores discretos para um sinal cuja amplitude varia entre valores contínuos.

Recodificação (*Winsorizing*): Substituição de *outliers* pelo maior valor não-*outlier*.

Regularização: Técnica antisobreajuste (*overfitting*), que consiste em penalizar valores elevados dos parâmetros do modelo tal que, ao término da minimização da função objetivo tem-se um vetor de parâmetros da função discriminante linear \mathbf{w} de magnitude reduzida, que no espaço original de características de um problema não linearmente separável, se traduz numa região de separação menos sinuosa e, por consequência, menos aderente ao conjunto de dados de treinamento e com maior potencial de generalização para novos dados.

Remoção (*trimming*): Remoção de valores extremos do conjunto de dados.

Sobremodulação: Condição na qual o nível instantâneo do sinal modulante excede o valor necessário para a produção de 100% de modulação na portadora.

PUBLICAÇÕES

Segue uma relação de eventos e periódicos de divulgação científica nos quais os avanços do trabalho foram progressivamente relatados:

- PIRES, L. d. S. Medição de picos verdadeiros de sinais de áudio. In: **V Simpósio de Processamento de Sinais da UNICAMP**, 2014.
- PIRES, L. d. S.; VIEIRA, M. N.; YEHIA, H. C. *Dimensionality Reduction and Support Vector Machine Parameter Adjustments applied to speech impairments*. In: **5th EICEFALA International Meeting on Speech Sciences**, 2015.
- PIRES, L. d. S.; VIEIRA, M. N.; YEHIA, H. C. Controle automático de *loudness* em conteúdo de formato curto para radiodifusão. In: **Anais do 14o. Congresso de Engenharia de Áudio da AES Brasil**, 2016.
- PIRES, L. d. S.; VIEIRA, M. N.; YEHIA, H. C. *Automatic loudness control in short-form content for broadcasting*. **The Journal of the Acoustical Society of America**, v. 141, n. 3, p. EL287-EL292, 2017.
- PIRES, L. d. S.; VIEIRA, M. N.; YEHIA, H. C.; PASQUAL A. M. Medição de *loudness* em áudio imersivo para radiodifusão através de técnicas de auralização. In: **XXVII Encontro da Sociedade Brasileira de Acústica**, 2017.

- PIRES, L. d. S.; VIEIRA, M. N.; YEHIA, H. C.; PASQUAL A. M.; BROOKES, T. S.; MASON, R. D. Modelo de distância auditiva percebida para o algoritmo de *loudness* ITU-R BS.1770. In: **XXVIII Encontro da Sociedade Brasileira de Acústica**, 2018.
- PIRES, L. d. S.; VIEIRA, M. N.; YEHIA, H. C. Ajuste do algoritmo de *loudness* ITU-R BS.1770 baseado em reverberação. In: **XXXVII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais**, 2019 (*pendente de publicação até a data desta redação*).

LINKS INTERESSANTES

<https://www.itu.int/en/ITU-R/study-groups/rsg6/rwp6c/Pages/default.aspx>

Página do Grupo de Trabalho em produção de programas e avaliação de qualidade do ITU-R, do qual o grupo relator de *loudness* (RG 32) é parte integrante;

<https://tech.ebu.ch/loudness> Página do grupo de trabalho em *loudness* P-LOUD da EBU, responsável pela Recomendação de *loudness* R.128 e por elaborar guias práticos voltados a radiodifusores e produtores de programas;

<http://www.anatel.gov.br/legislacao/leis/467-lei-10222> Lei nº 10.222, de 9 de maio de 2001, que padroniza o volume de áudio das transmissões de rádio e televisão nos espaços dedicados à propaganda e dá outras providências;

<https://www2.camara.leg.br/comunicacao/rede-legislativa-radio-tv/> Página da Rede Legislativa de Rádio e TV, que hospeda a Portaria nº 354 de 11 de julho de 2012, do antigo Ministério das Comunicações, que regulamenta a padronização do volume de áudio nos intervalos comerciais da programação dos serviços de radiodifusão sonora e de sons e imagens nos termos da Lei nº 10.222, de 9 de maio de 2001;

www.sinaprosp.org.br/download.php?arquivo=adm/upload/n072013B.pdf

Comunicado da Associação Brasileira das Produtoras de Fonogramas Pu-

blicitários (Aprosom) sobre normas para controle de *loudness* e seu efeito na produção do áudio publicitário.

DOCUMENTOS DE APOIO

Este anexo é composto dos documentos que deram suporte aos experimentos realizados, conforme relação abaixo:

Formulário de avaliação de riscos

Formulário de auto-avaliação ética

Experimento de distância: folha de recrutamento

Experimento de distância: folha de informações

Experimento de distância: formulário de consentimento

Experimento de reverberação: folha de informações ¹

Experimento de reverberação: formulário de consentimento

Experimento de direção: folha de recrutamento

Experimento de direção: folha de informações

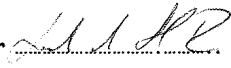
Experimento de direção: formulário de consentimento

¹ Neste experimento, o recrutamento foi registrado online em <http://doodle.com>.

Date assessment issued: 12 January 2018	Description of area or activity being assessed: <ul style="list-style-type: none"> The activity consists on evaluating loudness of sounds played at different distances. Eight listeners will be asked to match the loudness of each stimulus to that of a provided reference during one-hour sessions. Location: 43BC02 				
Planned review date:					
Retention period (+4 yrs from issue):					
Activity Hazard Level - Reference "Lone and Hazardous Working Policy" (see note 1)	Low Hazard	X	Medium Hazard	High Hazard	

Declaration:

To the best of my/our knowledge this document is an accurate assessment of the known and foreseeable risks and of the safety precautions which are to be followed.

Signature of assessor  Name (print)

LEANDRO PIRES

Date: 15 Jan 2018

Managers Approval:

I have reviewed this risk assessment in consultation with the assessor and accept the issues identified. The actions defined in this risk assessment will be taken in order to reduce residual risks to a level that is as low as reasonably practicable.

Signature of manager  Name (print)

TIM BROOKES

Date: 15 Jan 2018

Required only for High Hazard Activity (reference - Lone and Hazardous Working Policy)**Head of Department (or equivalent e.g. Director) Approval:**

I have reviewed this risk assessment in consultation with the Faculty/Area Health & Safety Advisor and accept the issues identified. The actions defined in this risk assessment will be taken in order to reduce residual risks to a level that are as low as reasonably practicable.

Signature of manager Name (print)

Date:

Signature of Safety Advisor Name (print)

Date:

DESCRIPTION OF ACTIVITY or FACILITY AND ITS USE

Use this area to describe the area and/or the main activities to be covered by this risk assessment:

- Five loudspeakers will be placed in fixed distances from the listener position (Max distance = 4.5 m) and the test/reference programme stimuli will be presented to listeners in a random fashion. The process is called "method of adjustment" in Psychophysics, where subjects are asked to adjust the level of a stimulus until is detectable against background noise or is the same level of a reference stimulus. The experiment will focus on the latter.
- 43BC02 is a reverberant room, wide enough to spread loudspeakers, place equipment and comfortably accommodate listeners.

SAFETY RULES AND GENERAL COMMENTSKey Findings:

N/A

Key Comments:

N/A

note: see detailed assessment and actions list below

HAZARDS

Identify significant hazards relevant to this risk assessment

Flammable / Explosive Substances	Hazardous Waste Disposal		Storage / Housekeeping	Temperature	Travel Health	
Ionising / Non-Ionising Radiation	Discharge / Spill		Falling Objects	Humidity	Stress	
Exposure to Hazardous Substances	Slips, Trips & Falls		Machinery / Power Tools	Lighting	Out of Hours Working / Lone Working	
Biological Hazards	Electrical Safety	X	Hygiene	Noise	X	Personal Security
Cryogenic Hazard	Manual Handling	X	Welfare	Vibration		
Chemical Storage	Working at Height		Pressure / Vacuum Systems	Access / Egress	Display Screen Equipment	

WHO IS AT RISK

Identify groups of individuals who need to be considered as part of this risk assessment

Staff	Contractors	Visitors	Others	Higher Risk groups
Employees	Cleaners	Visitors	Neighbors	Young Persons
Temporary Staff	Maintenance Engineers	Customers	Members of the Public	Disabled Persons
Operatives	Security	Delivery Staff	Environment	Children
Academics	Catering Staff		Wildlife	Pregnant & Nursing Mothers
Students	X Contractors			Lone Workers


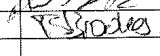
ASSESSMENT OF RISK

Assess the risks against each significant hazard group identified

Significant Hazard	Perceived Nature of Risk	Existing Control Measures	Residual Risk Low/Medium/ High	Further action required Y/N
Noise	Sound pressure levels of noise exposure (ISO 1999): 140 dB (peak limit), 85 dB (upper exposure action limit for 8h work) and 80 (lower exposure action limit for 8h work)	<ul style="list-style-type: none"> ■ Limited sessions per participant ■ Sound reproduction comfortably within safety levels (~70 dB SPL_A at listener position) 	Low	N
Electrical Safety	Mains-powered electrical equipment will be used.	All mains-powered equipment will be checked to ensure that it has an in-date PAT 'passed' sticker.	Low	N
Manual Handling	Installation may involve moving loudspeakers and tables.	Lightweight (<4kg) loudspeakers will be used. Manual handling guidelines will be followed.	Low	Y

ACTION PLAN

Develop a prioritised action plan to support the risk assessment

Action to be taken to further reduce risk	Person responsible for completing action	Target completion date (Prioritized on risk)		Action closure	
		Date	Priority	Signature	Date
Familiarise self with HSE manual handling guidelines.	L. PIRES	15 Jan 2018			15 Jan 2018
Send HSE manual handling guidelines to Leandro.	T. BROOKES	15 Jan 2018			15 Jan 2018

Subsequent assessment review: Risk assessments require review and in some cases revision to ensure the assessment continues to reflect current working practices e.g. a review should be initiated in response to significant changes to the area / activity or if an accident / incident has occurred.

Review undertaken on:

Comments:

NOTES:

Note 1.

Allocate this activity to one of the 3 Hazard Levels (see Lone and Hazardous Working Policy for full details)

e.g.

Low Hazard - Office based lone working, Lone working collecting routine data/non-hazardous procedure in Laboratory (Avoid use of hazardous materials, sharps, chemicals, etc)

Medium Hazard - Working within Workshop/Laboratory

High Hazard - Work within specialised Unit e.g. Containment Level 3, High Toxicity, High Voltage

Category Control Measures

The following risk control measures should be applied for each hazard category along with any other additional measures identified by the risk assessment. The overriding principle is to reduce the risk to the lowest level achievable.

Low Hazard Activities

- Can only be undertaken by lone workers who are familiar with the premises and are aware of the emergency procedures.
- Must be authorised and suitably controlled by the line manager/supervisor (or equivalent e.g. Principal Investigator) and can be verbal.

Medium Hazard Activities

- Can only be undertaken by competent persons if there is at least one other person in the vicinity (either in the same area or close by) who is competent to make safe any work being undertaken and is also familiar with any emergency procedures for the area.
- Must be authorised (using risk assessment form) by the line manager/supervisor (or equivalent e.g. Principal Investigator).

High Hazard Activities

- Can only be undertaken by competent persons if there is at least one other person in the same location who is competent to make safe any work being undertaken and is also familiar with any emergency procedures.
- Suitable emergency arrangements, such as the provision of adequate first aid or fire safety measures must be in place. (additional measures may be required depending on the time of day when the work is undertaken)
- Must be authorised (using risk assessment form) by the Head of Department (or equivalent e.g. Director) after consultation with the Faculty/Area Health & Safety Advisor.

SAFE

Response ID	Completion date
353003-352994-36010101	6 Jun 2018, 10:04 (BST)

1	Project title	Directional loudness
----------	----------------------	----------------------

2	Principal Investigator	Leandro da Silva Pires
2.a	Email address	l.dasilvapires@surrey.ac.uk

3	Level of research	PhD or EngD
3.b	If this is a PhD or EngD study, please provide the names of your supervisors.	Tim Brookes Russell Mason

4	School/ Department. External applicants should list their affiliated body or institution.	Department of Music and Media
----------	--	-------------------------------

5	Proposed start date for data collection	07/06/2018
----------	--	------------

6	End date for data collection	20/06/2018
----------	-------------------------------------	------------

7	Does the study fall into any of the following categories?	f. None of the above
----------	--	----------------------

8	Are there any procedures involving more than minimal risk to a participant's health or well-being?	No
----------	---	----

9	Does the study involve the use of surveys, questionnaires and any research, the nature of which might be offensive, distressing or deeply personal for the particular target group, where the participants will be identifiable to the researchers e.g. interviews, focus groups?	No
----------	--	----

10	Does the study involve children under the age of 16 or other vulnerable groups, or those who may feel under pressure to take part due to their connection with the researcher?	No
-----------	---	----

11	Does the study involve prisoners or young offenders?	No
-----------	---	----

12	Does the study involve the new collection or donation of human tissue, as defined by the Human Tissue Act, from a living person or the recently deceased according to the Human Tissue Authority?	No
-----------	--	----

13	Does the research require participants to take part in the study without their knowledge and/or consent at the time?	No
-----------	---	----

14	Does the study involve deception other than withholding information about the aims of the research until the debriefing?	No
----	---	----

15	Do you think that any other significant ethical concerns may arise, or does your external funding body or sponsor require full ethical review to be undertaken?	No
----	--	----

16	Have you answered Yes to any of the questions on this page?	No- I don't need to apply for full review
----	--	---

25	Does the study involve the use of surveys, questionnaires and any research, the nature of which might be offensive, distressing or deeply personal for the particular target group, where the participants will not be identifiable to the researchers e.g. online surveys, anonymous questionnaires?	No
----	--	----

26	Are you planning to access records of and/or collect personal confidential data, concerning identifiable individuals as defined by data protection legislation?	No
-----------	--	----

27	Are you linking or sharing personal data, special category data (sensitive personal data) or confidential information beyond the initial consent given (including linked data gathered outside of the UK)?	No
-----------	---	----

28	Will you collect or access audio recordings, video recordings, photographs or quotations within which participants may be identifiable and with the intention to disseminate them beyond the research team?	No
-----------	--	----

29	Do you plan to offer incentives which may unduly influence participants' decision to participate?	No
-----------	--	----

30	Does the study involve activities where the safety/wellbeing of the researcher may be in question?	No
-----------	---	----

31	Could the behavioural/physiological intervention possibly lead to discovery of ill health or concerns about wellbeing in a participant incidentally, even if the intervention in itself causes no more than minimal stress is to the research participant?	No
-----------	---	----

32	<p>Are you investigating existing working or professional practices among participants, identifiable to yourself as the researcher at your own place of work (this may be the University of Surrey or another organisation where you, your supervisor or co-investigator works)?</p>	No
----	---	----

33	<p>Have you answered Yes to any of the questions on this page?</p>	No- I don't need to apply for proportionate review
----	---	--

37	<p>According to the answers you have submitted your research project does not require review by the UEC.</p>	I confirm that I have answered No to all questions. I understand that my completed form may be audited.
----	---	---

38	<p>I, the undersigned, confirm that I have read and will comply with the Ethics Handbook for Teaching and Research and the Code on Good Research Practice. I understand that the project may be monitored and audited by the University of Surrey to ensure that it is carried out in</p>	I agree
----	--	---------

accordance with good practice, legal and ethical requirements and any other guidelines. I understand that the protocol and any associated documents such as information sheets and consent forms should have version numbers and dates. If I make any significant changes to my protocol I understand that I should complete the self-assessment again. I am also aware that any knowingly wrong answer to any of the questions below and any research misconduct reported may lead to disciplinary measures after investigation. In case of dissertation projects or theses, the provision of knowingly incorrect information or proven research misconduct may affect academic progression.

38.a

Name

Leandro da Silva Pires

Loudness Listening Test (Dec 2017 / Jan 2018)

TB7 / 09BC03 -- University of Surrey

The test will take place in two locations:

- 1) Listening room: room 7 of Teaching Block building (TB7), from Dec 11th to 15th (Week 11)
- 2) IoSR research group: room 9, floor 3 of the James Joule building (09BC03), from Jan 15th to 19th (Week 13)

Please:

- Sign out for two one-hour slots, one in each table/location.
- Write your contact details on the back of this sheet

You will receive a £5 Amazon gift card for completing both sessions.
Thanks for your time and ears.

Teaching Block building (TB7)

Mon	Tue	Wed	Thur	Fri
9:00	9:00	9:00	9:00	9:00
10:30	10:30	10:30	10:30	10:30
12:00	12:00	12:00	12:00	12:00
1:30	1:30	1:30	1:30	1:30
3:00	3:00	3:00	3:00	3:00
4:30	4:30	4:30	4:30	4:30

James Joule building (09BC03)

Mon	Tue	Wed	Thur	Fri
9:00	9:00	9:00	9:00	9:00
10:30	10:30	10:30	10:30	10:30
12:00	12:00	12:00	12:00	12:00
1:30	1:30	1:30	1:30	1:30
3:00	3:00	3:00	3:00	3:00
4:30	4:30	4:30	4:30	4:30

Listening Tests: General Information for Participants

Thank you for agreeing to take part in one of our listening tests. Listening tests are an important part of our research into reproduced sound quality and we rely upon a committed panel of listeners to provide us with their responses to a range of different sounds.

The sorts of responses that you will typically be asked for fall into the following categories:

- Ratings of sound quality attributes on scales provided by the experimenter
- Descriptions of sound quality using your own terms or drawings
- Preference or 'liking' responses
- Judgements of differences and/or similarities between sounds

Sometimes sound reproduction will be accompanied by video pictures and you may be asked to undertake a task relating to the picture or the sound, such as playing a game, following an object or identifying features in the scene.

We may ask you to undergo a screening exercise prior to selecting you for a listening panel. This is normally to determine factors such as your consistency of judgement and your sensitivity to the attributes under investigation. All data relating to such screening tests is stored confidentially and anonymously. In some cases you will subsequently be trained in the identification and scaling of the sound quality attributes under investigation.

There are normally no right or wrong answers during the listening test proper – in other words it is not you that is under test. We are interested in your responses to the sounds that we present because of what they tell us about the way sound signals are perceived and/or described.

Health and Safety

Listening tests are structured in such a way as to allow for breaks so as to avoid listener fatigue. The precise time commitment required of you will be agreed beforehand.

Sound reproduction levels are controlled so as to be within safe limits as recommended by UK legislation.

Although we naturally hope for your continued commitment to the listening test, you are free to opt out of it at any time without giving a reason and without prejudice.

Data Archival

The data we gather from you during the course of the listening test will be stored anonymously and may be made publicly available.

Listening Test Consent Form

This form is to be completed by any subject that agrees to take part in a listening test, before the test begins.

I the undersigned voluntarily agree to take part in the study on

_____ (project title)

I have read and understood the information sheet provided. I have been given a full explanation by the investigators of the nature, purpose, location and likely duration of the listening test, and of what I will be expected to do. I have been given the opportunity to ask questions on all aspects of the listening test and have understood the advice and information given as a result.

I agree to comply with any instructions given to me during the listening test and to cooperate fully with the investigators.

I understand that all documentation held on a volunteer is in the strictest confidence and complies with the Data Protection Act (1998). I agree that provided that my anonymity is preserved I will not seek to restrict the use of the results and consent to them being made publicly available.

I understand that I am free to withdraw from the listening test at any time without needing to justify my decision and without prejudice.

I confirm that I have read and understood the above and freely consent to participating in this study. I have been given adequate time to consider my participation and agree to comply with the instructions and restrictions of the study.

Name of volunteer _____

(block capitals)

Signed _____

Date _____

Name of witness _____

(block capitals)

Signed _____

Date _____

FOLHA DE INFORMAÇÕES PARA PARTICIPANTES¹

Título do estudo: Intensidade Sonora Percebida (*Loudness*) Direcional

VOCÊ LEVARÁ UMA CÓPIA DESTA FOLHA DE INFORMAÇÕES

Parágrafo de convite

Eu sou estudante do Programa de Pós-Graduação em Engenharia Elétrica. Gostaria de convidá-lo para participar neste projeto que é parte integrante da minha pesquisa de doutorado. Você deverá participar somente se assim o desejar. Escolher não participar não lhe trará desvantagens de qualquer natureza. Antes de decidir se gostaria de fazer parte, é importante que você entenda o porquê da condução desta pesquisa e o que sua participação envolverá. Por favor, tire um tempo para ler cuidadosamente as informações a seguir e discutí-las com outras pessoas se quiser. Caso algo não esteja claro ou você necessitar de mais informações, pergunte-me.

Qual é o propósito do estudo?

O objetivo deste estudo é investigar como o algoritmo de medição de intensidade sonora percebida (*loudness*) da União Internacional de Telecomunicações (UIT) poderia contemplar medidas de objetos sonoros de localização dinâmica. Eu estou especificamente interessado nas sensibilidades de *loudness* provocadas por fontes sonoras estáticas. Isso envolverá tarefas de casamento de *loudness* com o propósito de quantificar o quão distinta a intensidade sonora pode ser percebida ao longo de diferentes direções.

Por que eu fui convidado(a) para fazer parte?

Eu estou convidando ouvintes com audição normal e com a habilidade tanto de entender as instruções quanto de executar as tarefas solicitadas.

Eu tenho que participar?

A participação é voluntária. Você não é obrigado(a) a participar. É importante ler esta folha de informações e, se você tiver quaisquer dúvidas, pergunte-me.

O que acontecerá comigo se eu participar?

Se você decidir participar, uma cópia desta folha de informações lhe será entregue e você será requisitado(a) a assinar um formulário de consentimento. Eu então irei instruí-lo(a) na tarefa de casamento de *loudness* e o(a) direcionarei até a cabine de áudio onde o teste será conduzido. Os procedimentos levarão duas sessões de aproximadamente uma hora cada, que serão realizadas na Sala 2505 da Escola de Engenharia

¹Esclarecimentos prestados em conformidade com Art. 10 da Resolução COEP nº 510/2016

da UFMG. Nenhuma captura de áudio ou vídeo será feita. Somente seu nome completo, informações para contato, e assinatura serão colhidos para registrar seu consentimento. Você também poderá ser contatado(a) para pesquisa futura, se necessário.

Quais os possíveis benefícios e riscos da participação?

As informações extraídas nesse estudo darão suporte ao Grupo Relator de Medidas de Loudness da UIT para futuras revisões do algoritmo de medida, auxiliando no aprimoramento das recomendações vigentes no assunto.

A desvantagem principal de participar deste estudo é alguma possibilidade de irritação. Testes de escuta são planejados para permitirem intervalos de modo a se evitar a fadiga do ouvinte. Os níveis de reprodução sonora são controlados de tal forma a se enquadrarem nos limites estipulados pela Lei Ordinária nº 9.505/2008 do Município de Belo Horizonte.

Minha participação será mantida em sigilo?

O que é dito em entrevista e/ou o que é coletado é considerado estritamente confidencial e todos os dados para análise serão anonimizados e pseudoaleatorizados. Em futuros relatos sobre as descobertas da pesquisa, eu não revelarei os nomes de quaisquer participantes ou da instituição onde você estuda/trabalha.

Todos os dados relacionados a esta etapa da pesquisa (ex.: formulário de consentimento) serão guardados por 5 anos. Seus dados pessoais serão guardados e processados em conformidade com a Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709 de 14/08/2018).

O que acontecerá com os dados e os resultados do estudo?

Eu produzirei um relato resumindo as principais descobertas. Também planejo disseminar as descobertas da pesquisa por meio de publicação e conferências. Estes dados também poderão ser utilizados em pesquisas futuras.

Quem eu devo contatar para obter informações adicionais?

Se você tiver quaisquer dúvidas ou necessitar de mais informações sobre este estudo, por favor contate-me no endereço de correio eletrônico *leandropires@ufmg.br*.

E se algo der errado?

Se desejar fazer uma reclamação sobre a condução do estudo, você poderá contatar o Professor Hani Yehia:

hani@cpdee.ufmg.br

Centro de Estudos da Fala, Acústica, Linguagem e música – CEFALA

Departamento de Engenharia Elétrica

Obrigado por ler esta folha de informações e por considerar fazer parte desta pesquisa.

FORMULÁRIO DE CONSENTIMENTO PARA PARTICIPANTES DE EXPERIMENTOS¹

Por favor complete este formulário depois de ler a Folha de Informações e/ou ouvir uma explanação sobre a pesquisa.

Título do estudo: Intensidade Sonora Percebida (*Loudness*) Direcional

Obrigado por considerar sua participação neste experimento. O pesquisador deve explicar-lhe o projeto antes de sua concordância em participar. Se você tem alguma pergunta oriunda da Folha de Informações ou da explanação feita, por favor pergunte ao pesquisador antes que você decida por integrar o experimento. Você terá uma via deste Formulário de Consentimento para guarda e referência a qualquer tempo.

Ao assinalar/marcas cada caixa, você está consentindo com o elemento correspondente. Presumir-se-á que caixas não assinaladas/marcadas signifiquem que você NÃO consente com o elemento correspondente e que você poderá ser considerado inelegível para o estudo.

- Eu confirmo que li e compreendi a folha de informações datada de 4 de setembro de 2018 para o estudo acima. Tive a oportunidade de considerar as informações e de tirar dúvidas as quais foram satisfatoriamente esclarecidas.
- Eu entendo que minha participação é voluntária e que sou livre para me retirar a qualquer tempo durante o estudo sem justificativa e sem ser prejudicado de maneira alguma. Além disso, entendo que estarei apto a retirar meus dados por até um mês depois do teste.
- Eu consinto com o processamento das minhas informações pessoais para os propósitos a mim explicados. Entendo que a lida com tais informações se dará em conformidade com regulações vigentes de proteção de dados.
- Eu entendo que minhas informações poderão estar sujeitas à revisão por servidores competentes da UFMG e/ou por reguladores para fins de monitoramento e auditoria.
- Eu entendo que confidencialidade e anonimidade serão mantidos e que o pesquisador não me identificará em qualquer produto da pesquisa.
- Eu concordo em ser contatado no futuro por pesquisadores da UFMG que gostariam de me convidar para participar de estudos de continuidade deste projeto, ou em estudos futuros de natureza similar.
- Eu concordo que o núcleo de pesquisa pode usar meus dados anonimizados para pesquisas futuras e entendo que qualquer uso de dados identificáveis poderá ser revisado e aprovado por um comitê de ética de pesquisa.
- Eu entendo que não devo participar caso me enquadre em algum dos critérios de exclusão detalhados na folha de informações e a mim explicados pelo pesquisador.

Nome do Participante: _____

Data: _____ Assinatura: _____

¹Registro de consentimento em conformidade com Art. 10 da Resolução COEP nº 510/2016

LOUDNESS LISTENING TEST (JUN/2018)

TEACHING BLOCK ROOM 7 (TB7) – UNIVERSITY OF SURREY

I would like to invite you to participate in this experiment. I am specifically interested in directional loudness sensitivities of static sound sources. This will involve loudness matching tasks with the purpose of quantifying how different the perceived sound intensity is among different directions. The procedures take two sessions of approximately one hour each and they will be based at Teaching Block Room 7 (TB7). Please:

- Sign out for two one-hour slots.
- Write your contact details in the back of this sheet

You will receive a £5 Amazon gift card for each session. Thank you for your time and ears.
Leandro

THURSDAY	FRIDAY	SATURDAY	SUNDAY	MONDAY	TUESDAY	WEDNESDAY
JUNE 7 <hr/> 1:00pm-2:00pm Name: _____ <hr/> 2:20pm-3:20pm Name: _____ <hr/> 3:40pm-4:40pm Name: _____ <hr/> 5:00pm-6:00pm Name: _____	JUNE 8 <hr/> 9:00am-10:00am Name: _____ <hr/> 10:20am-11:20am Name: _____ <hr/> 11:40am-12:40pm Name: _____ <hr/> 1:00pm-2:00pm Name: _____ <hr/> 2:20pm-3:20pm Name: _____ <hr/> 3:40pm-4:40pm Name: _____ <hr/> 5:00pm-6:00pm Name: _____	JUNE 9 <hr/> 9:00am-10:00am Name: _____ <hr/> 10:20am-11:20am Name: _____ <hr/> 11:40am-12:40pm Name: _____ <hr/> 1:00pm-2:00pm Name: _____ <hr/> 2:20pm-3:20pm Name: _____ <hr/> 3:40pm-4:40pm Name: _____ <hr/> 5:00pm-6:00pm Name: _____	JUNE 10 <hr/> 9:00am-10:00am Name: _____ <hr/> 10:20am-11:20am Name: _____ <hr/> 11:40am-12:40pm Name: _____ <hr/> 1:00pm-2:00pm Name: _____ <hr/> 2:20pm-3:20pm Name: _____ <hr/> 3:40pm-4:40pm Name: _____ <hr/> 5:00pm-6:00pm Name: _____	JUNE 11 <hr/> 9:00am-10:00am Name: _____ <hr/> 10:20am-11:20am Name: _____ <hr/> 11:40am-12:40pm Name: _____ <hr/> 1:00pm-2:00pm Name: _____ <hr/> 2:20pm-3:20pm Name: _____ <hr/> 3:40pm-4:40pm Name: _____ <hr/> 5:00pm-6:00pm Name: _____	JUNE 12 <hr/> 9:00am-10:00am Name: _____ <hr/> 10:20am-11:20am Name: _____ <hr/> 11:40am-12:40pm Name: _____ <hr/> 1:00pm-2:00pm Name: _____ <hr/> 2:20pm-3:20pm Name: _____ <hr/> 3:40pm-4:40pm Name: _____ <hr/> 5:00pm-6:00pm Name: _____	JUNE 13 <hr/> 9:00am-10:00am Name: _____ <hr/> 10:20am-11:20am Name: _____ <hr/> 11:40am-12:40pm Name: _____ <hr/> 1:00pm-2:00pm Name: _____ <hr/> 2:20pm-3:20pm Name: _____ <hr/> 3:40pm-4:40pm Name: _____ <hr/> 5:00pm-6:00pm Name: _____
THURSDAY	FRIDAY	SATURDAY	SUNDAY	MONDAY	TUESDAY	WEDNESDAY
JUNE 14 <hr/> 9:00am-10:00am Name: _____ <hr/> 10:20am-11:20am Name: _____ <hr/> 11:40am-12:40pm Name: _____ <hr/> 1:00pm-2:00pm Name: _____ <hr/> 2:20pm-3:20pm Name: _____ <hr/> 3:40pm-4:40pm Name: _____ <hr/> 5:00pm-6:00pm Name: _____	JUNE 15 <hr/> 9:00am-10:00am Name: _____ <hr/> 10:20am-11:20am Name: _____ <hr/> 11:40am-12:40pm Name: _____ <hr/> 1:00pm-2:00pm Name: _____ <hr/> 2:20pm-3:20pm Name: _____ <hr/> 3:40pm-4:40pm Name: _____ <hr/> 5:00pm-6:00pm Name: _____	JUNE 16 <hr/> 9:00am-10:00am Name: _____ <hr/> 10:20am-11:20am Name: _____ <hr/> 11:40am-12:40pm Name: _____ <hr/> 1:00pm-2:00pm Name: _____ <hr/> 2:20pm-3:20pm Name: _____ <hr/> 3:40pm-4:40pm Name: _____ <hr/> 5:00pm-6:00pm Name: _____	JUNE 17 <hr/> 9:00am-10:00am Name: _____ <hr/> 10:20am-11:20am Name: _____ <hr/> 11:40am-12:40pm Name: _____ <hr/> 1:00pm-2:00pm Name: _____ <hr/> 2:20pm-3:20pm Name: _____ <hr/> 3:40pm-4:40pm Name: _____ <hr/> 5:00pm-6:00pm Name: _____	JUNE 18 <hr/> 9:00am-10:00am Name: _____ <hr/> 10:20am-11:20am Name: _____ <hr/> 11:40am-12:40pm Name: _____ <hr/> 1:00pm-2:00pm Name: _____ <hr/> 2:20pm-3:20pm Name: _____ <hr/> 3:40pm-4:40pm Name: _____ <hr/> 5:00pm-6:00pm Name: _____	JUNE 19 <hr/> 9:00am-10:00am Name: _____ <hr/> 10:20am-11:20am Name: _____ <hr/> 11:40am-12:40pm Name: _____ <hr/> 1:00pm-2:00pm Name: _____ <hr/> 2:20pm-3:20pm Name: _____ <hr/> 3:40pm-4:40pm Name: _____ <hr/> 5:00pm-6:00pm Name: _____	JUNE 20 <hr/> 9:00am-10:00am Name: _____ <hr/> 10:20am-11:20am Name: _____ <hr/> 11:40am-12:40pm Name: _____ <hr/> 1:00pm-2:00pm Name: _____ <hr/> 2:20pm-3:20pm Name: _____ <hr/> 3:40pm-4:40pm Name: _____ <hr/> 5:00pm-6:00pm Name: _____



INFORMATION SHEET FOR PARTICIPANTS

Title of Study: Directional Loudness

University of Surrey Ref:: 353003-352994-36010101

YOU WILL BE GIVEN A COPY OF THIS INFORMATION SHEET

Invitation Paragraph

I am a visiting postgraduate researcher at the Department of Music and Media. I would like to invite you to participate in this research project which forms part of my PhD research. You should only participate if you want to; choosing not to take part will not disadvantage you in any way. Before you decide whether you want to take part, it is important for you to understand why the research is being done and what your participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask me if there is anything that is not clear or if you would like more information.

What is the purpose of the study?

The aim of this study is to investigate how the International Telecommunication Union (ITU) loudness measurement algorithm should include measurement of sound objects that have dynamic locations. I am specifically interested in directional loudness sensitivities of static sound sources. This will involve loudness matching tasks with the purpose of quantifying how different the perceived sound intensity is among different directions.

Why have I been invited to take part?

I am inviting normal-hearing listeners with the ability to understand the instructions and execute the tasks.

Do I have to take part?

Participation is voluntary. You do not have to take part. You should read this information sheet and if you have any questions you should ask the research team.

What will happen to me if I take part?

If you decide to take part you will be given this information sheet to keep and will be asked to sign a consent form. I will then instruct you on the loudness matching task and direct you to the listening room where the test is to be conducted. The procedures take two sessions of approximately one hour each and they will be based at Teaching Block Room 7 (TB7). No audio/video recordings will be taken. Only your

full name, contact details and signature will be taken to record your consent. You may also be contacted for future research if needed.

What are the possible benefits and risks of taking part?

The information we will get from the study will support the ITU Rapporteur Group on Loudness Measurement on future revisions of the measurement algorithm, helping to improve current recommendations on the subject.

The main disadvantage to taking part in the study is some possibility of annoyance. Listening tests are structured in such a way as to allow for breaks so as to avoid listener fatigue. Sound reproduction levels are controlled so as to be within safe limits as recommended by UK legislation.

Will my taking part be kept confidential?

What is said in the interview and/or data collected is regarded strictly confidential and all data for analysis will be anonymised/pseudonymised. In reporting on the research findings, I will not reveal the names of any participants or the organisation where you work.

All project data related to the administration of the project, (e.g. consent form) will be held for at least 6 years and all research data for at least 10 years in accordance with University policy. Your personal data will be held and processed in the strictest confidence, and in accordance with current data protection regulations.

All information gathered will be held for long-term storage on University approved secure servers. Hard files will be kept on University approved secure premises. No identifiable data will be accessed by anyone other than me, members of the research team and authorised personnel from the University and regulators for monitoring purposes.

This study has been given a favourable ethical opinion by the xxxxx Research Ethics Committee

What will happen to the data and results of the study?

I will produce a final report summarising the main findings. I also plan to disseminate the research findings through publication and conferences. De-identified data will be deposited or submitted to an open source online research data repository at the end of the study. This data may be used for future research.

Who should I contact for further information?

If you have any questions or require more information about this study, please contact me using the following contact details:

l.dasilvapires@surrey.ac.uk
Institute of Sound Recording research group
James Joule Building, 3rd floor, room 8 (08 BC 03)
+44 (0)1483 683050



CONSENT FORM FOR PARTICIPANTS IN RESEARCH STUDIES

Please complete this form after you have read the Information Sheet and/or listened to an explanation about the research.

Title of Study: Directional Loudness

University of Surrey Ref.: 353003-352994-36010101

Thank you for considering taking part in this research. The person organising the research must explain the project to you before you agree to take part. If you have any questions arising from the Information Sheet or explanation already given to you, please ask the researcher before you decide whether to join in. You will be given a copy of this Consent Form to keep and refer to at any time.

By ticking/initialling each box you are consenting to this element of the study. It will be assumed that un-ticked/un-initialled boxes mean that you DO NOT consent to that part of the study and you may be deemed ineligible for the study.

- I confirm that I have read and understood the information sheet dated June 6, 2018 for the above study. I have had the opportunity to consider the information and asked questions which have been answered satisfactorily.
- I understand that my participation is voluntary and that I am free to withdraw at any time during the study without giving any reason and without being disadvantaged in any way. Furthermore, I understand that I will be able to withdraw my data up to one month after the test.
- I consent to the processing of my personal information for the purposes explained to me. I understand that such information will be handled in accordance with current data protection regulations.
- I understand that my information may be subject to review by responsible individuals from the University of Surrey and/or regulators for monitoring and audit purposes.
- I understand that confidentiality and anonymity will be maintained and the researcher will not identify me in any research output.
- I agree to be contacted in the future by University of Surrey researchers who would like to invite me to participate in follow up studies to this project, or in future studies of a similar nature.
- I agree that the research team may use my anonymised data for future research and understand that any use of identifiable data would be reviewed and approved by a research ethics committee. (In such cases, as with this project, data would not be identifiable in any report).
- I understand that I must not take part if I fall under the exclusion criteria as detailed in the information sheet and explained to me by the researcher.

Name of Participant:

Date: Signature: