Anais do 14º Congresso de Engenharia de Áudio da AES Brasil
20ª Convenção Nacional da AES Brasil

Proceedings of the 14th AES Brazil Conference
17-19 Maio/May, 2016

Pavilhão Amarelo . Expo Center Norte . São Paulo - SP - Brasil

AUDIO ENGINEERING SOCIETY

BRAZIL SECTION

Anais do 14º Congresso de Engenharia de Áudio da AES Brasil

Proceedings of the 14th AES Brazil Conference

17-19 Maio/*May*, 2016

Pavilhão Amarelo do Expo Center Norte – São Paulo – SP – Brasil

Copyright ©2016

AUDIO ENGINEERING SOCIETY
BRAZIL SECTION

C76a Congresso de Engenharia de Áudio, 14., São Paulo, 2016.
Anais do 14º Congresso de Engenharia de Áudio AES Brasil e 20ª Convenção Nacional AES Brasil
Editado por Paulo R. A. Marins
Rio de Janeiro: Sociedade de Engenharia de Áudio, 2016.

Disponível em: www.aesbrasil.org/congressos/anais/

1. Engenharia de áudio – Congressos. 2. Processamento de sinais – Congressos. 3. Computação musical – Congressos. 4. Acústica – Congressos. I. Paulo R. A. Marins. II. Convenção Nacional AES Brasil, 20., São Paulo, 2016. III. Sociedade de Engenharia de Áudio. IV. Título.

ISSN 2177-529X

Índices para Catálogo Sistemático

1. Engenharia de áudio – Congressos – 621.3893
2. Processamento de sinais – Congressos – 621.38043
3. Computação musical – Congressos – 789.9
4. Acústica – Congressos – 620.21

Os artigos publicados nestes anais foram editorados a partir dos originais finais entregues pelos autores, sem edições, correções ou considerações feitas pelo comitê técnico. A AES Brasil não se responsabiliza pelo conteúdo. Outros artigos podem ser adquiridos através da Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA, (*www.aes.org*). Informações sobre a seção Brasileira podem ser obtidas em (*www.aesbrasil.org*). Todos os direitos são reservados. Não é permitida a reprodução total ou parcial deste artigo sem autorização expressa da AES Brasil.

Impresso no Brasil. *Printed in Brazil.*

Editoração e arte / *Publishing and art*
Bárbara R. Rosa

totum
agenciatotum.com.br

PATROCÍNIO

CAPES

APOIO

USP

SBC Sociedade Brasileira de Computação

UFRJ Universidade Federal do Rio de Janeiro

UNICAMP

UFSC

SÃO JUDAS UNIVERSIDADE

UnB

SOBRAC 30 anos Sociedade Brasileira de Acústica

Universidade Federal de Santa Maria 1960

UERJ Universidade do Estado do Rio de Janeiro

REALIZAÇÃO

AUDIO AES

AUDIO ENGINEERING SOCIETY
BRAZIL SECTION

Sociedade de Engenharia de Áudio
AES – Audio Engineering Society – Brazil Section
Rua Carlos Machado 164, sala 305, Pólo Rio de Cine e Vídeo – Barra da Tijuca
Rio de Janeiro, Brasil – Cep. 22775-042 | e-mail: aesbrasil@aes.org | www.aesbrasil.org
telefone: +55(21) 2421-0112 | fax: +55(21)2421-0112

ADMINISTRAÇÃO

Presidente / *Chairman:*	Armando Vicente Baldassarra
Vice- Presidente / *Vice- Chairman:*	José Pereira Anselmo Júnior
Secretário / *Secretary:*	José Carlos Giner
Tesoureiro / *Treasurer:*	Joel Vieira de Brito
Coordenador Acadêmico / *Academic Coordinator:*	Júlio Lucchi
Conselheiros / *Counselors:*	Aldo Soares
	José Augusto Mannis
	Marcelo Claret
	Julio Lucchi
	João Americo Bezerra

Audio Engineering Society, Inc.
International headquarters
60 East 42nd St., Room 2520, New York, NY, 10165-2520, USA
e-mail: hq@aes.org | www.aes.org telephone: +1(212)661-8528 | fax: +1(212)661-7829

COMITÊ DE ORGANIZAÇÃO / *ORGANIZING COMMITTEE*

Coordenadores do Programa Técnico ***/ Program Committee Chairs***
José Augusto Mannis - (IA-UNICAMP / UFF)
Sidnei Noceti Filho (LINSE-EEL-CTC-UFSC)
Luiz Wagner Biscainho (COPPE/UFRJ)

Coordenador de Artigos ***/ Papers Chair***
Hugo T. de Carvalho (IME/UERJ & COPPE/UFRJ)

Coordenador Editorial ***/ Editorial Chair***
Paulo R. A. Marins (UnB)

Coordenador de infra-estrutura ***/ Infrastructure Chair***
Regis Rossi A. Faria (USP)

Coordenador Acadêmico / ***Academic Chai***r
Stephan Paul (UFSC)

Coordenador de Comunicação ***/ Communication Chair***
Márcio Brandão (UnB)
Júlio C. Lucchi (USJT-IMT-FEI-ITA)

Coordenador da Convenção ***/ Convention Chair***
José Carlos Giner (AES-Brasil)

COMITÊ DE PROGRAMA TÉCNICO (REVISORES)
TECHNICAL PROGRAM COMMITTEE (REVIEWERS)

Alexander Pasqual	Universidade Federal de Minas Gerais – UFMG
Amaro Lima	Centro Federal de Educação Tecnológica do Rio de Janeiro - CEFET-RJ
André Dalcastagnê	Instituto Federal de Santa Catarina – IF-SC
Carlos de Araújo	Universidade Federal da Paraíba - UFPB
Carlos Ynoguti	Instituto Nacional de Telecomunicações - Inatel
Dante Coaquira Begazo	Universidade de São Paulo - USP
Diego Haddad	Centro Federal de Educação Tecnológica do Rio de Janeiro – CEFET-RJ
Fernando Iazzetta	Universidade de São Paulo - USP
Flávio Ávila	Universidade do Estado do Rio de Janeiro – UERJ
Hugo Carvalho	Universidade Federal do Rio de Janeiro – UFRJ
Ignacio Irigaray	Universidad de la Republica
Isabela Apolinário	Universidade Federal do Rio de Janeiro – UFRJ
Ivandro Sanches	Departamento de Engenharia Elétrica do Centro Universitário da FEI
João P.S. Monnazzi	Universidade de São Paulo - USP
José A. Mannis	Univ. Estadual de Campinas – UNICAMP & Univ. Federal Fluminense - UFF
Julio Lucchi	Universidade Sao Judas Tadeu - USJT
Leonardo Nunes	Microsoft
Lucas Maia	Universidade Federal do Rio de Janeiro – UFRJ
Luiz Biscainho U	Universidade Federal do Rio de Janeiro – UFRJ
Marcelo Queiroz	Universidade de São Paulo – USP
Marcio Brandao	Universidade de Brasilia - UNB
Marcus Bittencourt	Universidade Estadual de Maringá - UEM
Mario Minami	Universidade Federal do ABC
Martín Rocamora	Universidad de la República
Maurício da Costa	Universidade Federal do Rio de Janeiro – UFRJ /COPPE
Micael Antunes da Silva	Universidade de São Paulo - USP
Miguel Ramírez	Universidade de São Paulo – USP
Mikhail Malt	IRCAM & IReMus – UMR8223 – Paris – Sorbonne
Paulo Marins	Universidade de Brasília - UNB
Phillip Burt	Universidade de São Paulo - USP
Rafael Mendes	Universidade Estadual de Campinas - Unicamp - FEEC
Regis Faria	Universidade de São Paulo - USP
Stephan Paul	Universidade Federal de Santa Catarina - UFSC
Thiago Prego	Centro Federal de Educação Tecnológica Celso Suckow da Fonseca
Walter Gontijo	Universidade Federal de Santa Catarina

EDITORIAL

O Congresso de Engenharia de Áudio da AES Brasil chega à sua 14a edição. Ao longo de mais de uma década de existência, acima de 200 (duzentos) artigos científicos inéditos foram publicados, o que faz com que o evento possa ser considerado um fórum consolidado de discussões científicas acerca das pesquisas na área de engenharia de áudio no Brasil.

Nesta 14a edição estão sendo publicados 12 (doze) trabalhos inéditos, sendo 8 (oito) artigos completos (full papers) e 4 (quatro) artigos convidados (invited papers) de palestrantes. Os temas desses trabalhos estão relacionadas aos seguintes tópicos: computação musical, processamento digital de sinais de áudio, tecnologia musical, sistemas eletroacústicos, psicoacústica, análise e síntese de voz e música, produção musical, dentre outros. Esta diversidade de temas abordados no artigos científicos reforça a capacidade deste fórum no que tange à produção de discussões acerca de pesquisas científicas na área engenharia de áudio.

É importante destacar também as 4 (quatro) palestras apresentadas no Congresso: "Implementações de Efeitos de Áudio utilizando Arduino DUE e PedalSHIELD" por Walter Antonio Gontijo, Andrew Henrique Pavei e Sidnei Noceti Filho, da Universidade Federal de Santa Catarina; "Um relato atual da prática da bioacústica de mamíferos no Brasil" por Patrícia Monticelli, Bruna Campos Paula e Regis Rossi A. Faria, da Universidade de São Paulo; "Qual o futuro do MP3? Áudio espacial e codificação orientada a objetos", por Bruno Masiero, da Universidade Estadual de Campinas, e "Análise modal e representação musical aplicadas ao design de estúdios de produção fonográfica" por José Augusto Mannis das Universidades Estadual de Campinas e Federal Fluminense e Tiago F. Mattos, do Estúdio Codimuc.

Outro fato que é importante salientar é a concomitância da realização do 14o Congresso de Engenharia de Áudio da AES Brasil com a 20a Convenção Nacional da AES Brasil. O fato de os dois eventos acontecerem no mesmo espaço e simultaneamente possibilita aos participantes não só estar presentes em discussões científicas, como também a participar de outras atividades, tais como: assistir a palestras proferidas por profissionais de engenharia de áudio, presenciar demonstrações de fabricantes, ou mesmo tomar conhecimento acerca das novidades tecnológicas no que tange a equipamentos de áudio.

Por fim, gostaríamos de ressaltar a crescente relevância e repercussão do evento, visto que o número de participantes e envolvidos no congresso vem aumentando a cada edição, o que denota um significante aumento do interesse por parte da comunidade acadêmica e dos profissionais da área de áudio no que concerne aos temas abordados no 14o Congresso de Engenharia de Áudio da AES Brasil .

Paulo R. A. Marins - Coordenador Editorial - Universidade de Brasília - UnB

PROGRAMAÇÃO
Terça Feira / Thuesday, Maio / May 17, 2016

10:30AM
12:30PM
Auditório D
Room D

Sessão de Artigos 1/ *Paper Session 1*
ARTIGOS COMPLETOS / *FULL PAPERS*
Chair: Hugo Carvalho

PROGRAMAÇÃO LÓGICA INDUTIVA APLICADA À COMPUTAÇÃO MUSICAL: UMA REVISÃO SISTEMÁTICA
Cienio Gonçalves Jr. / Murilo Homem
UFSCAR - Universidade Federal de São Carlos - Sorocaba, Brasil

AN IMPROVED METHOD FOR DETERMINING PITCH RELEVANCE IN POLYPHONIC AUDIO
Isabela Apolinário / Luiz Biscainho
UFRJ - Universidade Federal do Rio de Janeiro, Brasil

AJUSTE DE MODELOS ANALÍTICOS DE MICROFONE PIEZOELÉTRICO A PARTIR DE MODELOS NUMÉRICOS
Bernardo Henrique Murta / Eric Brandão / Paulo Mareze / Willian Fonseca
UFSM - Universidade Federal de Santa Maria, Brasil
Homero Sette Silva
Pesquisador Independente, Brasil
Júlio Apolinário Cordiolli
UFSC - Universidade Federal de Santa Catarina, Brasil

GUITAR AMI: DESENVOLVIMENTO, IMPLEMENTAÇÃO E PERFORMANCE DE UM INSTRUMENTO MUSICAL AUMENTADO
Eduardo Meneses / José Fornari
UNICAMP - Universidade Estadual de Campinas, Brasil

PROGRAMAÇÃO
Quarta-Feira/ Wednesday Maio / May 18, 2016

9:30AM
11:30PM
Auditório D
Room D

Sessão de Artigos 2 / *Paper Session 2*
ARTIGOS COMPLETOS / *FULL PAPERS*
Chair: Luiz Biscainho

CONTROLE AUTOMÁTICO DE LOUDNESS EM CONTEÚDO DE FORMATO CURTO PARA RADIOFUSÃO
Leandro Pires / Maurílio Vieira / Hani Yehia
UFMG - Universidade Federal de Minas Gerais, Brasil

SÍNTESE DE CAMPOS SONOROS EM PAINÉIS RÍGIDOS ATRAVÉS DE UM ARRANJO DE FONTES ELEMENTARES
Clinton Merlo
UNIFEI - Universidade Federal de Itajubá, Brasil
Alexander Pasqual / Eduardo Bauzer Medeiros
UFMG - Universidade Federal de Minas Gerais, Brasil

IDENTIFICAÇÃO DE DOENÇAS DE LARINGE A PARTIR DE SINAIS DE VOCALIZAÇÃO E REDES NEURAIS ARTIFCIAIS
Bruno Sodré / Marcelo Rosa /Marco da Silva
UTFPR - Universidade Tecnológica Federal do Paraná, Brasil
Ana Dassie Leite
UNICENTRO - Universidade Estadual do Centro-Oeste, Brasil

EXTRAÇÃO DE DESCRITORES SONOROS A PARTIR DA SÍNTESE ADITIVA FRACTAL
Thiago Roque / Rafael Mendes
UNICAMP - FEEC - Universidade Estadual de Campinas, Brasil

11:30 am
12:40 pm
Auditório D
Room D

Palestra / Lecture

IMPLEMENTAÇÕES DE EFEITOS DE ÁUDIO UTILIZANDO ARDUINO DUE E PEDALSHIELD
Sidnei Noceti Filho / Walter Gontijo / Andrew Pavei
UFSC - Universidade Federal de Santa Catarina, Brasil

PROGRAMAÇÃO
Quinta-feira/ Thursday, Maio / May 19, 2016

09:30 am
10:40 am
Auditório D
Room D

Palestra / Lecture

UM RELATO ATUAL DA PRÁTICA DA BIOACÚSTICA DE MAMÍFEROS NO BRASIL
Regis Rossi Alves Faria
Patrícia Monticelli / Bruna Paula
FFCLRP USP- Universidade de São Paulo, Brasil - Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Brasil

10:40 am
11:50 am
Auditório D
Room D

Palestra / Lecture

QUAL O FUTURO DO MP3? ÁUDIO ESPACIAL E CODIFICAÇAO ORIENTADA A OBJETOS
Bruno Masiero
UNICAMP - Universidade Estadual de Campinas

11:50am
13:00pm
Auditório D
Room D

Palestra / Lecture

ANÁLISE MODAL E REPRESENTAÇÃO MUSICAL APLICADAS AO DESIGN DE ESTÚDIOS DE PRODUÇÃO FONOGRÁFICA
José Augusto Mannis
UNICAMP - Universidade Estadual de Campinas, Brasil & UFF - Universidade Federal Fluminense
Tiago F.Mattos
Estúdio Codimuc

SUMÁRIO / *CONTENTS*

COMITÊ DE ORGANIZAÇÃO / *ORGANIZING COMMITTEE* IV

COMITÊ DE PROGRAMA TÉCNICO (REVISORES) / *TECHNICAL PROGRAM COMMITTEE (REVIEWERS)* V

EDITORIAL VI

PROGRAMAÇÃO / *SCHEDULE* VII

ARTIGOS COMPLETOS / *FULL PAPERS*

PROGRAMAÇÃO LÓGICA INDUTIVA APLICADA À COMPUTAÇÃO MUSICAL: UMA REVISÃO SISTEMÁTICA
Cienio Gonçalves Jr. / Murilo Homem 14

SÍNTESE DE CAMPOS SONOROS EM PAINÉIS RÍGIDOS ATRAVÉS DE UM ARRANJO DE FONTES ELEMENTARES
Alexander Pasqual / Eduardo Bauzer Medeiros / Clinton Merlo 23

AJUSTE DE MODELOS ANALÍTICOS DE MICROFONE PIEZOELÉTRICO A PARTIRDE MODELOS NUMÉRICOS
Bernardo Henrique Murta / Eric Brandão / Homero Sette Silva / Júlio Apolinário Cordiolli 33

GUITARAMI: DESENVOLVIMENTO, IMPLEMENTAÇÃO E PERFORMANCE DE UM INSTRUMENTO MUSICAL AUMENTADO
Eduardo Meneses / José Fornari 39

CONTROLE AUTOMÁTICO DE LOUDNESS EM CONTEÚDO DE FORMATO CURTO PARA RADIOFUSÃO
Leandro Pires / Maurílio Vieira / Hani Yehia 46

IDENTIFICAÇÃO DE DOENÇAS DE LARINGE A PARTIR DE SINAIS DE VOCALIZAÇÃO E REDES NEURAIS ARTIFCIAIS
Bruno Sodré / Marcelo Rosa /Marco da Silva /Ana Dassie Leite 56

EXTRAÇÃO DE DESCRITORES SONOROS A PARTIR DA SÍNTESE ADITIVA FRACTAL
Thiago Roque / Rafael Mendes 64

AN IMPROVED METHOD FOR DETERMINING PITCH RELEVANCE IN POLYPHONIC AUDIO
Isabela Apolinário / Luiz Biscainho 72

PALESTRAS / *LECTURES*

IMPLEMENTAÇÕES DE EFEITOS DE ÁUDIO UTILIZANDO ARDUINO DUE E PEDALSHIELD
Sidnei Noceti Filho / Walter Gontijo / Andrew Pavei 81

UM RELATO ATUAL DA PRÁTICA DA BIOACÚSTICA DE MAMÍFEROS NO BRASIL
Regis Rossi Alves Faria / Patrícia Monticelli / Bruna Paula 90

QUAL O FUTURO DO MP3? ÁUDIO ESPACIAL E CODIFICAÇAO ORIENTADA A OBJETOS
Bruno Masiero 96

ANÁLISE MODAL E REPRESENTAÇÃO MUSICAL APLICADAS AO DESIGN DE ESTÚDIOS DE PRODUÇÃO FONOGRÁFICA
José Augusto Mannis / Tiago F. Mattos 106

ÍNDICE DE AUTORES / AUTHORS INDEX 118

ARTIGOS COMPLETOS / FULL PAPERS

Sociedade de Engenharia de Áudio
Artigo de Congresso

Apresentado no 14º Congresso de Engenharia de Áudio
20ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2016, São Paulo, SP

Este artigo foi reproduzido do original final entregue pelo autor, sem edições, correções ou considerações feitas pelo comitê técnico. A AES Brasil não se responsabiliza pelo conteúdo. Outros artigos podem ser adquiridos através da Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA, www.aes.org. Informações sobre a seção Brasileira podem ser obtidas em www.aesbrasil.org. Todos os direitos são reservados. Não é permitida a reprodução total ou parcial deste artigo sem autorização expressa da AES Brasil.

Programação Lógica Indutiva aplicada à Computação Musical: uma revisão sistemática

Clenio B. Gonçalves Junior[1,2] e Murillo R. P. Homem[1,3]

[1] Universidade Federal de São Carlos - Campus Sorocaba, Programa de Pós-Graduação em Ciência da Computação de Sorocaba, Departamento de Computação (DComp) Sorocaba
Sorocaba, São Paulo, 18052-780, Brasil

[2] Instituto Federal de Educação, Ciência e Tecnologia de São Paulo - Campus São Roque
São Roque, São Paulo, 18136-540, Brasil

[3] Universidade Federal de São Carlos, Departamento de Computação
São Carlos, São Paulo, 13565-905, Brasil

clenio@ifsp.edu.br, murillo@ufscar.br

RESUMO

Em Computação Musical, o processo de representação de conhecimento constitui um elemento essencial para o desenvolvimento de sistemas. Métodos têm sido aplicados visando fornecer ao computador a capacidade de gerar conclusões a partir da experiência e definições previamente estabelecidas. Neste sentido, a Programação Lógica Indutiva apresenta-se como um campo de pesquisa que incorpora conceitos de Programação em Lógica e Aprendizado de Máquina, seu caráter declarativo possibilita que o conhecimento musical seja apresentado a usuários não especialistas de modo naturalmente compreensível. O presente trabalho realiza uma revisão sistemática baseada em abordagens que utilizam a Programação Lógica Indutiva na representação de conhecimento musical. Foram levantadas questões que esses trabalhos procuram atender, bem como identificados aspectos característicos ligados à sua aplicação.

0 INTRODUÇÃO

O campo da Computação Musical tem se destacado ao longo do tempo, entre outros fatores, por sua característica multidisciplinar [1]. Estudos realizados nessa área têm fornecido auxílio na compreensão de conceitos envolvendo interação humano-computador, paradigmas de programação, processos cognitivos, etc. A área de Inteligência Artificial, particularmente, con-

centra grande variedade de trabalhos que mantêm estreita relação e cuja afinidade remonta várias décadas atrás [2, 3, 4, 5, 6].

Intimamente ligado à abordagem simbólica em inteligência artificial, existe o conceito de representação de conhecimento. Este campo de pesquisa trata elementos de modelagem da informação, de forma que o computador tenha a capacidade de gerar conclusões lógicas de maneira eficiente, possibilitando a execução de atividades complexas que envolvam elementos como raciocínio e criatividade. A definição de uma ferramenta adequada para representação de conhecimento é fundamental para que o desenvolvimento de um sistema de inteligência artificial seja bem sucedido [7].

Sendo assim, a modelagem da estrutura de representação de conhecimento é de suma importância em atividades de computação musical [8], sendo que os processos envolvendo tal atividade devem compor a primeira fase no desenvolvimento de um sistema [9]. De forma mais específica, a modelagem declarativa, por meio da Lógica de Primeira Ordem, constitui um importante método. Tanto sua estrutura descritiva como os mecanismos de aplicação levantam questões particularmente interessantes [10]. Esse modelo apresenta um formalismo natural para a definição de componentes musicais, possibilitando que inferências sejam realizadas sobre uma base musical previamente definida, além de propiciar a descoberta de novos padrões estruturais.

Nesse contexto, a Programação Lógica Indutiva (PLI) apresenta-se como um campo de pesquisa crescente que combina os conceitos de Programação em Lógica e Aprendizado de Máquina [11]. PLI baseia-se na lógica de primeira ordem, o que lhe confere um caráter declarativo, isso permite que o resultado do processamento gerado por um sistema possa ser apresentado a usuários não especialistas de uma forma simples e intuitiva [12]. Por meio da programação lógica indutiva, novo conhecimento musical pode ser gerado automaticamente a partir da derivação de estruturas e regras expressas na forma de cláusulas de Horn. Uma série de modelos têm sido propostos em trabalhos envolvendo conceitos como contraponto, expressividade performática, representação harmônica, inferência modal, entre outros. Adicionalmente, a PLI tem sido aplicada como um eficiente recurso para representação de conhecimento no que tange à caracterização de gêneros e estilos musicais [13, 14, 15, 12, 16]. Variados sistemas que implementam indução lógica têm sido utilizados em aplicações musicais específicas, tais como Aleph [17], TILDE [18] e PAL [19]. Tais sistemas realizam o processo indutivo a partir de linguagens como Prolog. Além disso, teorias cognitivas têm sido aplicadas na análise musical, como Narmour [20], visando a compreensão de estruturas melódicas.

Desse modo, destaca-se a importância da realização de um levantamento considerando abordagens que apliquem a PLI no processo de representação de conhecimento musical. Esse levantamento deve investigar critérios e práticas adotados, bem como a utilização de recursos computacionais, além de conceitos envolvidos no campo da Musicologia. Tal atividade deve fornecer subsídios, baseados em evidências, que levem ao delineamento de diretrizes na elaboração de trabalhos relacionados [21].

O presente trabalho realiza uma uma extensão de [22], apresentando uma revisão sistemática da literatura a partir de abordagens que aplicam a Programação Lógica Indutiva no processo de representação de conhecimento musical. O restante do artigo é organizado da seguinte forma: a Seção 1 descreve a metodologia utilizada. A Seção 2 apresenta e analisa os resultados relacionados às questões de pesquisa. A Seção 3 finaliza o artigo, descrevendo as principais conclusões obtidas.

1 METODOLOGIA

Com o intuito de realizar um levantamento sobre trabalhos que abordem a representação de conhecimento musical por meio da Programação Lógica Indutiva, foi elaborada uma revisão sistemática de literatura. A metodologia de pesquisa foi aplicada de acordo com critérios e recomendações apresentados em [23]. Foram levantadas duas questões para servir de diretriz ao trabalho:

Q1: Qual arquitetura é proposta para representação do conhecimento musical?

Q2: Como a abordagem aplica os recursos computacionais ligados à PLI?

Deste modo, buscou-se encontrar abordagens que possibilitem o estudo da técnica em questão. A seguir, são descritas as etapas realizadas para a seleção dos trabalhos.

1.1 ESTRATÉGIA DE PESQUISA

A pesquisa foi realizada a partir de publicações de trabalhos acadêmicos, artigos em periódicos científicos e participações em conferências. As seguintes bases de consulta foram utilizadas:

- Academic Search Premier
- ACM Digital Library
- Cambridge Journals Online
- Computers and Applied Sciences
- Oxford Journals
- Computer and Information Systems
- SpringerLink
- Web of Science
- ScienceDirect
- SCOPUS
- IEEE Xplore
- Google Scholar

Com relação à estratégia de pesquisa, o trabalho foi baseada nas línguas Inglesa e Portuguesa. Foram elencados os principais termos de acordo com as questões levantadas. As atividades foram realizadas nos mesmos moldes das abordagens apresentadas em [24], [25], [26] e [27]. Apenas trabalhos realizados a partir de 1990 foram considerados, devido ao fato de que nesse ano foi introduzido o formalismo da Lógica Indutiva de Primeira Ordem (FOIL) [28].

1.2 *Strings* de busca

Duas expressões iniciais foram tomadas como base para a busca, enfatizando o elemento musi-

cal e a programação lógica indutiva. Ambas foram aplicadas em conjunção a componentes voltados à representação de conhecimento. Em seguida foram incluídas expressões literais equivalentes e adicionadas derivações de pronúncia. Os elementos fundamentais para a seleção foram unidos pelo operador ”AND”e as variações alternativas foram conectadas pelo operador ”OR”. Deste modo, considerando as duas línguas em questão, as seguintes *strings* de busca foram elaboradas:

1: `("knowledge representation" AND musi*) OR ("musi* representation" AND knowledge) OR ("musi* knowledge" AND representation) OR knowledge AND (representation AND "computer musi*")`

2: `("representacao de conhecimento" AND musi*) OR ("representacao musi*" AND conhecimento) OR ("conhecimento musi*" AND representacao) OR (conhecimento AND representacao AND "computacao musical")`

3: `"inductive logic programming" AND knowledge AND representation`

4: `"programação lógica indutiva" AND conhecimento AND representação`

1.3 Seleção de Trabalhos

Para a seleção dos trabalhos, foram aplicados filtros com o intuito de destacar os estudos com maior importância de acordo com a finalidade proposta. Para um adequado refinamento dos tópicos de relevância, a busca foi direcionada de modo mais específico aos seguintes assuntos:

> Artificial intelligence, Expert systems, Automatic composer, Algorithmic composition, Music representation, Logic programming, Answer set programming, Prolog, Logical language, Automatic composition, Constraint programming, Declarative languages, Declarative system, Knowledge-based system.

A Figura 1 apresenta os critérios utilizados na elaboração dos filtros, bem como a quantidade de trabalhos selecionados em cada etapa.

A partir de uma quantidade inicial de 7075 trabalhos, chegou-se ao número de 9 trabalhos a serem estudados. A desigualdade entre esses valores é explicada em [23]. A Tabela 1 apresenta as abordagens, cada uma recebeu uma identificação numérica e um nome formado pelas duas primeiras palavras do título.

2 RESULTADOS E ANÁLISE

Dando continuidade aos processos de pesquisa e filtragem de trabalhos, as 9 abordagens selecionadas foram analisadas do ponto de vista dos componentes relevantes à revisão. Foram identificados elementos fundamentais bem como aspectos característicos de acordo com as questões básicas. A seguir apresenta-se uma discussão sobre as abordagens feitas com relação a tais questões. Considerando o inter-relacionamento entre os componentes levantados, alguns aspectos serão analisado juntamento com sua descrição.

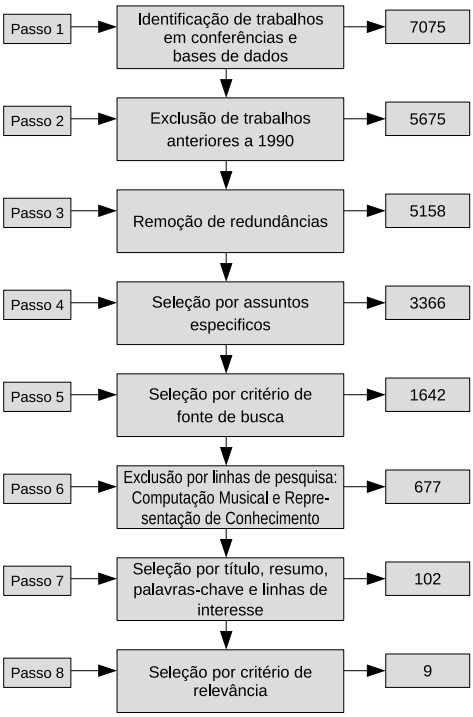


Figura 1: Etapas na estratégia de busca

2.1 Qual arquitetura é proposta para representação do conhecimento musical?

Ao se tratar os elementos relacionados à representação musical por meio do computador, componentes como notação simbólica, formatos de arquivo, recursos para codificação, técnicas de programação e abstração de dados têm sido aplicados no nível do processamento da informação. Adicionalmente, o trabalho envolvendo representação de conhecimento musical aborda aspectos referentes a paradigmas de programação e componentes ligados a lógicas formais, além de métodos de busca e tomada de decisão eficientes.

SymbolicRepresentation [29] utiliza um sistema baseado em regras para incorporar o conhecimento aplicado no processo de interpretação da informação musical. As regras são formuladas a partir de considerações como deslocamento da fundamental, movimentação do baixo e fatores ligados à condução da voz principal. Visando modelar o movimento do campo harmônico, a estrutura gramatical é representada por regras da teoria musical ocidental vigente. Em contraste aos algoritmos tradicionais do tipo “se-então-senão” é adotada uma estrutura de árvore de decisão, a qual possibilita resultados mais eficientes. É feita a implementação de duas categorias principais de regras. A primeira define restrições com o objetivo de direcionar o foco do processo de inferência para conjuntos específicos de progressões harmônicas. Assim, se houver a necessidade de fazer distinção entre acordes cromáticos e

Tabela 1: Relação de trabalhos revisados

ID	ABORDAGEM	TÍTULO	REFERÊNCIA
1	SymbolicRepresentation	Symbolic Representation of Chords for Rule-Based Evaluation of Tonal Progressions	Chong and Ding, 2014
2	ImprovingMusic	Improving Music Genre Classification Using Automatically Induced Harmony Rules	Anglade et al., 2010
3	ProbabilisticLogic	Probabilistic and Logic-Based Modelling of Harmony	Dixon et al., 2011
4	ApplicationILP	Application of ILP in a musical database: learning to compose the two-voice counterpoint	Pompe et al., 1996
5	LearningMusical	Learning Musical Rules	Morales and Morales, 1995
6	ExpressiveConcatenative	Expressive Concatenative Synthesis by Reusing Samples from Real Performances Recordings	Maestre et al., 2009
7	ConstructiveAdaptive	Constructive Adaptive User Interfaces - Composing Music Based on Human Feelings	Numao et al., 2002
8	ModelingMoods	Modeling Moods in Violin Performances	Perez et al., 2008
9	ModelingExpressive	Modeling Expressive Music Performance in Jazz	Ramirez and Hazan, 2005

diatônicos, pode-se criar uma regra em que todos os acordes de uma progressão devam pertencer à mesma clave. A segunda categoria realiza considerações harmônicas sobre a condução da voz principal. Neste caso, um conjunto de regras pode ser implementado para garantir o tratamento apropriado para resoluções de acordes cadenciais.

A abordagem de ImprovingMusic [15] é voltada para a identificação de estilos e gêneros musicas. Progressões harmônicas são tratadas utilizando a descrição sequencial de acordes. Realiza-se a combinação de descritores de áudio de baixo nível com uma ferramenta de aprendizado de máquina, o que possibilita a classificação baseada em cadências. As sequências servem como base para o treinamento e são obtidas de modo indutivo a partir de transcrições automáticas. Desse modo, pode-se obter um desempenho satisfatório na classificação de gêneros musicais. Um conjunto de atributos é utilizado na definição dos parâmetros representativos. Esse conjunto envolve elementos como tempo e dinâmica, além do espectro sonoro e descritores de altura. Trechos musicais são representados por listas que incorporam os acordes neles contidos. Gêneros são representados por conjuntos de trechos. Para cada gênero, busca-se encontrar regras harmônicas que descrevam sequências características em acordes presentes nas músicas armazenadas. Estas regras harmônicas constituem uma Gramática Livre de Contexto. Apenas as características harmônicas pertencentes aos gêneros conhecidos são consideradas neste modelo. Dessa forma, acordes com finalidades puramente ornamentais e sequências não características aos estilos mantidos são desconsiderados. Nesse formalismo, os acordes são representados por letras do alfabeto e rotulados usando a notação de cifras. As propriedades dos acordes são descritas por meio de predicados lógicos que retornam valores verdadeiros ou falsos.

Integrando-se ao mecanismo indutivo, alguns elementos estruturais foram utilizados nas abordagens. Tais componentes foram aplicados de modo a compor o modelo de representação de conhecimento específico. A Tabela 2 apresenta esses elementos.

ProbabilisticLogic [12] apresenta duas abordagens voltadas à modelagem harmônica: probabilística e lógica. Com isso pode-se obter tanto a categoria de conhecimento musical, como também o raciocínio utilizado por um músico ao realizar tarefa análoga. Acordes são transcritos a partir de gravações de áudio. O sistema realiza a modelagem de contexto musical em alto nível. Nesse modelo, elementos como acordes, claves, posicionamento métrico, nota de baixo, características cromáticas e estruturas de repetição são integrados em um sistema Bayasiano. Esse sistema produz o conteúdo de uma pauta-mestre que contém uma sequência simbólica representando acordes. Cada símbolo inclui variações de clave e modulações no transcorrer do tempo. Além disso, o sistema concentra-se na descrição lógica de sequências harmônicas com o intuito de caracterizar estilos musicais e gêneros particulares. As relações harmônicas são representadas por meio do formalismo da lógica de primeira ordem. A abordagem baseia-se em representações por árvores de decisão, utilizadas para classificar amostras não visitadas ou fornecer sugestões com respeito a características variadas da base de dados.

ApplicationILP [13] desenvolveu uma ferramenta com a finalidade de realizar o processo de descoberta de conhecimento em uma base de dados com dezenas de milhares de instâncias. O sistema foi implementado de modo que a indução de uma hipótese é tratada como um problema de otimização. Tal abordagem é direcionada ao tratamento da problemática envolvendo a

Tabela 2: Elementos estruturais na representação de conhecimento

	1	2	3	4	5	6	7	8	9
Utiliza o padrão MIDI			■			■	■		■
Realiza coleta analógica de dados								■	■
Realiza pré-processamento		■							■
Realiza síntese aditiva						■			
Estrutura básica: melodia						■			■
Estrutura básica: contraponto				■	■				
Estrutura básica: harmonia	■	■	■						
Técnica integrada: Algoritmos genéticos							■		
Técnica integrada: Mineração de dados									■
Técnica integrada: Processo estocástico			■						
Técnica integrada: Restrições lógicas	■				■				
Representação de textura polifônica						■			

composição do contraponto a duas vozes. A aplicação de técnicas de aprendizado de máquina mostrou-se adequada a esse subdomínio da música tonal que possui características imprecisas, envolvendo aspectos não convencionais e exigentes do ponto de vista da acurácia. Dois predicados alvo foram utilizados para representar o conhecimento prévio a partir do universo de constantes musicais. Adicionalmente, foi realizado o treinamento tanto das instâncias positivas como das negativas, com o propósito de aplicar a indução por meio de regras.

LearningMusical [11] utiliza a lógica de primeira ordem para expressar regras de contraponto, possibilitando que relações entre estados musicas sejam descritas de uma maneira compacta e compreensível. O sistema executa a análise a partir de um conjunto finito de regras bem estabelecidas. O aprendizado de regras é implementado por meio do sistema PAL [19]. Este ambiente é alimentado por um subconjunto de cláusulas de Horn. Essas cláusulas contém conhecimento armazenado a partir de duas origens: amostras e informações prévias, sendo em ambos os casos expressas de modo declarativo. PAL incorpora regras de contraponto e as utiliza na análise musical e geração de notas de acordo com as convenções do *cantus firmus*. Desse modo, uma sequência de notas individuais é utilizada como base para a aplicação das regras de contraponto. Com isso, novas notas são geradas de acordo com as restrições harmônicas estabelecidas. Por meio do sistema PAL, as restrições podem ser utilizadas como diretrizes na busca por hipóteses, tanto no contexto imediato, como também de modo mais extensivo.

O modelo de ExpressiveConcatenative [30] realiza a síntese expressiva por meio da aquisição de conhecimento obtido em gravações de áudio. A partir destas gravações, amostras de notas pontuais são criteriosamente concatenadas, formando melodias com um corpo direcionado à sua execução. O sistema busca produzir uma sequência de áudio a partir da análise de partituras. Para isso, utiliza um modelo performático indutivo, o qual é integrado a uma base de dados gerada a partir de execuções reais. A abordagem aplica a ideia de utilizar o mesmo banco de dados nas principais etapas do processo. Desse modo, tanto na introdução do modelo de performance como na concatenação sonora das amostras, a mesma base de conhecimento é manipulada. Assim, fica estabelecida uma conexão durante o processo de síntese entre o som do instrumento em execução e as características modeladas a partir desta execução.

ConstructiveAdaptive [14] lida com estruturas musicais capazes de causar sentimentos humanos característicos. Foi construído um sistema para arranjo e composição de modo automático, que gera peças musicais com a capacidade de produzir sentimentos específicos em uma pessoa. Inicialmente, o sistema realiza a coleta de sentimentos a partir de obras musicais. Baseando-se em tal informação, são extraídas estruturas musicais que servirão de base para a produção dos sentimentos. Com relação à geração musical, o sistema é direcionado para duas possibilidades: elaborar arranjos em músicas pré-existentes e criar novas composições de maneira automática. Em ambos os casos, o resultado produzido tem a capacidade de produzir em um indivíduo sentimentos previamente determinados.

A abordagem de ModelingMoods [16] é voltada ao aspecto instrumental de obras executadas no violino. O sistema armazena padrões expressivos que são adquiridos de modo automático. Realiza a modelagem de aspectos interpretativos como variações respectivas ao tempo, dinâmica e altura. Além disso, o modelo recebe informações relativas ao controle gestual, com ênfase no que diz respeito à direção do arco e posicionamento dos dedos. O sistema baseia-se em quatro tipos de humores performáticos: Tristeza, Felicidade, Medo e Raiva. As características expressivas analisadas são duas: o tempo e um conjunto de descritores em nível de nota. Estes descritores compreendem qualidades ligadas ao ataque, duração, energia, direção do arco

e corda sendo tangida. Tanto as sequências de notas como as variações de expressividade na execução são definidas de um modo estruturado por meio de predicados da lógica de primeira ordem. Há dois grupos de predicados para descrever o contexto musical de cada nota e as variações expressivas. Os predicados para contexto musical definem informações sobre propriedades intrínsecas das notas, incluindo duração e posição métrica. Também armazenam suas notas predecessora e subsequente, além da extensão e direção dos intervalos. Os predicados de variações expressivas definem o fator de extensão de uma nota no que diz respeito à sua duração na pauta. Adicionalmente, possuem a capacidade de incorporar o ponto de mudança na direção do arco, a corda em que a nota é tocada e o nível de energia na execução.

Em ModelingExpressive [8] a abordagem tem o intuito de investigar a performance musical em melodias de *Jazz*. Utiliza técnicas de aprendizado de máquina para extrair movimentos regulares e padrões de desempenho. Para tanto, são aplicados métodos de classificação e regressão em bases de dados obtidas a partir de execuções *jazzísticas* reais. Considerando os métodos de regressão, busca-se a precisão preditiva para gerar boas soluções do ponto de vista da acurácia. No caso da classificação, sua utilização tem o objetivo de possibilitar uma descrição compreensível com relação às predições do sistema. Os modelos baseados em regressão são introduzidos com o objetivo de implementar transformações que possibilitem performances características. Os modelos de classificação são utilizados para possibilitar a compreensão dos princípios e critérios no que diz respeito à execução dos trechos musicais.

2.2 Como a abordagem aplica os recursos computacionais ligados à PLI?

Os recursos computacionais utilizados envolvem mecanismos para indução lógica, ferramentas de pré-processamento, linguagens de programação, conceituação teórica para percepção e cognição melódica. Inicialmente, o mecanismo de inferência lógica TILDE (*Top-down Induction of Logical Decision Trees*) [18] fundamenta-se na lógica de primeira ordem e aplica a indução por meio de árvores de decisão. Sua funcionalidade é considerada uma extensão ao algoritmo C4.5 [31], que ao invés de testar valores de atributos em nós de uma árvore, testa predicados lógicos [30]. Nesse modelo, cada árvore trata um problema de classificação específico. Em ImprovingMusic, regras descrevendo padrões harmônicos de um dado gênero podem coexistir com regras de outros gêneros em uma mesma árvore. ProbabilisticLogic realiza a descrição de acordes em termos de sua nota fundamental, grau na escala e categorias de intervalos entre notas sucessivas. A utilização de TILDE em ExpressiveConcatenative possibilita vantagens no sentido de se obter árvores proposicionais de decisão com eficiência e técnicas de poda, além do uso da lógica de primeira ordem e a possibilidade de incluir conhecimento prévio no processo de aprendizagem. ModelingMoods realiza a mineração de dados estruturados por meio da abordagem indutiva, aplicando o algoritmo *top-down* em árvores de decisão. Os recursos e ferramentas computacionais utilizados em cada abordagem são apresentados na Tabela 3.

O sistema de programação lógica indutiva Aleph (*A Learning Engine for Proposing Hypotheses*) [17] foi desenvolvido com o propósito de explorar ideias expressas por cláusulas de Horn com alta capacidade representativa. Escrito em Prolog, possibilita a descrição de expressões complexas, incorporando simultaneamente o conhecimento adquirido, com a capacidade de escolha da ordem de geração das regras, alteração nas funções de avaliação e busca [32]. Em ProbabilisticLogic, esses conceitos são aplicados para se encontrar um conjunto mínimo de regras que seja capaz de descrever a totalidade das amostras positivas e um número mínimo de amostras negativas. Essas regras cobrem todas as sequências de 4 acordes, sendo que cada sequência é tratada uma única vez. ModelingExpressive utiliza seu algoritmo padrão para manipulação em conjuntos com a finalidade de construir hipóteses individuais. Deste modo, Aleph utiliza a primeira amostra positiva não visitada como semente para a busca e cobre o espaço de cláusulas de acordo com uma restrição de comprimento previamente definida.

Narmour é uma teoria para percepção e cognição de melodias aplicada à análise musical. Auxilia na compreensão tanto do significado melódico como do conhecimento envolvido em sua criação [20]. ExpressiveConcatenative utiliza o modelo de implicação/realização dessa teoria, onde cada nota pertence a uma estrutura Narmour. ModelingMoods utiliza o contexto Narmour, onde são definidos grupos específicos com os quais uma nota mantenha uma relação de pertinência. ModelingExpressive faz uso extensivo deste conceito, incorporando informações sobre notas prévias e sucessoras, além de propriedades intrínsecas a intervalos.

ModelingExpressive utiliza SMSTools [33] para realizar o preprocessamento, criando uma descrição em alto nível de gravações em áudio. Também gera uma sonorização expressiva de acordo com transformações obtidas por aprendizado de máquina. Essa abordagem utiliza o algoritmo C4.5 [31] para obter um conjunto de regras de classificação diretamente da árvore de decisão gerada. O mecanismo de regras aplicado em SymbolicRepresentation baseia-se em JBoss Drools [34] que possibilita a operação de acordo com linhas de desenvolvimento pré-definidas. Esse mecanismo utiliza o algoritmo Rete [35] para realizar o emparelhamento de padrões. LearningMusical fundamenta-se no sistema PAL [19] para o aprendizado lógico. Esse sistema é utilizado no processo de análise de contraponto para executar o aprendizado de regras de transição. Tais regras

Tabela 3: Recursos computacionais ligados à PLI

	1	2	3	4	5	6	7	8	9
Aleph		■	■						■
C4.5									■
SFOIL				■					
Narmour						■		■	■
SMSTools									■
TILDE		■	■			■		■	
Pal					■				
JBoss Drools	■								
Indução a partir de Prolog		■	■		■		■	■	■
Utiliza WEKA									■
Utiliza base de treinamento adicional		■	■			■	■	■	■

são expressas na forma de cláusulas de Horn a partir de pares de estados musicais (conjuntos de notas), adicionalmente representando o conhecimento musical de propósito geral.

O sistema FOIL [28] é um passo inicial no sentido de se obter ferramentas eficientes para lidar com problemas do mundo real. É considerado um marco no campo da PLI [13, 36, 37]. ApplicationILP utiliza o algoritmo SFOIL (*Stochastic* FOIL), esse algoritmo combina a eficiência de FOIL com a implementação de uma estratégia de busca estocástica. Essa abordagem utiliza uma representação extensional de experiências, baseando-se em predicados alvo. Tal representação possibilitou o processo de descoberta de conhecimento em uma base de dados superior a 10.000 instâncias.

2.3 Características adicionais

Alguns fatores adicionais foram levantados a partir dos direcionamentos aplicados pelas abordagens. Com relação aos componentes de Musicologia tratados, destacam-se a caracterização de gênero e estilo, aplicados à performance expressiva, composição automática e instrumentação. No que diz respeito aos experimentos e testes, realizaram-se considerações sobre complexidade computacional, *benchmarks* e avaliação de acurácia, entre outros. A Tabela 4 apresenta os elementos abordados com relação tanto aos conceitos em Musicologia como aos experimentos e testes.

Com relação ao sistema de representação de regras lógicas, considera-se a utilização da linguagem Prolog, cuja importância decorre de seu aspecto fortemente declarativo. O mecanismo de unificação de Prolog - cujo processo de inferência possui características imperativas [38] - é acionado em segundo plano, devido aos algoritmos de indução lógica aplicados sobre ele. Nesse sentido, destacam-se as ferramentas Aleph e TILDE. Ambas realizam a indução lógica sobre os predicados de Prolog. A teoria Narmour também destaca-se como um importante recurso descritivo. Por meio dessa modelagem, melodias podem ser expressas por listas de estruturas sobrepostas (sequências de Narmour). O conceito de contexto Narmour possibilita a representação conveniente de agrupamentos, onde pode-se expressar tanto a noção de notas sucessoras como a de membresia em grupos específicos.

3 CONCLUSÃO

A presente revisão sistemática possibilitou a avaliação de trabalhos que realizam a representação de conhecimento musical por meio da Programação Lógica Indutiva. A metodologia de pesquisa foi descrita na segunda seção e os resultados foram analisados na terceira seção de acordo com as questões de pesquisa. Uma série de características abordadas pelos trabalhos foi identificada, de modo que a revisão fornece uma visão abrangendo componentes fundamentais na representação de conhecimento musical. Com relação à arquitetura para representação de conhecimento, observou-se a importância fundamental de 3 elementos: linguagem de programação declarativa, mecanismo para implementação da indução lógica e técnica integrada para aquisição de conhecimento. Nesse sentido, a linguagem Prolog destaca-se pela característica expressiva, a qual, aliada a um mecanismo de indução como Aleph ou TILDE, possibilita elevada capacidade representativa, além de simplicidade ao se apresentar tal conhecimento. Os métodos envolvendo algoritmos genéticos, restrições lógicas e mineração de dados tiveram aplicação análoga pelas abordagens estudadas com relação à aquisição de conhecimento. Pode-se observar que os trabalhos analisados encontram-se em diferentes estágios de validação. Como é característico nesse campo, poucas consideração foram desenvolvidas com relação à complexidade computacional. Testes de *benchmark* e avaliações de acurácia foram realizados, entretanto não se observou estudo com relação ao direcionamento a públicos específicos, tanto leigos como especializados. Semelhantemente, não foram realizadas avaliações do ponto de vista da interação humano-computador. Em estudos futuros torna-se im-

Tabela 4: Musicologia, Experimentos e Testes

	1	2	3	4	5	6	7	8	9
Elementos da Musicologia									
Caracterização de estilo			■						
Caracterização de gênero		■	■						
Análise em nível de áudio						■			
Análise em nível melódico/harmônico		■	■						■
Funções tonais	■					■	■		
Graus diatônicos	■	■							
Representação de partitura							■	■	
Sistema Modal					■				
Processamento interativo	■		■						
Voltado à performance expressiva						■	■	■	■
Voltado à composição automática				■	■		■		
Voltado à instrumentação						■		■	
Propósito educacional	■								
Experimentos e Testes									
Considerações sobre complexidade computacional		■				■			
Estudo por processos empíricos		■							
Avaliação por músicos e não músicos				■					
Realiza benchmark			■	■		■	■		■
Realiza beta-teste	■	■						■	
Realiza testes completos		■		■				■	
Avaliação de acurácia		■	■			■		■	■

portante o levantamento de critérios voltados a uma melhor compreensão de tais aspectos.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Flores Fritsch Pimenta Vicari Miletto, Costalonga, "Introdução a computação musical," in *X Congresso Brasileiro de Computação, 4., Itajaí, 2004. Itajaí, SC - Brasil*, 2004, pp. 883–902.

[2] Jose D Fernández and Francisco Vico, "Ai methods in algorithmic composition: A comprehensive survey," *Journal of Artificial Intelligence Research*, pp. 513–582, 2013.

[3] Marvin Minsky, "Music, mind, and meaning," *Computer Music Journal*, pp. 28–44, 1981.

[4] Geraint A Wiggins and Alan Smaill, "Musical knowledge: what can artificial intelligence bring to the musician?," 2000.

[5] Miguel Delgado, Waldo Fajardo, and Miguel Molina-Solana, "A state of the art on computational music performance," *Expert Systems with Applications*, vol. 38, no. 1, pp. 155–160, 2011.

[6] Curtis Roads, "Research in music and artificial intelligence," *ACM Computing Surveys (CSUR)*, vol. 17, no. 2, pp. 163–190, 1985.

[7] Jesus L Alvaro, Eduardo R Miranda, and Beatriz Barros, "Ev: Multilevel music knowlegde representation and programming," *Proceedings of SBCM, Belo Horizonte, Brazil*, 2005.

[8] Rafael Ramirez and Amaury Hazan, "Modeling expressive music performance in jazz.," in *FLAIRS Conference*, 2005, pp. 86–91.

[9] Mira Balaban, "The music structures approach in knowledge representation for," *Computer Music Journal*, vol. 20, no. 2, pp. 96–111, 1996.

[10] Ian Whalley, "International computer music conference 2004: Papers," in *Proceedings International Computer Music Conference, Miami, ICMA Press.* 2005, JSTOR.

[11] Eduardo Morales and Roberto Morales, "Learning musical rules," in *Proceedings of the International Joint Conference on Artificial Inteligence*. Citeseer, 1995.

[12] Simon Dixon, Matthias Mauch, and Amélie Anglade, "Probabilistic and logic-based modelling of harmony," in *Exploring Music Contents*, pp. 1–19. Springer, 2011.

[13] Uros Pompe, Igor Kononenko, and Tjasa Makse, "An application of ilp in a musical database: Learning to compose the two-voice counterpoint," 1996.

[14] Masayuki Numao, Shoichi Takagi, and Keisuke Nakamura, "Constructive adaptive user interfaces-composing music based on human feelings," in *Proceedings of the National Conference on Artificial Intelligence*. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2002, pp. 193–198.

[15] Amélie Anglade, Emmanouil Benetos, Matthias Mauch, and Simon Dixon, "Improving music genre classification using automatically induced harmony rules," *Journal of New Music Research*, vol. 39, no. 4, pp. 349–361, 2010.

[16] Alfonso Perez, Rafael Ramirez, and Stefan Kersten, "Modeling moods in violin performances," in *SMC 08: 5th Sound and Music Computing Conference: Sound in Space-Space in Sound, July 31st-August 3rd, 2008, Berlin, Germany: Proceedings*, 2008, p. 30.

[17] Ashwin Srinivasan, "The aleph manual: version 4 and above," 2007.

[18] Hendrik Blockeel, "Top-down induction of first order logical decision trees," *Aicommunications*, vol. 12, no. 1-2, pp. 119–120, 1999.

[19] Eduardo Morales, "Learning patterns for playing strategies," *International Computer Chess Association Journal*, vol. 17, no. 1, pp. 15–26, 1994.

[20] Yashoda N Sakhare and Mr DB Hanchate, "Comparative study of musical performance by machine learning," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 2, no. 2, pp. 347–352, 2014.

[21] Staffs Keele, "Guidelines for performing systematic literature reviews in software engineering," in *Technical report, Ver. 2.3 EBSE Technical Report. EBSE*. 2007.

[22] Clenio B Gonçalves Jr. and Murillo R P Homem, "Representação de conhecimento musical e programação lógica indutiva - uma revisão sistemática," in *15º Simpósio Brasileiro de ComputaçãO Musical – SBCM*. NUCOM, 2015, pp. 138–141.

[23] Barbara Kitchenham, O Pearl Brereton, David Budgen, Mark Turner, John Bailey, and Stephen Linkman, "Systematic literature reviews in software engineering," *Information and software technology*, vol. 51, no. 1, pp. 7–15, 2009.

[24] Gustavo HRP Tomas, Welington M da Silva, Paulo Anselmo da Mota Silveira Neto, Vinicius Cardoso Garcia, Alexandre Alvaro, and Kiev Gama, "Smart cities architectures-a systematic review.," in *ICEIS (2)*, 2013, pp. 410–417.

[25] Mahvish Khurum and Tony Gorschek, "A systematic review of domain analysis solutions for product lines," *Journal of Systems and Software*, vol. 82, no. 12, pp. 1982–2003, 2009.

[26] Gabriel Abdalla, Carlos Diego Nascimento Damasceno, and Elisa Yumi Nakagawa, "A systematic literature review on systems-of-systems knowledge representation," 2015.

[27] Lianping Chen, Muhammad Ali Babar, and Nour Ali, "Variability management in software product lines," in *Proceedings of the 13th International Software Product Line Conference*. Carnegie Mellon University, 2009, pp. 81–90.

[28] J. Ross Quinlan, "Learning logical definitions from relations," *Machine learning*, vol. 5, no. 3, pp. 239–266, 1990.

[29] Eddy Kwong Mei Chong and Qin Ding, "Symbolic representation of chords for rule-based evaluation of tonal progressions," 2014.

[30] Esteban Maestre, Rafael Ramírez, Stefan Kersten, and Xavier Serra, "Expressive concatenative synthesis by reusing samples from real performance recordings," *Computer Music Journal*, vol. 33, no. 4, pp. 23–42, 2009.

[31] StevenL. Salzberg, "C4.5: Programs for machine learning by j. ross quinlan. morgan kaufmann publishers, inc., 1993," *Machine Learning*, vol. 16, no. 3, pp. 235–240, 1994.

[32] João Paulo Duarte Conceição, "The aleph system made easy," *Integrated Master in Electrical and Computers Engineering, Faculdade de Engenharia da Universidade do Porto*, 2008.

[33] J Salamon, *Melody extraction from polyphonic music signals*, Ph.D. thesis, Department of Information and Communication Technologies Universitat Pompeu Fabra, Barcelona, Spain, 2013.

[34] Erwin De Ley and Dirk Jacobs, "Rules-based analysis with jboss drools: adding intelligence to automation," *Proceedings of ICALEPCS 2011*, pp. 790–793, 2011.

[35] Di Liu, Tao Gu, and Jiang-Ping Xue, "Rule engine based on improvement rete algorithm," in *Apperceiving Computing and Intelligence Analysis (ICACIA), 2010 International Conference on*. IEEE, 2010, pp. 346–349.

[36] Jignesh Patel and Bhavesh A Oza, "Survey on graph pattern mining approach," in *International Journal of Engineering Development and Research*. IJEDR, 2014, vol. 2.

[37] Andrew Koster, Jordi Sabater-Mir, and Marco Schorlemmer, "Trust alignment: a sine qua non of open multi-agent systems," in *On the Move to Meaningful Internet Systems: OTM 2011*, pp. 182–199. Springer, 2011.

[38] José O Guimarães, "Programming languages paradigms," pp. 96–114. UFSCar Sorocaba, 2015.

Sociedade de Engenharia de Áudio

Artigo de Congresso

Apresentado no 14º Congresso de Engenharia de Áudio
20ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2016, São Paulo, SP

Este artigo foi reproduzido do original final entregue pelo autor, sem edições, correções ou considerações feitas pelo comitê técnico. A AES Brasil não se responsabiliza pelo conteúdo. Outros artigos podem ser adquiridos através da Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA, www.aes.org. Informações sobre a seção Brasileira podem ser obtidas em www.aesbrasil.org. Todos os direitos são reservados. Não é permitida a reprodução total ou parcial deste artigo sem autorização expressa da AES Brasil.

Síntese de Campos Sonoros em Painéis Rígidos através de um Arranjo de Fontes Elementares

Clinton André Merlo[1, 2], Alexander Mattioli Pasqual[1] e Eduardo Bauzer Medeiros[1]

[1] Universidade Federal de Minas Gerais, Departamento de Engenharia Mecânica, Belo Horizonte, Minas Gerais, Brasil
[2] Universidade Federal de Itajubá, Departamento de Matemática, Itabira, Minas Gerais, Brasil

clintonmerlo@unifei.edu.br, ampasqual@demec.ufmg.br, bauzer@demec.ufmg.br

RESUMO

Este trabalho investiga a capacidade de arranjos planos de alto-falantes, modelados como monopolos, em sintetizar campos acústicos incidentes em superfícies planas e rígidas. A aplicação vislumbrada é em ensaios de transparência acústica de painéis. Através de simulações numéricas, estuda-se o desempenho de arranjos de monopolos na síntese de ondas planas com incidência oblíqua sobre a superfície de teste. O método das imagens é usado para modelar o efeito do painel rígido no campo sonoro produzido pelas fontes. As intensidades dos monopolos são obtidas a partir da distribuição de pressão acústica almejada sobre o painel através da solução do problema inverso via método dos quadrados mínimos e decomposição em valores singulares truncada. Os efeitos de alguns parâmetros do modelo sobre a qualidade da solução são analisados, tais como, frequência, ângulo desejado de incidência da onda plana, e distância entre o arranjo de fontes e o plano rígido. Os exemplos numéricos apresentados mostram que a síntese de campos sonoros envolve o equilíbrio entre dois resultados desejáveis, porém geralmente conflitantes, a saber, fidelidade da síntese e bom condicionamento do modelo.

0. INTRODUÇÃO

A obtenção de sistemas acústicos de alto desempenho requer a modelagem detalhada dos fenômenos vibroacústicos intervenientes e o estabelecimento de técnicas de caracterização experimental avançadas, as quais forneçam as informações demandadas pelos modelos computacionais mais sofisticados. Tradicionalmente, a avaliação experimental da transparência acústica em painéis é feita utilizando duas salas contíguas acusticamente condicionadas, podendo tratar-se de duas câmaras reverberantes, ou uma câmara reverberante acoplada a uma câmara anecóica. O painel a ser ensaiado é montado na interface entre as salas e a fonte excitadora localizada na câmara reverberante com o intuito de produzir um campo difuso na superfície do painel, fazendo com que

todas as direções de incidência sonora sejam igualmente prováveis [1].

Entretanto, no método clássico a excitação difusa pode não corresponder ao campo incidente no painel durante suas condições reais de uso, comprometendo a qualidade dos resultados experimentais obtidos, como é o caso de painéis submetidos a excitações sonoras provocadas por escoamentos turbulentos [2].

Outra opção seria o desenvolvimento de um sistema para caracterização experimental de painéis que permitisse recriar o campo sonoro incidente na estrutura a ser ensaiada, tais como os gerados por camadas limite turbulentas [3-11], através de um arranjo de alto-falantes adequadamente controlados, ou seja, através da síntese das ondas acústicas que incidiriam no painel em suas condições reais de uso, suplantando assim a metodologia clássica de ensaios, e tornando desnecessária a realização de testes onerosos, tais como medições em vôo e em túneis de vento anecóico. Ressalta-se que esta abordagem tem despertado um interesse crescente por parte da comunidade científica, conforme atestado em muitos trabalhos [12-18].

Teoricamente, qualquer campo sonoro pode ser reconstruído perfeitamente em uma dada região do espaço utilizando um arranjo contínuo de monopolos/dipolos, mas isto não é, obviamente, possível na prática.

O objetivo deste trabalho consiste em simular, numericamente, uma distribuição específica de pressão acústica (onda plana com determinada inclinação e frequência) sobre uma superfície plana, finita e rígida (plano de reprodução representando o painel a ser testado) usando um arranjo plano e finito de monopolos (plano fonte) em campo livre e seu respectivo arranjo de fontes imagem. Desta forma, será possível identificar algumas características e parâmetros do modelo de propagação tais como faixas de frequências, inclinações da onda plana e distâncias entre o arranjo de monopolos e o plano de reprodução que produzirão os melhores resultados associados com o problema da síntese de campos sonoros.

As fontes sonoras individuais do arranjo (os alto-falantes) serão modeladas como monopolos devido à facilidade de lidar com o método das fontes imagens e a simplicidade na formulação matemática e implementação computacional do modelo linear de propagação acústica.

A escolha pelas ondas planas como "alvos" da síntese sonora é devido ao fato de que qualquer campo sonoro, por mais complexo que seja, pode ser decomposto em um conjunto de ondas planas viajando em diferentes direções. Desta forma, se for viável a síntese sonora de um campo de pressão gerado por uma onda plana com determinada inclinação e frequência, por um arranjo de fontes do tipo monopolos, então será possível a síntese de campos sonoros mais complexos.

1. FUNDAMENTOS TEÓRICOS

1.1 Decomposição de Campos Sonoros em Ondas Planas

Ondas harmônicas planas caracterizam-se por apresentar amplitude e fase constantes em qualquer plano perpendicular à sua direção de propagação [19]. O campo de pressão sonora correspondente pode ser expresso como

$$p(x,y,z) = Pe^{j(k_x x + k_y y + k_z z - \omega t)} \tag{1}$$

onde P é a amplitude, t é o tempo, ω é a frequência angular, $j=\sqrt{-1}$, e k_x, k_y e k_z são os componentes em coordenadas cartesianas do vetor número de onda **k** (ver Figura 1). Note que, neste trabalho, assume-se uma dependência temporal da forma $e^{-j\omega t}$.

Como um caso especial, para uma onda plana cujas superfícies de fase constante são paralelas ao eixo z, a eq. (1) se reduz a

$$p(x,y) = Pe^{j(k_x x + k_y y - \omega t)} \tag{2}$$

onde $k_x = k\cos\theta$, $k_y = k\,sen\theta$, $k = |\mathbf{k}|$ é o número de onda e θ o ângulo de inclinação da onda plana com relação ao eixo x [19].

Pela Figura 1, observamos que a onda plana tem um comprimento de onda aparente na direção x de $\lambda_x = 2\pi/k_x$, e como $\lambda/\lambda_x = \cos\theta$, então $k_x = k\cos\theta$. Analogamente, $k_y = k\,sen\theta$.

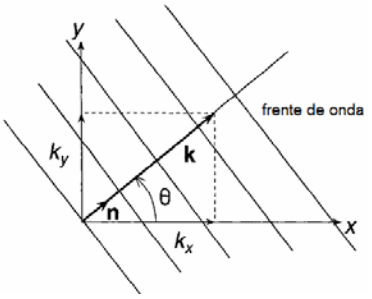


Figura 1: Frente de onda plana em duas dimensões [20].

As superfícies de fase constante são dadas por

$$y = -(k_x/k_y)x + \text{constante} \tag{3}$$

as quais descrevem superfícies planas paralelas ao eixo z com inclinação de $-(k_x/k_y)$ no plano xOy [19].

Uma distribuição espacial qualquer de pressão sonora, em regime permanente com $z>0$, pode ser expressa por uma soma de ondas planas propagativas e evanescentes de diferentes amplitudes e fases [21] da forma

$$p(x,y,z) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_y\, P(k_x,k_y)e^{j(k_x x + k_y y + k_z z)} \tag{4}$$

onde $k_z = \sqrt{k^2 - k_x^2 - k_y^2}$ e o termo $P(k_x, k_y)$ representa o espectro angular [21] da pressão sonora, calculada a partir da pressão no plano $z = 0$, o qual pode ser expresso por

$$P(k_x, k_y) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \; p(x, y, 0) e^{-j(k_x x + k_y y)} \tag{5}$$

Note que o campo de pressão pode ser escrito pela transformada inversa bidimensional de Fourier [21].

As Equações 4 e 5 mostram que praticamente qualquer campo sonoro, por mais complexo que seja, pode ser decomposto através do seu espectro angular em um conjunto de ondas planas viajando em diferentes direções. A importância deste fato que motivou a escolha das ondas planas como "alvos" da síntese sonora neste trabalho.

1.2 Monopolos e Método das Imagens

Fontes ideais do tipo monopolar são de fundamental importância na formulação de soluções para os problemas mais elaborados de radiação e síntese acústica. A amplitude complexa da pressão sonora gerada por um monopolo em campo livre é

$$p(r, k) = -j\omega\rho_0 Q \frac{e^{jkr}}{4\pi r} \tag{6}$$

onde ρ_0 é a densidade de equilíbrio do meio, $(-j\omega\rho_0 Q)$ é a aceleração de volume da fonte, ou simplesmente, amplitude do monopolo e Q é a velocidade de volume da fonte ou sua vazão em m^3/s [21].

O método das imagens pode ser usado para calcular o campo acústico de uma fonte próxima de uma fronteira plana e rígida, representando, por exemplo, o painel a ser testado em estudos de transparência acústica.

Considere um plano perfeitamente rígido situado em $x = 0$, bem como uma fonte monopolar harmônica localizada em $(+d, 0, 0)$. Na ausência do plano rígido, esta fonte produz ondas que se propagam através de todo o espaço, sendo o campo de pressão dado por

$$p(r^-, k) = -j\omega\rho_0 Q \frac{e^{jkr^-}}{4\pi r^-} \tag{7}$$

onde $r^- = [(x-d)^2 + y^2 + z^2]^{1/2}$ é a distância do ponto considerado à fonte [19]. O efeito do plano pode ser modelado substituindo-o por uma segunda fonte, a imagem, operando nas mesmas condições da primeira, porém localizada em $(-d, 0, 0)$. Seu campo de pressão será dado por

$$p(r^+, k) = -j\omega\rho_0 Q \frac{e^{jkr^+}}{4\pi r^+} \tag{8}$$

onde $r^+ = [(x+d)^2 + y^2 + z^2]^{1/2}$ [21].

O campo de pressão resultante em uma região $x > 0$ será dado pela superposição dos respectivos campos indicados nas Equações 7 e 8 [19], ou seja, pela soma

$$p(r^-, k) + p(r^+, k) = -j\omega\rho_0 Q \left(\frac{e^{jkr^-}}{4\pi r^-} + \frac{e^{jkr^+}}{4\pi r^+} \right) \tag{9}$$

Além disso, pode-se mostrar que a componente normal da velocidade de partícula anula-se sobre o plano yOz, conforme requer a condição de plano rígido [19].

2 SÍNTESE DE CAMPOS SONOROS

2.1 Modelagem do Arranjo de Fontes

As fontes sonoras individuais do arranjo (os alto-falantes) serão modeladas neste trabalho como monopolos devido à facilidade de lidar com o método das imagens, a simplicidade na formulação matemática, a fácil implementação computacional, e a eficiência comprovada de modelar fontes de ruído com geometrias arbitrárias. Além disso, será utilizada uma distribuição simétrica e equidistante das fontes no modelo de propagação por ser mais eficiente na estimativa de campos de pressão.

Para a construção do modelo de propagação sonora, consideraremos um arranjo de n monopolos, distribuídos uniformemente, oscilando em um mesmo número de onda k e com velocidades de volume Q_i onde $i = 1, \ldots, n$. Além disso, consideraremos o respectivo arranjo de monopolos imagem, contendo também n monopolos. Assim, utilizando a eq. (9) e o princípio da superposição, a pressão sonora gerada pelo arranjo na posição (x_j, y_j, z_j) é dada por

$$p(x_j, y_j, z_j, k) = -j\omega\rho_0 \sum_{i=1}^{n} Q_i \left[\frac{e^{jkr_{ij}^-}}{4\pi r_{ij}^-} + \frac{e^{jkr_{ij}^+}}{4\pi r_{ij}^+} \right] \tag{10}$$

onde $r_{ij}^- = [(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2]^{1/2}$ é a distância entre o monopolo localizado em (x_i, y_i, z_i) e o ponto de "medição" (x_j, y_j, z_j). Além disso, $r_{ij}^+ = [(x_j + x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2]^{1/2}$ é a distância entre o monopolo imagem em $(-x_i, y_i, z_i)$ e o ponto em (x_j, y_j, z_j).

Assim, considerando m pontos de medição sobre o painel rígido a ser testado, obtém-se o seguinte

sistema linear que representa o modelo de propagação sonora:

$$\mathbf{Aq} = \mathbf{p} \tag{11}$$

onde $\mathbf{A}$ é a matriz de tranferência $m \times n$ do modelo de propagação, cujos elementos, extraídos da eq. (10), são

$$a_{ij} = -j\omega\rho_0 \left[\frac{e^{jkr_{ij}^-}}{4\pi r_{ij}^-} + \frac{e^{jkr_{ij}^+}}{4\pi r_{ij}^+} \right] \tag{12}$$

$\mathbf{q}$ é o vetor $n \times 1$ de velocidade de volume dos monopolos e $\mathbf{p}$ é o vetor $m \times 1$ das amplitudes do campo de pressão sonora medidas no arranjo de microfones. A matriz $\mathbf{A}$ contém todas as características do modelo de propagação, sendo complexa e somente será quadrada quando o número de fontes for igual ao número de microfones. Além disso, os vetores $\mathbf{q}$ e $\mathbf{p}$ também são complexos.

2.2 Solução do Problema Inverso

Considerando um campo sonoro arbitrário que incide em um arranjo de m microfones, com $m > n$, teremos o vetor das amplitudes complexa de pressão sonora $\mathbf{p}$. Para sintetizar o campo sonoro através do modelo de propagação, dado pela eq. (11), recorremos ao Método dos Mínimos Quadrados, uma vez que o sistema linear é super determinado e não pode ser resolvido exatamente ($m > n$). Assim, temos que encontrar uma estimativa do vetor $\mathbf{q}$ das velocidades de volume do arranjo, que minimiza o erro da aproximação

$$\min \left\| \mathbf{Aq} - \mathbf{p} \right\|_2 \tag{13}$$

Quando $\mathbf{A}$ tem posto completo o Problema de Minimização 13, pelo método dos mínimos quadrados, tem solução ótima dada por

$$\mathbf{q}_{MQ} = \left(\mathbf{A}^H \mathbf{A} \right)^{-1} \mathbf{A}^H \mathbf{p} \tag{14}$$

A matriz $\left(\mathbf{A}^H \mathbf{A} \right)^{-1} \mathbf{A}^H$ é a pseudo-inversa de $\mathbf{A}$, representando a generalização padrão de inversão de matrizes retangulares. Quando $\mathbf{A}$ for não-singular e quadrada, sua pseudo-inversa será igual a sua inversa, neste caso a matriz $\mathbf{A}^H \mathbf{A}$ será Hermitiana positiva definida de modo que a inversa existe e a solução dada pela eq. (14) será única.

Para verificar a precisão do modelo de propagação em sintetizar o campo de pressão incidente no arranjo de microfones, podemos calcular o erro relativo da aproximação por

$$\text{Erro}_{MQ} = \frac{\left\| \mathbf{Aq}_{MQ} - \mathbf{p} \right\|_2}{\| \mathbf{p} \|_2} \times 100 \tag{15}$$

Além disso, através da norma do vetor de velocidades de volume das fontes $\left\| \mathbf{q}_{MQ} \right\|_2$ é possível verificar se a síntese do campo sonoro incidente no painel de teste não está sobrecarregando, demasiadamente, os monopolos. Em geral, isto ocorre devido ao mal-condicionamento da matriz do modelo de propagação e também pela incompatibilidade do campo sonoro alvo com o campo que se pode produzir usando arranjos de monopolos.

O número de condição ou condicionamento κ_A é um parâmetro obtido pela razão entre o maior e o menor valor singular da matriz $\mathbf{A}$ do modelo de propagação a ser invertida, $\kappa_A = \sigma_{\max} / \sigma_{\min}$. Um alto valor do número de condição pode indicar grande sensibilidade do sistema linear a erro nos dados utilizados, ruídos de medição e truncamento numérico. Além disso, a matriz $\mathbf{A}^H \mathbf{A}$ pode ter um comportamento quase singular o que cria um enorme problema na inversão numérica dada pela eq. (14). Desta forma, pequenas perturbações em $\mathbf{p}$ podem fazem com que o sistema de equações apresente soluções completamente diferentes. No entanto, é importante perceber que o mal-condicionamento da eq. (11) não é apenas numérico, o próprio sistema físico que gerou as equações pode ser sensível a pequenas alterações (por exemplo, pequenos deslocamentos nos microfones poderia fazer as velocidades de volumes das fontes mudarem de forma exagerada, assim como variações nas distâncias entre os arranjos de fontes e microfones e também a baixa frequência dos monopolos).

2.3 Solução Regularizada do Problema Inverso

Em geral, o mal-condicionamento da matriz $\mathbf{A}$ na eq. (11) está associado com valores singulares muito baixos, os quais representam modos de radição ineficientes do arranjo de fontes. Desta forma, usando a técnica de regularização TSVD (Decomposição em Valores Singulares Truncada) será possível trabalhar apenas com os valores singulares mais altos, os quais estão associados com modos de radiação eficientes. Estes modos eficientes de radiação vão permitir determinar velocidades de volume mais próximas da realidade para os monopolos, porém o erro da aproximação (resíduo) aumentará devido a não estarmos mais trabalhando com a solução ótima original.

Existem dois tipos de classes de problemas relacionados com o mal-condicionamento da matriz dos coeficientes de um sistema linear. Os problemas de posto-deficiente, que são caracterizados por uma matriz $\mathbf{A}$ que tem um conjunto de pequenos valores singulares, e existe uma lacuna bem determinada entre os valores singulares grandes e pequenos. A

outra classe é a de problemas discretos mal colocados, onde todos os valores singulares da matriz $\mathbf{A}$ decrescem gradualmente a zero. Em ambas as classes de problemas, existe uma forte relação entre a quantidade de informação linearmente independente extraída, a norma da solução e o correspondente resíduo [22].

A solução pelo método dos mínimos quadrados para o problema de minimização da eq. (13), usando a decomposição em valores singulares, será dada por

$$\mathbf{q}_{MQ} = \sum_{i=1}^{posto(\mathbf{A})} \sigma_i^{-1} (\mathbf{u}_i^H \mathbf{p}) \mathbf{v}_i \tag{16}$$

Observamos que a solução $\mathbf{q} = \mathbf{A}^{-1}\mathbf{p}$ é um caso particular da eq. (16).

Raramente, os coeficientes da matriz $\mathbf{A}$ tem exatamente posto-deficiente, mas, numericamente, pode apresentar posto-deficiente, ou seja, problemas para os quais existe uma grande diferença entre os pequenos e grandes valores singulares [22], tal que $r_\varepsilon < posto(\mathbf{A}) = n$.

Definimos o posto numérico de uma matriz $\mathbf{A}$, com respeito a uma certa tolerância ε por

$$r_\varepsilon \equiv \min_{\|\mathbf{E}\|_2 \le \varepsilon} posto(\mathbf{A} + \mathbf{E}) \tag{17}$$

onde $\mathbf{E}$ é a matriz de perturbação. Ou seja, o posto r_ε será o número de colunas, realmente, linearmente independentes para qualquer pertubação de $\mathbf{A}$ com norma menor ou igual a tolerância ε [22].

A regularização de problemas mal-condicionados consiste em substituir $\mathbf{A}$ por uma matriz $\mathbf{A}_k$ que esteja próxima e que seja, matematicamente, posto-deficiente, dada por

$$\mathbf{A}_k \equiv \sum_{i=1}^{k} \mathbf{u}_i \sigma_i \mathbf{v}_i^H \tag{18}$$

Neste caso, todos os pequenos valores singulares não-zeros $\sigma_{k+1}, \ldots, \sigma_n$ são substituídos exatamente por zeros. Em geral, a matriz $\mathbf{A}_k$ tem posto $k = r_\varepsilon$, pois se $k < r_\varepsilon$ teremos a perda de informação associada com valores singulares maiores e se $k > r_\varepsilon$ teremos soluções com norma muito elevadas [22]. Desta forma, o novo problema de mínimos quadrados $\min\|\mathbf{A}_k\mathbf{q} - \mathbf{p}\|_2$ terá solução regularizada única, chamada de TSVD (Decomposição em Valores Singulares Truncada), dada por

$$\mathbf{q}_{TSVD} = \sum_{i=1}^{k} \sigma_i^{-1} (\mathbf{u}_i^H \mathbf{p}) \mathbf{v}_i \tag{19}$$

ou equivalentemente,

$$\mathbf{q}_{TSVD} = \left(\mathbf{A}_k^H \mathbf{A}_k\right)^{-1} \mathbf{A}_k^H \mathbf{p} \tag{20}$$

Em geral, nos problemas de regularização a norma $\|\mathbf{q}_{TSVD}\|_2$ poderá ser muito menor que a norma $\|\mathbf{q}_{MQ}\|_2$, o que implicará em um erro de aproximação regularizado Erro_{TSVD} maior que um erro não regularizado Erro_{MQ} [22].

3. RESULTADOS

Este trabalho consistiu na síntese numérica de um campo de pressão sonora (uma onda plana com determinada amplitude, inclinação e frequência) que incide em um arranjo plano e quadrado (painel rígido a ser testado, por exemplo) discretizado por 1600 pontos de medida (representando os microfones) igualmente espaçados de 0,01 m. O campo alvo foi sintetizado utilizando um arranjo plano e quadrado de 16 monopolos (representando as fontes) igualmente espaçados de 0,02 m e o seu respectivo arranjo imagem, ambos a uma distância d do painel rígido, paralelos a este e oscilando na mesma frequência da onda plana almejada.

Para identificar os possíveis parâmetros e as principais dificuldades associadas com o problema de síntese sonora foram feitas diversas simulações numéricas com o modelo de propagação para verificar o comportamento do erro relativo da aproximação Erro_{MQ} (resíduo relativo), da norma $\|\mathbf{q}_{MQ}\|$ do vetor das velocidades de volume dos monopolos e do número de condição da matriz do modelo de propagação κ_A.

Diversos parâmetros e variáveis podem afetar a síntese de campos sonoros, tais como frequência, fase, distância entre o plano de síntese e o arranjo de monopolos, quantidade de monopolos e pontos de discretização no painel, distância entre os pontos de discretização, distância entre os monopolos e distribuição espacial das fontes. Para o caso da síntese de ondas planas em superfícies planas, a frequência e a inclinação θ desempenham um papel importante, como veremos a seguir.

Como exemplo, na Figura 2 temos a superfície representando o campo de pressão acústico gerado sobre o painel rígido por uma onda plana oscilando a 500 Hz, de amplitude $P = 1$ e inclinada de $\theta = 90°$ com relação ao eixo x, onde a cor indica o ângulo (graus) de fase da onda plana incidente. A síntese do campo de pressão acústico pelo arranjo de monopolos distante $d = 0{,}5$ m do painel rígido é apresentada na Figura 3. Além disso, os valores obtidos das medidas de controle associadas com a

síntese sonora foram: $\text{Erro}_{MQ} = 3{,}3\,\%$, $\|\mathbf{q}_{MQ}\| = 418{,}64$ e $\kappa_A = 9{,}86 \times 10^9$.

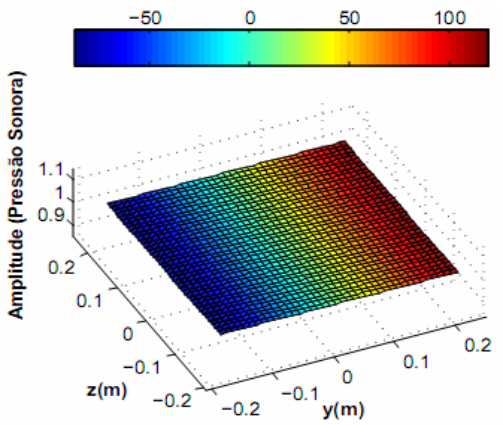


Figura 2: Campo de pressão gerado por uma onda plana.

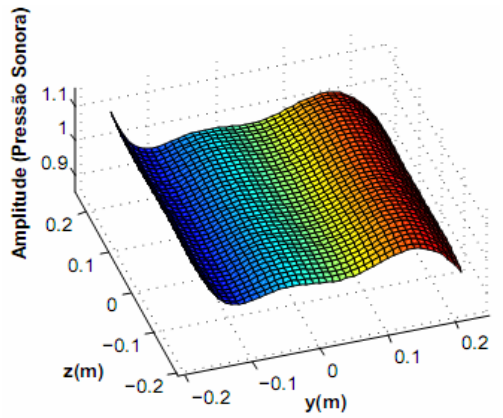


Figura 3: Síntese do campo de pressão por um arranjo de 16 monopolos.

Pela Figura 4 observa-se o erro de síntese de uma onda plana pelo arranjo de monopolos com $d = 0{,}5$ m em função de θ e da frequência. Devido ao *aliasing* espacial, o arranjo de monopolos oscilando em altas frequências produz campos de pressão muito complexos quando comparados aos produzidos por ondas planas. Desta forma, os campos de pressão gerados por ondas planas em altas frequências são mais difíceis de serem sintetizados por arranjos de monopolos, principalmente, para inclinações mais elevadas próximas de 90 graus, conforme mostrado na Figura 4.

No geral, a síntese sonora melhora para frequências e inclinações mais baixas, pois apresenta menores valores de Erro_{MQ} e velocidades de volume $\|\mathbf{q}_{MQ}\|$, conforme observado nas Figuras 4 e 5. Note que $\theta = 0°$ corresponde ao caso de incidência normal.

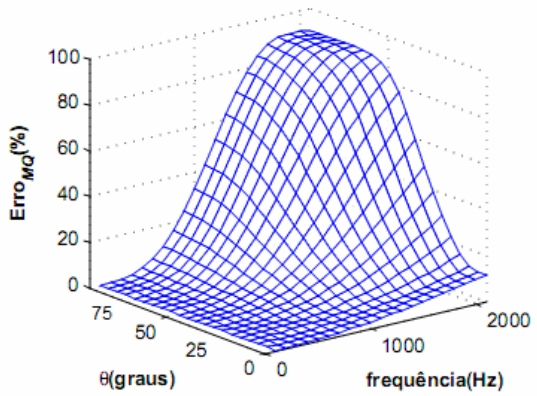


Figura 4: Variação do Erro_{MQ} para uma distância d = 0,5 m entre o arranjo de fontes e o plano de síntese.

O campo de pressão a ser reproduzido se torna mais complexo e irregular à medida que a onda plana assume inclinações mais elevadas, isto faz com que o modelo de propagação sobrecarregue as fontes, aumentando a norma $\|\mathbf{q}_{MQ}\|$ e consequentemente o Erro_{MQ}, conforme observado pelas Figuras 4 e 5.

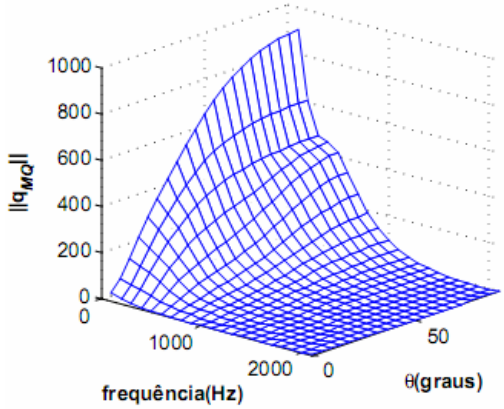


Figura 5: Variação da $\|\mathbf{q}_{MQ}\|$ para uma distância d = 0,5 m entre o arranjo de fontes e o plano de síntese.

Com o aumento da distância entre o arranjo de monopolos e o plano de síntese, ocorrerá naturalmente uma suavização do campo sonoro sintetizado, principalmente nas frequências mais baixas, contribuindo para a diminuição do erro da aproximação, conforme pode ser observado na Figura 6 para algumas distâncias d entre o plano de síntese e o arranjo de fontes oscilando a 500 Hz. Porém, para compensar o aumento da distância entre o arranjo de fontes e o plano de síntese o modelo de propagação aumenta as velocidades de volume das fontes, o que eleva a norma $\|\mathbf{q}_{MQ}\|$, conforme pode ser observado na Figura 7 para uma frequência de 500 Hz.

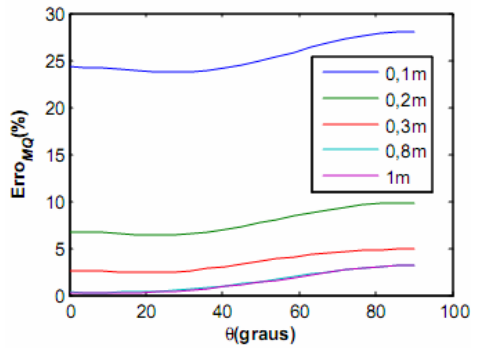


Figura 6: Erro_{MQ} em função de θ para várias distâncias *d*.

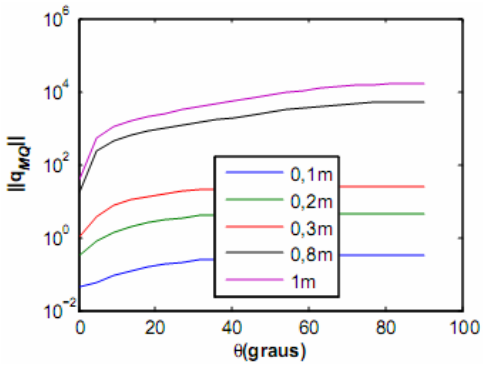


Figura 7: $\|\mathbf{q}_{MQ}\|$ em função de θ para várias distâncias *d*.

Na solução do problema inverso podem surgir valores elevados de $\|\mathbf{q}_{MQ}\|$ devido, em geral, ao modelo de propagação ser mal condicionado e também incompatível com o campo sonoro que se quer sintetizar.

Na Figura 8, observamos que a matriz $\mathbf{A}$ do modelo de propagação apresenta um número de condição κ_A muito elevado, o qual decresce com o aumento da frequência e diminuição da distância entre o arranjo e o plano de síntese. Em geral, se o modelo de propagação for mal-condicionado, sua matriz $\mathbf{A}$ terá um κ_A muito elevado.

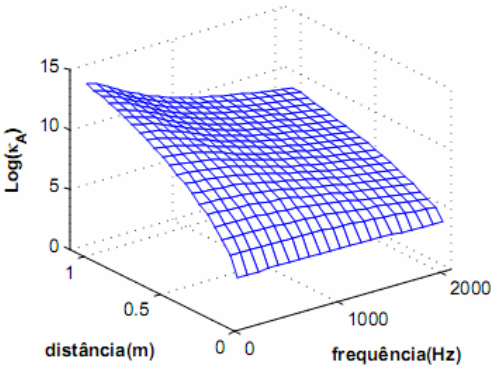


Figura 8: κ_A em função da distância e frequência.

Desta forma, usando a técnica de regularização TSVD (Decomposição em Valores Singulares Truncada) no problema de síntese da onda plana inclinada de $\theta = 90°$, amplitude $P = 1$ e frequência de 500 Hz pelo arranjo de monopolos distante $d = 0{,}5$ m do painel rígido, reduziu-se consideravelmente a norma do vetor de velocidades de volume das fontes de $\|\mathbf{q}_{MQ}\| = 418{,}64$ para $\|\mathbf{q}_{TSVD}\| = 4{,}05$ ao custo de um pequeno aumento no $\text{Erro}_{MQ} = 3{,}3\%$ para $\text{Erro}_{TSVD} = 5{,}3\%$.

O aumento no erro de aproximação se deve em parte à escolha da faixa de valores singulares muito pequenos que serão descartados durante o processo de regularização TSVD. Quanto mais ampla for a faixa de valores singulares pequenos que serão descartados maior será o erro da aproximação pois não estaremos mais trabalhando com a solução ótima original.

Geralmente, o processo de regularização TSVD contribuirá para o aumento do erro de aproximação e a diminuição da velocidade de volume das fontes. Este fato pode ser observado pela comparação entre as Figuras 9 e 6 e pelas Figuras 10 e 7, respectivamente, para diferentes distâncias entre o plano de síntese e o arranjo de monopolos oscilando a 500 Hz. Além disso, a regularização TSVD é mais eficiente em reduzir a velocidade de volume das fontes apenas quando se tem distâncias maiores entre o arranjo de monopolos e o plano de síntese, pois nessas condições o mal-condicionamento de $\mathbf{A}$ é mais severo, conforme pode ser observado na Figura 8.

Portanto, dependendo da precisão requerida na síntese de campos sonoros o processo de regularização TSVD deve ser utilizado de forma a encontrar um ponto de equilíbrio entre a redução da velocidade de volume das fontes e o consequente aumento do erro de aproximação.

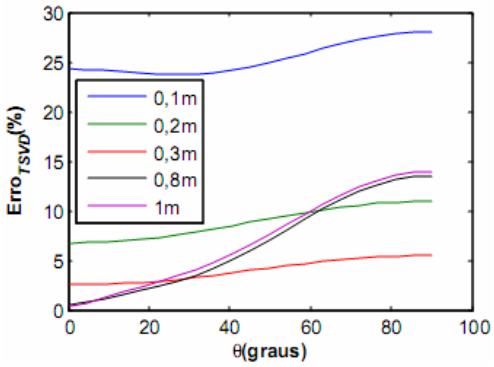


Figura 9: Erro_{TSVD} em função de θ para várias distâncias *d* entre o arranjo de fontes e o plano de síntese.

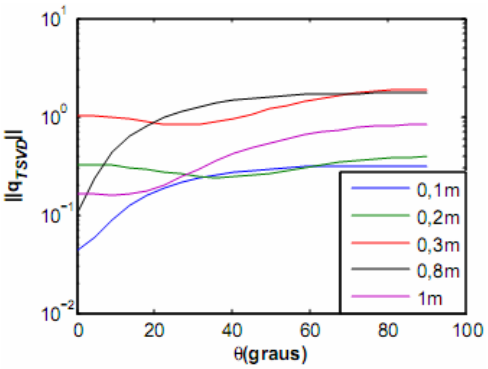


Figura 10: $\|\mathbf{q}_{TSVD}\|$ em função de θ para várias distâncias d entre o arranjo de fontes e o plano de síntese.

Outra maneira de reduzir o Erro_{MQ} para faixas de frequências e inclinações mais elevadas é aumentar a quantidade de monopolos. Porém, com o aumento da quantidade de monopolos, ocorre naturalmente o aumento da norma $\|\mathbf{q}_{MQ}\|$, mas também o aumento da velocidade de volume Q_i nas próprias fontes tomadas individualmente. Pelas Figuras 11 e 12 observa-se o aparente aumento do valor absoluto das velocidades de volume das fontes para arranjos de quatro e nove monopolos, respectivamente. Optou-se por colocar os arranjos a uma distância $d = 0{,}5$ m do plano de síntese e onda plana inclinada de $\theta = 30°$.

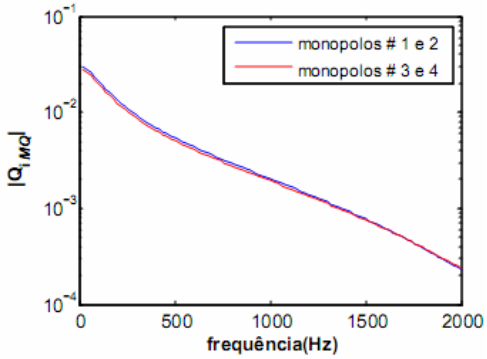


Figura 11: $|Q_{i\,MQ}|$ para o arranjo de 4 monopolos.

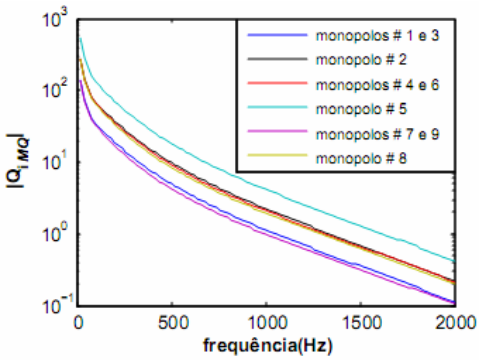


Figura 12: $|Q_{i\,MQ}|$ para o arranjo de 9 monopolos.

Simulações numéricas, obtidas pelo arranjo de monopolos distante $d = 0{,}5$ m do plano de síntese e $\theta = 90°$, mostram que o modelo de propagação também é influenciado por outras variáveis como distância entre os pontos de discretização, quantidade destes pontos e distância entre os monopolos.

A redução da distância entre os pontos de discretização contribui para a diminuição do Erro_{MQ} e o aumento do condicionamento da matriz do modelo de propagação, conforme observado pelas Figuras 13 e 14. Consequentemente, para κ_A mais elevado tem-se o aumento da norma $\|\mathbf{q}_{MQ}\|$. Além disso, para uma amostragem adequada do campo de pressão a distância entre os pontos de discretização deve ser menor que o comprimento de onda.

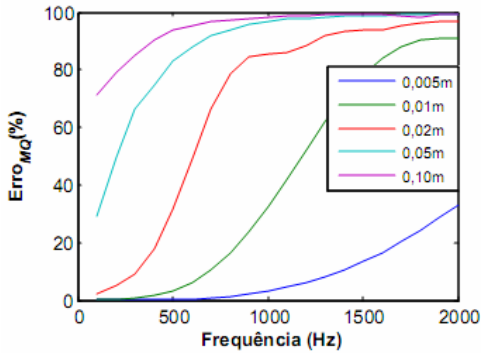


Figura 13: Erro_{MQ} em função da frequência para várias distâncias entre os pontos de discretização.

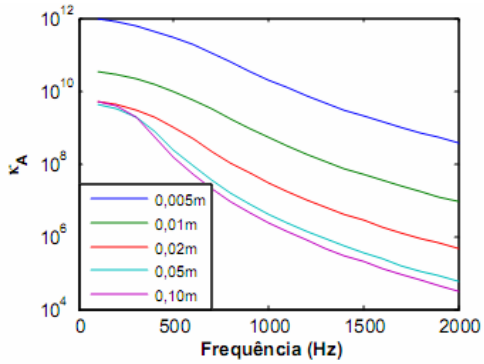


Figura 14: κ_A em função da frequência para várias distâncias entre os pontos de discretização.

Por outro lado, com o aumento da quantidade de pontos de discretização ocorre a elevação do Erro_{MQ} e a consequente redução do condicionamento κ_A da matriz do modelo de propagação, conforme observado pelas Figuras 15 e 16. Desta forma, a norma do vetor de velocidades de volume $\|\mathbf{q}_{MQ}\|$ também será reduzida.

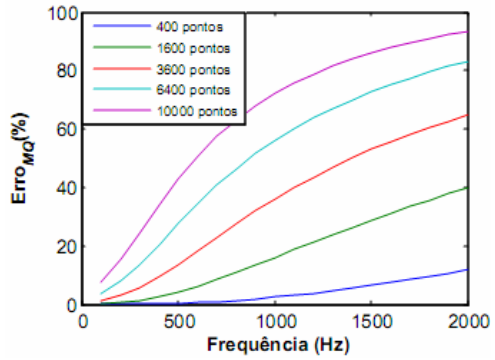


Figura 15: Erro_{MQ} em função da frequência para várias quantidades de pontos de discretização.

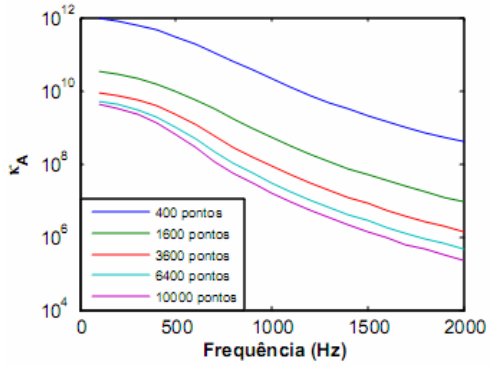


Figura 16: κ_A em função da frequência para várias quantidades de pontos de discretização.

Distâncias muito pequenas entre as fontes podem ser usadas no modelo de propagação uma vez que os monopolos são fontes pontuais idealizadas. Assim, reduzindo a distância entre as fontes ocorre a elevação do condicionamento do sistema, porém sem variação substancial do Erro_{MQ}, conforme observado pelas Figuras 17 e 18. Consequentemente com a elevação do condicionamento tem-se o aumento da norma $\|\mathbf{q}_{MQ}\|$.

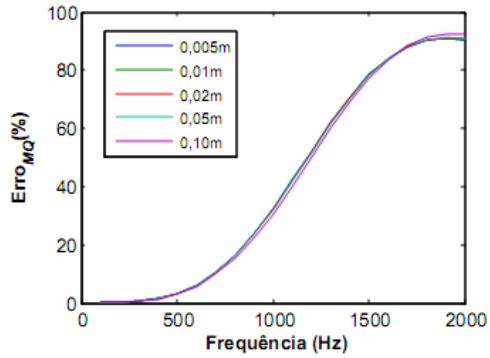


Figura 17: Erro_{MQ} em função da frequência para várias distâncias entre os monopolos.

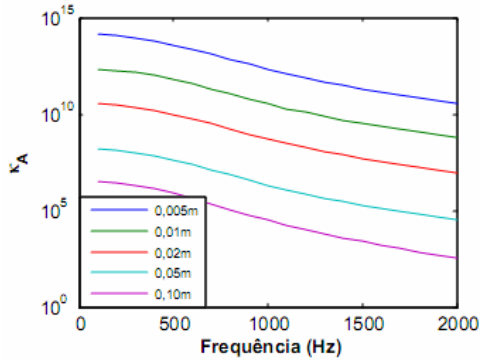


Figura 18: κ_A em função da frequência para várias distâncias entre os monopolos.

4. CONCLUSÕES

Uma das principais características de se usar fontes do tipo monopolo é a simplificação matemática do modelo de propagação e também de sua fácil implementação computacional.

Os resultados obtidos neste trabalho usando o modelo de propagação por arranjo de monopolos sugerem que é possível sintetizar, em campo próximo, a distribuição de pressão gerada por uma onda plana sobre um painel finito e rígido para diversas inclinações e frequências de forma razoável, mesmo sabendo que o modelo não é totalmente compatível com este tipo de distribuição de pressão acústica.

Pela análise do erro de aproximação e da norma do vetor das velocidades de volume dos monopolos para os casos com e sem regularização, é possível identificar as melhores faixas de frequências, inclinações, quantidade de monopolos e distâncias entre o arranjo de fontes e o plano de síntese que produzirá soluções mais satisfatórias.

Através de simulações numéricas, mostrou-se que tanto o erro de síntese quanto a norma do vetor de amplitudes das fontes aumentam com a inclinação almejada para a onda plana. Pôde-se observar que o aumento da distância entre o arranjo de fontes e o plano de síntese resulta em um crescimento do número de condição κ_A da matriz do modelo de propagação. Resultado semelhante é obtido com o aumento da quantidade de monopolos usados no modelo. Porém, quanto maior a quantidade de fontes, melhor será o resultado obtido na síntese sonora.

Portanto, concluiu-se que o condicionamento do sistema linear não está associado com a qualidade da solução estimada pelo modelo, mas sim com a estabilidade da solução do sistema de equações algébricas. Desta forma, nem sempre o modelo com melhor condicionamento apresentará melhor resultado na síntese sonora. Em geral, um elevado

κ_A está associado a uma instabilidade da solução do modelo de propagação a perturbações no campo de pressão alvo.

Agradecimentos

Os autores agradecem ao PPGMEC (Programa de Pós-Graduação em Engenharia Mecânica) da UFMG pelo apoio financeiro.

5. REFERÊNCIAS BIBLIOGRÁFICAS

[1] BS EN ISO 140-3. "Acoustics, measurement of sound insulation in buildings and of building elements. Part 3: Laboratory measurements of airborne sound insulation of building elements", 1995.

[2] T. Bravo; C. Maury. "A synthesis approach for reproducing the response of aircraft panels to a turbulent boundary layer excitation", *Journal of the Acoustical Society of America*, vol. 129, no. 1, pp. 1-11, 2011.

[3] G. M. Corcos. "Resolution of pressure in turbulence", *Journal of the Acoustical Society of America*, vol. 35, no. 2, pp. 192-199, 1963.

[4] R. Lyon, "Boundary layer noise response simulation with a sound field," *NASA*, CR-56074, 1964.

[5] F. Fahy, "On simulating the transmission through structures of noise from turbulent boundary layer pressure fluctuations," *Journal of the Acoustical Society of America*, vol. 3, pp. 57–81, 1966.

[6] M. S. Howe. "Surface pressures and sound produced by turbulent flow over smooth and rough walls", *Journal of the Acoustical Society of America*, vol. 90, no. 2, pp. 1041-1047, 1991.

[7] W. R. Graham. "A comparison of models for the wavenumber-frequency spectrum of turbulent boundary layer pressures", *Journal of Sound and Vibration*, vol. 206, no. 4, pp. 541-565, 1997.

[8] M. Goody. "Empirical spectral model of surface pressure fluctuations". *AIAA Journal*, vol. 42, no. 9, pp. 1788-1794, 2004.

[9] X. Gloerfelt; J. Berland. "Direct computation of turbulent boundary layer noise", Proc. 15th AIAA/CEAS Aero Acoustics Conference, paper 2009-3401, 2009.

[10] B. Arguillat; D. Ricot; C. Bailly; G. Robert. "Measured wavenumber: Frequency spectrum associated with acoustic and aerodynamic wall pressure fluctuations", *Journal of the Acoustical Society of America*, vol. 128, no. 4, pp. 1647-1655, 2010.

[11] J. Antunes; X. Delaune; P. Piteau. "Time-domain modeling of the random vibrations of tubes subjected to turbulence-conveying flows," *Proc. ASME - Pressure vessels & piping conference*, 2011.

[12] D. H. Trivett; L. D. Luker; S. Petrie; A. L. Van Buren; J. E. Lue. "A planar array for the generation of evanescent waves", *Journal of the Acoustical Society of America*, vol. 87, no. 6, pp. 2535-2540, 1990.

[13] O. Kirkeby; P. A. Nelson. "Reproduction of plane wave sound fields," *Journal of the Acoustical Society of America*, vol. 94, pp. 2992–3000, 1993.

[14] C. Maury; T. Bravo. "The experimental synthesis of random pressure fields, Part II: Practical feasibility", *Journal of the Acoustical Society of America*, vol. 120, no. 5, pp. 2712-2723, 2006.

[15] T. Bravo; C. Maury. "Enhancing low frequency sound transmission measurements using a synthesis method", *Journal of the Acoustical Society of America*, vol. 122, no. 2, pp. 869-880, 2007.

[16] T. Bravo; C. Maury. "A synthesis approach for reproducing the response of aircraft panels to a turbulent boundary layer excitation", *Journal of the Acoustical Society of America*, vol. 129, no. 1, pp. 1-11, 2011.

[17] O. Robin; A. Berry; S. Moreau. "Reproduction of random pressure fields based on planar near field acoustic holography", *Journal of the Acoustical Society of America*, vol. 133, no. 6, pp. 3885-3899, 2013.

[18] A. Pasqual; Ph. Herzog. "Pilotage d'un réseau de sources acoustiques à partir de ses modes de rayonnement", *Actes du 12e Congrès Français d'Acoustique*, pp. 1885–1891, 2014.

[19] L. E. Kinsler; A. R. Frey; A. B. Coppens; J. V. Sanders. Fundamentals of Acoustics. 3 ed. New York, John Wiley & Sons. 1982.

[20] F. Fahy. Foundations of Engineering Acoustics. 1.ed. Elsevier, 2001.

[21] E. G. Williams. Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography. Academic Press. 1999.

[22] P. C. Hansen. Rank-deficient and Discrete ill-posed Problems: Numerical Aspects of Linear Inversion. Philadelphia. SIAM. 1998.

Sociedade de Engenharia de Áudio

Artigo de Congresso

Apresentado no 14º Congresso de Engenharia de Áudio
20ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2016, São Paulo, SP

Este artigo foi reproduzido do original final entregue pelo autor, sem edições, correções ou considerações feitas pelo comitê técnico. A AES Brasil não se responsabiliza pelo conteúdo. Outros artigos podem ser adquiridos através da Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA, www.aes.org. Informações sobre a seção Brasileira podem ser obtidas em www.aesbrasil.org. Todos os direitos são reservados. Não é permitida a reprodução total ou parcial deste artigo sem autorização expressa da AES Brasil.

Ajuste de modelos analíticos de microfone piezoelétrico a partir de modelos numéricos

Bernardo H. P. Murta,[1] Eric Brandão,[1] Homero S. Silva,[2] Paulo H. Mareze,[1] William D'A. Fonseca[1] e Julio A. Cordioli[3]

[1] Universidade Federal de Santa Maria, Curso de Engenharia Acústica, DECC
Santa Maria, RS, 97105-900, Brasil

[2] Pesquisador Independente

[3] Universidade Federal de Santa Catarina, Laboratório de Vibrações e Acústica,
Florianópolis, SC, 88040-900, Brasil

be.murta@gmail.com, eric.brandao@eac.ufsm.br, homerosette@hotmail.com,
paulo.mareze@eac.ufsm.br, will.fonseca@eac.ufsm.br, julio.cordioli@ufsc.br

RESUMO

O presente artigo visa a exploração de técnicas de ajuste de modelos analíticos a partir de existentes modelos numéricos. Esses procedimentos são úteis pois modelos analíticos possuem menor custo computacional, o que alavanca o projeto de transdutores eletroacústicos possibilitando otimizações para definição de materiais e geometrias. Técnicas de medição dos parâmetros Thiele-Small de alto-falantes são adaptadas nos cálculos, que utilizam dados da relação entre pressão sonora incidente e a tensão elétrica de saída, ou seja, da função de transferência do dispositivo. Com isso, pode-se desenvolver modelos analíticos, baseados na teoria de circuitos equivalentes, mais exatos, bem como direcionar maneiras de medir propriedades específicas, como compliância mecânica, de tais transdutores efetivamente. O modelo numérico estudado, de um microfone piezoelétrico, foi validado experimentalmente, o que sustenta a relevância do presente estudo. Trabalhos futuros visam a extensão do estudo para ajustar outras propriedades, como a impedância elétrica do transdutor, o fator de transformação de energia mecânica em elétrica e os componentes resistivos correspondentes às perdas do sistema.

0 INTRODUÇÃO

Atualmente, etapas de desenvolvimento de transdutores eletroacústicos são cada vez mais baseadas em modelos numéricos e analíticos, fator que contribui na diminuição dos custos de produção. Desse modo, pode-se projetar microfones e alto-falantes tendo como meta características específicas, como faixa de frequência plana, alta sensibilidade ou largas bandas de atuação. Métodos de otimização, como algorítimos genéticos, são geralmente utilizados junto a esses modelos, permitindo ao projetista definir as geometrias e materiais mais adequados para atingir os parâmetros desejados [1, 2].

Métodos matemáticos robustos, como o de elementos finitos (FEM) e o de elementos de contorno (BEM), são implementados por diversos *software* comerciais que permitem a modelagem e análise de transdutores com geometrias complexas e acoplamentos entre diversos ramos da física, como acústica, mecânica e elétrica [3]. Entretanto, o alto custo computacional e frequentes longos tempos de cálculo, podem se tornar obstáculos quando há interesse de se aplicar técnicas de otimização, que consistem de diversas resoluções de modelos semelhantes. Uma alternativa utilizada é o desenvolvimento de modelos analíticos, que apesar de levarem em conta diversas simplificações e aproximações, geralmente possuem baixo custo computacional. Desse modo, pode-se obter com um mais curto tempo de cálculo resultados satisfatoriamente semelhantes aos de modelos numéricos, desde que devidamente ajustados.

Modelos analíticos baseados em analogias por parâmetros concentrados revolucionaram o desenvolvimento de transdutores nas décadas de 60 e 70 devido ao advento dos computadores [4]. Nessa abordagem, as equações diferenciais que regem o modelo do transdutor estudado são substituídas por equações de braço de elementos passivos de circuitos elétricos, representando as impedâncias mecânicas e acústicas por meio de resistores, capacitores e indutores. Com isso, pode-se desenvolver um único circuito que compreende todo o acoplamento eletro-mecano-acústico e resolvê-los com base nas Leis de Kirchhoff [5].

Por sua vez, deve-se ter em mente as limitações das analogias eletroacústicas, dentre elas a necessidade dos comprimentos de onda analisados terem dimensões suficiente maiores que qualquer dimensão relevante do transdutor modelado. Assim, a energia do sinal é distribuída na geometria com pouca variação espacial [6, 7], ou seja, o diafragma apenas se move como um pistão, no caso de um microfone ou alto-falante. Outro limitante é a complexidade dos elementos do transdutor, como no caso de diafragma composto de mais de um material ou geometrias mais elaboradas e sem simetria.

Para contornar essas limitações, autores vêm desenvolvendo modelos mais complexos, considerando o tipo de material, condições de engaste e a própria geometria [6, 8]. Desse modo, é possível modelar pequenos detalhes de microfones, *e.g.*, o efeito da presença de uma placa (*backplate*) perfurada [9] (paralela ao diafragma), na resposta de transdutores de medição [10]. Assim, modelos analíticos cada vez mais elaborados são desenvolvidos e utilizados na indústria [11] como uma ferramenta relevante em projetos. Ainda há limitações, como no desenvolvimento de transdutores mais modernos, como os *microelectromecanical systems*[1] (MEMS) [7, 12] que possuem geometrias não usuais, em relação a microfones comuns, e camadas com diversos materiais de propriedades diferentes compondo seus diafragmas; bem como de transdutores piezoelétricos, que compartilham dessa complexidade de materiais no diafragma [13].

Este estudo visa, então, à avaliação da aplicabilidade de técnicas de ajuste de modelos analíticos de parâmetros concentrados por meio de modelos numéricos em FEM. Mais especificamente, estimou-se a compliância mecânica do diafragma bimorfo[2] de um microfone piezoelétrico modelado em elementos finitos e validado experimentalmente [13]. Decidiu-se por estimar esse parâmetro por ser uma incógnita devido à falta de fórmulas matemáticas para a sua caracterização, considerando a geometria em questão. A compliância mecânica é um parâmetro importante, pois a frequência de ressonância do sistema é sensível à mesma. Outros parâmetros como a resistência mecânica e resistência acústica, que alteram a proeminência do pico ressonante na resposta em frequência (FRF[3]) ou a capacitância elétrica do transdutor, que altera a sensibilidade, são também incógnitas, porém não foram abordadas neste trabalho.

A Seção 1 descreve os modelos numérico e analítico, bem como as técnicas utilizadas na estimativa da compliância mecânica utilizada no ajuste dos modelos. Os resultados são apresentados na Seção 2 e discutidos na Seção 3.

1 METODOLOGIA

Na presente seção os modelos analítico e numérico que basearam o estudo do transdutor são apresentados. O microfone analisado possui geometria simples, composto de um diafragma bimorfo e uma cavidade fixada na extremidade oposta ao piezoelétrico, vide Figura 1. Com variações das dimensões da cavidade acústica, pode-se estimar a compliância mecânica, com base nos métodos de medição de parâmetros de Thiele-Small de alto-falantes [14].

1.1 Modelo Analítico

A modelagem analítica segue os passos de montagem dos circuitos de cada domínio desenvolvidos por

[1] Sistemas eletromecânicos de dimensões microscópicas.

[2] Diafragma composto por dois materiais, nesse caso de um disco metálico em latão sob um disco de material piezoelétrico.

[3] A Função de Resposta em Frequência também pode ser chamada de função de transferência.

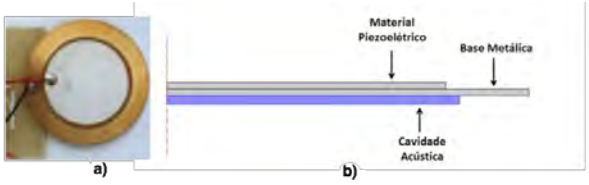


Figura 1: **a)** Exemplo do diafragma estudado; **b)** Detalhes da geometria do modelo FEM, acoplada à uma cavidade. O modelo usa axi-simetria para diminuir o custo computacional [13].

Rossi [6] e adaptados por Brandão [15]. Cada domínio é, então, montado separadamente e depois acoplado por transformadores ideais, cuja razão de número de espiras corresponde ao fator de conversão de cada acoplamento. Em seguida, os elementos do domínio acústico P, pressão incidente, Z_{rad}, impedância de radiação do diafragma e C_{a}, compliância acústica da cavidade, são rebatidos ao domínio mecânico, de acordo com o fator de acoplamento S_d, a área do diafragma, vide circuito exposto na Figura 2. O circuito foi representado segundo os análogos de impedância, de modo que no domínio mecânico a corrente representa a velocidade e a tensão sobre um elemento representa a força atuante no mesmo[4].

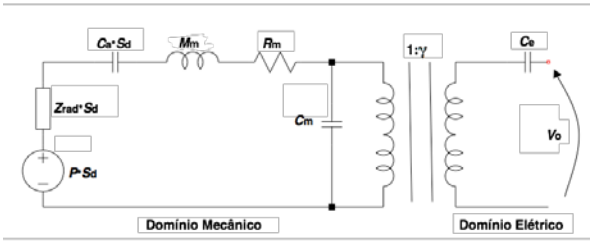


Figura 2: Circuito equivalente do transdutor com as impedâncias do domínio acústico rebatidas para o domínio mecânico. O transformador representa o acoplamento com o lado elétrico com um fator de transformação γ.

O acoplamento com o domínio elétrico, por sua vez, é definido por Beranek [4] como uma razão entre a capacitância elétrica do transdutor C_{e} e a sua constante piezoelétrica, no caso de deformações em apenas uma direção e diferença de potencial gerada na mesma direção d_{33}, definido como:

$$\gamma = \frac{d_{33}}{C_{\mathrm{e}}} \left[\frac{\mathrm{V}}{\mathrm{N}}\right] . \tag{1}$$

Entretanto, como o movimento de um disco excitado por uma pressão sonora incidente se dá de maneira radial e transversal, esse acoplamento se torna mais complexo e deve ser melhor representado em trabalhos futuros. Essa simplificação faz com que os valores de sensibilidade encontrados no modelo analítico sofram um desvio em relação aos valores reais. Porém, ao se calcular a função de transferência de circuito aberto, variações desse fator de acoplamento, que é constante na frequência, apenas deslocam a curva de FRF como um todo para cima ou para baixo, variando apenas a sensibilidade do transdutor.

As impedâncias foram modeladas de acordo com as seguintes equações:

Compliância[5] da cavidade

Cavidades que delimitam volumes V de ar, que possuem uma impedância com característica capacitiva C_{a} $\left[\frac{\mathrm{m}^3}{\mathrm{Pa}}\right]$ [6], são dadas por:

$$C_{\mathrm{a}} = \frac{V}{\rho_0 c_0^2} , \tag{2}$$

em que ρ_0 é a densidade do ar e c_0 a velocidade do som nesse meio.

Impedância de radiação

A impedância de radiação de um pistão de raio a no fim de um tubo pode ser utilizada para representar o fenômeno físico no presente estudo. Beranek [4] deriva o circuito no domínio acústico, Figura 3, para descrever esse comportamento. Os componentes devem ser modelados com as seguintes propriedades:

$$R_{\mathrm{a1}} = \frac{0,1977\rho_0 c_0}{a^2} , \tag{3}$$

$$R_{\mathrm{a2}} = \frac{\rho_0 c_0}{S_{\mathrm{d}}} , \tag{4}$$

$$C_{\mathrm{a1}} = \frac{\pi a^3}{\rho_0 c_0^2} \tag{5}$$

e

$$M_{\mathrm{a1}} = \frac{2\rho_0}{3\pi^2 a} . \tag{6}$$

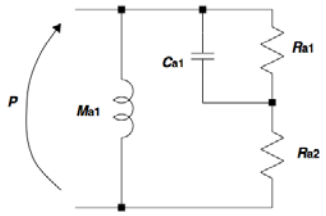


Figura 3: Circuito equivalente da impedância de radiação no domínio acústico.

A impedância de radiação Z_{rad} pode então ser derivada a partir do equivalente Z_{radEq1} entre os resistores e o capacitor

$$Z_{\mathrm{radEq1}} = \left(\frac{1}{R_{\mathrm{a1}}} + \mathrm{j}\omega C_{\mathrm{a1}}\right)^{-1} + R_{\mathrm{a2}} , \tag{7}$$

[4]Considerando o fator de conversão acústico-mecânico representado pela área do diafragma, tem-se que no domínio acústico a tensão sobre um elemento corresponde à pressão e a corrente à velocidade de volume.

[5]Inverso da constante de elasticidade.

portanto, descrita por

$$Z_{\mathrm{rad}} = \left(\frac{1}{Z_{\mathrm{radEq1}}} + \frac{1}{\mathrm{j}\omega M_{\mathrm{a1}}} \right)^{-1} . \quad (8)$$

Massa mecânica

Segundo Rossi [6], a massa mecânica M_{m} [kg], com características indutivas, de um diafragma circular rígido fixado nas extremidades é equivalente a um quinto de sua massa total. Essa aproximação foi utilizada devido às condições de contorno impostas ao sistema no modelo numérico.

Resistência mecânica

R_{m} [Ns/m] refere às perdas do sistema, representada por um resistor. Esse parâmetro determina o Q_{m}, que caracteriza a largura de banda da ressonância, ou seja, o amortecimento do sistema. Entretanto, muitos transdutores comercializados possuem pequenas dimensões, o que pode inviabilizar a obtenção da resistência mecânica experimentalmente. Portanto, o parâmetro é definido por meio de ajustes do modelo analítico em relação ao modelo numérico validado[6]. A impedância R_{m}, nesse circuito, é o único componente exclusivamente resistivo. Com isso, ela abrange tanto as perdas relativas aos efeitos visco-térmicos acústicos [16], devido às pequenas dimensões da cavidade, como as perdas mecânicas e elétricas, devido à dispersão de energia no elemento piezoelétrico anisotrópico.

Compliância mecânica

Há estudos [12] para modelar sistemas bimorfos, como o encontrado no microfone piezoelétrico estudado, nos quais a membrana é composta de um disco de cerâmica piezoelétrica apoiado sobre um disco metálico de raio ligeiramente maior. Entretanto, a complexidade desses modelos dificulta sua aplicação nas analogias eletroacústicas. A estimativa da compliância mecânica C_{m} é realizada a partir do modelo FEM, de maneira baseada na medição de parâmetros Thiele-Small, mais especificamente, o método *delta-compliance* [17].

Função de transferência

A função de transferência $H(\mathrm{j}\omega)$ é então derivada, utilizando a Lei de malhas de Kirchhoff, por:

$$H(\mathrm{j}\omega) = \frac{S_d}{\gamma\, \mathrm{j}\omega C_{\mathrm{m}}} \left[Z_{\mathrm{eq}} + \frac{1}{\mathrm{j}\omega C_{\mathrm{m}}} + \frac{S_d^2}{\mathrm{j}\omega C_{\mathrm{a}}} \right]^{-1} , \quad (9)$$

sendo que

$$Z_{\mathrm{eq}} = R_{\mathrm{m}} + \mathrm{j}\omega M_{\mathrm{m}} + S_d^2 Z_{\mathrm{rad}} . \quad (10)$$

1.2 Modelo Numérico

O modelo numérico foi criado no *software* Comsol Multiphysics [18] que permite a modelagem das três físicas que compreendem o sistema de maneira integrada, sendo elas: acústica (incluindo perdas visco-térmicas devido às pequenas dimensões da cavidade) mecânica e elétrica. A geometria, exposta na Figura 1, é criada em um ambiente 2D axi-simétrico, isto é, uma revolução no eixo de simetria[7], que em conseguinte, gera a geometria total em 3 dimensões (3D). Três retângulos, representando a cavidade, a base metálica e o material piezoelétrico, são desenhados com as respectivas dimensões, vide Tabela 1. Foi selecionado, dentro da pasta de materiais piezoelétricos da biblioteca do programa o *Lead Zirconate Titanate* (PZT-5A) [13], que possui várias propriedades pré-definidas como densidade, matriz de acoplamento, permissividade relativa, fator de perda, dentre outros. As propriedades piezoelétricas são geralmente definidas de forma matricial, já que esse tipo de material geralmente apresenta comportamento anisotrópico[8]. O metal, ou latão segundo [13], teve suas propriedades definidas de acordo com a Tabela 1, que apresenta também as propriedades piezoelétricas relativas às direções do potencial elétrico e das tensões aplicadas no material consideradas na simplificação.

Tabela 1: Propriedades básicas do microfone piezoelétrico (retirado de [13]).

Propriedade	Símbolo	Valor	Unidade
Raio da cavidade	r_{cp}	5,25	mm
Espessura da cavidade	t_{cp}	0,15	mm
Raio do PZT	r_{p}	5,00	mm
Espessura do PZT	t_{p}	0,12	mm
Raio Latão	r_{M}	6,50	mm
Espessura Latão	t_{M}	0,12	mm
Módulo Young (Latão)	E_{M}	110,00	GPa
Cof. de Poisson (Latão)	ν_{M}	0,32	-
Densidade (Latão)	ρ_{M}	8000,00	kg/m^3
Fator de perda (Latão)	η_{M}	0,025	-
Módulo Young (PZT E_{33})	E_{p}	66,00	GPa
Cof. de Poisson (PZT)	ν_{p}	0,32	-
Densidade (PZT)	ρ_{p}	7600,00	kg/m^3
Coeficiente Piezoelétrico	d_{33}	374	pm/V
Capacitância (PZT) [19]	C_{e}	22	nF
Densidade do ar	ρ_0	1,21	kg/m^3
Velocidade do som	c_0	343	m/s

Para estimar a compliância do sistema utilizou-se de uma técnica inspirada em métodos de medição de parâmetros de alto-falantes [17]. Assim, a espessura da cavidade foi aumentada em 50%, de modo a causar um deslocamento da frequência de ressonância f_r. Nessa análise, utilizou-se de um passo de frequências de 1 Hz. Desse modo, pode-se estimar a compliância mecânica C_{m} de acordo com as equações:

$$2\pi f_r = \frac{1}{\sqrt{C_{\mathrm{t}} M_{\mathrm{m}}}} , \quad (11)$$

na qual C_{t} corresponde à compliância total do sis-

[6] Esse ajuste é realizado empiricamente. Uma alternativa a ser investigada futuramente é sua aproximação a partir do fator de qualidade.

[7] Representado pela linha pontilhada na Figura 1.

[8] As propriedades do material se diferem de acordo com a direção, ou seja, não são homogêneas para todo ele.

tema, comportando-se como uma equivalente em série das compliâncias acústica C_a (devido à cavidade) e mecânica C_m, ou seja:

$$\frac{1}{C_\mathrm{t}} = \frac{S_d^2}{C_\mathrm{a}} + \frac{1}{C_\mathrm{m}} \,. \tag{12}$$

2 RESULTADOS

Extraiu-se do modelo numérico, para manipulação em MATLAB, a sensibilidade em V/Pa do transdutor em função da frequência. Na Figura 4 tem-se a curva obtida com a geometria original, cujas propriedades se encontram na Tabela 1 e a curva relativa à geometria modificada, cuja única modificação se trata da cavidade com altura 50% maior. Percebe-se que a atenuação devido à influência dos efeitos viscotérmicos foi menor na cavidade com maiores dimensões. Foi detectado um deslocamento do pico ressonante de 195 Hz. Utilizando os valores de M_m calculados como um quinto da massa total e a Equação 11 para cada situação, foi obtido um valor de C_m = 11,46 µm/N, com um erro de 6% entre as duas geometrias analisadas[9].

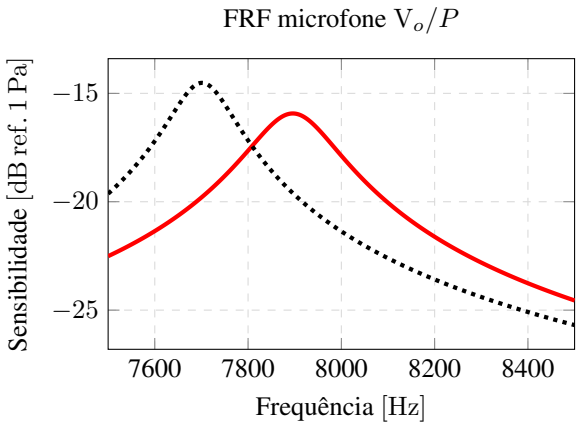


Figura 4: FRFs do transdutor piezoelétrico com a geometria original (——) e após a alteração da geometria (·····). Percebe-se um deslocamento de 195 Hz para baixo da frequência de ressonância do sistema após o aumento das dimensões da cavidade. Um passo de frequências de 1 Hz foi utilizado para aumentar a exatidão dos cálculos de C_m.

O valor de C_m médio, entre as duas curvas, foi então aplicado no modelo analítico de circuito equivalente desenvolvido e, com um ajuste da variável R_m[10], foram obtidas as curvas normalizadas pela sensibilidade em 1.000 Hz, expostas na Figura 5. As sensibilidades foram calculadas -21 dB (7,6 mV/Pa) no modelo analítico e -32 dB (0,6 mV/Pa) no modelo numérico. O ajuste de R_m deu-se apenas para modificar o valor máximo na ressonância, já que, como mostra a Equação 11, f_r não depende desse componente. De fato, ao se adicionar o valor de C_m médio calculado, as frequências de ressonância ficaram muito próximas. Para este caso, enquanto o modelo numérico leva cerca de 15 minutos para ser resolvido, o tempo de cálculo do modelo analítico é inferior a 0,25 segundos (para a mesma máquina de simulação).

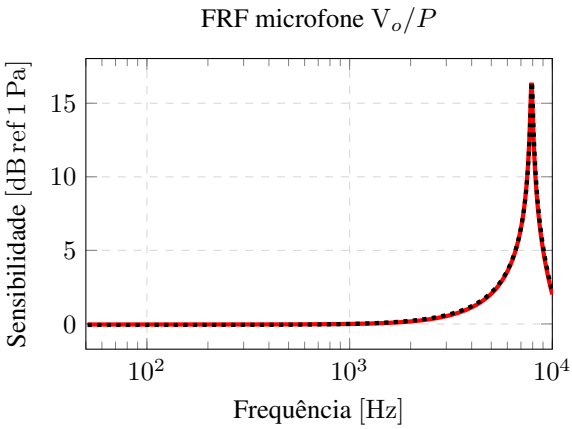


Figura 5: FRFs obtidas pelo modelo numérico (——) e pelo modelo analítico (·····) após o ajuste de R_m. Os modelos foram resolvidos com o mesmo passo de frequências a fim de garantir a correta representação da curva de sensibilidade.

3 CONSIDERAÇÕES FINAIS

Demonstrou-se a eficiência da técnica de ajuste do modelo analítico por parâmetros extraídos do modelo numérico, resultando em curvas de FRF normalizadas suficientemente equivalentes. Foram evidenciadas limitações na modelagem do acoplamento eletromecânico que causaram uma discrepância nas sensibilidades calculadas, o que deve ser estudado em trabalhos futuros. Para corrigir essas limitações, sugere-se estimar a capacitância do transdutor, bem como modelar apenas a conversão de energia mecânica, na forma de força, para energia elétrica, na forma de tensão elétrica. Uma possibilidade é a modelagem do diafragma livre, isto é, sem a inclusão de cavidades ou outros elementos acústicos.

Entretanto, com um tempo de cálculo muito inferior, o modelo analítico se mostra uma ferramenta importante, desde que devidamente ajustado. Propõe-se o desenvolvimento de técnicas para melhores estimativas da resistência mecânica, resistência por perdas viscotérmicas e das perdas elétricas. Um caminho é modelá-las separadamente em seus respectivos domínios e desenvolver técnicas de variação de cada uma isoladamente. Com isso, aumenta-se a robustez do modelo analítico e sua efetividade frente ao modelo FEM, o que sugere que o modelo com baixo custo computacional possui potencial para aplicação em técnicas de otimização. Como a compliância mecânica deve

[9] Erro possivelmente ligado à não inclusão da massa de radiação na Equação 11.

[10] Reforça-se que nesse modelo simplificado R_m representa todas as perdas do sistema, sendo elas acústicas ou mecânicas.

ser ajustada *a priori*, essas técnicas de otimização na etapa atual devem ser utilizadas apenas compreendendo variações no domínio acústico, variando a geometria da cavidade e/ou incluindo um *backplate* ou situações similares.

AGRADECIMENTOS

Os autores agradecem ao Mestre em Engenharia Israel Pereira pela Universidade Federal de Santa Catarina por gentilmente ceder o modelo FEM base do estudo.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Gustavo Martins, *Modelos Multi-físicos de Transdutores Piezoelétricos para Aparelhos Auditivos com Vistas à Aplicação de Técnicas de Otimização*, Tese de doutorado, Universidade Federal de Santa Catarina, Florianópolis, SC, Brasil, 2015.

[2] Gustavo Martins and Julio Cordioli, "Multiphysical finite element model of a hearing aid piezoelectric loudspeaker for optimization purposes," *Acta Acustica united with Acustica*, vol. 101, no. 5, pp. 993–1006, 2015.

[3] Mads Jakob Herring Jensen and Erling S. Olsen, "Virtual prototyping of condenser microphones using the finite element method for detailed electric, mechanic, and acoustic characterization," in *Proceedings of Meetings on Acoustics*. Acoustical Society of America, 2013, vol. 19, p. 030039.

[4] Leo L. Beranek and Tim Mellow, *Acoustics: sound fields and transducers*, Academic Press, 2012.

[5] Alexander Sadiku, *Fundamentals of electric circuits*, McGraw-Hill, New York, 2000.

[6] Mario Rossi, *Acoustics and electroacoustics*, Artech House Publishers, 1988.

[7] Stephen Horowitz, Toshikazu Nishida, Louis Cattafesta, and Mark Sheplak, "Development of a micromachined piezoelectric microphone for aeroacoustics applications," *The Journal of the Acoustical Society of America*, vol. 122, no. 6, pp. 3428–3436, 2007.

[8] W. Marshall Leach, *Introduction to electroacoustics and audio amplifier design*, Kendall/Hunt Publishing Company, 2003.

[9] Thomas Lavergne, Stéphane Durand, Michel Bruneau, Nicolas Joly, and Dominique Rodrigues, "Dynamic behavior of the circular membrane of an electrostatic microphone: Effect of holes in the backing electrode," *The Journal of the Acoustical Society of America*, vol. 128, no. 6, pp. 3459–3477, 2010.

[10] Brüel & Kjær, *Condenser microphones and microphone preamplifiers for acoustic measurements*, B&K, Denmak, 1982.

[11] Dorel Homentcovschi and Ronald N. Miles, "An analytical-numerical method for determining the mechanical response of a condenser microphone," *The Journal of the Acoustical Society of America*, vol. 130, no. 6, pp. 3698–3705, 2011.

[12] Valeriy Sharapov, Zhanna Sotula, and Larisa Kunickaya, *Piezo-Electric Electro-Acoustic Transducers*, Springer, 2014.

[13] Israel Pereira, *Caracterização Numérica e Experimental de um Protótipo de Microfone Piezoelétrico Visando Uso em Aparelhos Auditivos*, Dissertação de mestrado, Universidade Federal de Santa Catarina, Florianópolis, SC, Brasil, 2013.

[14] Richard H. Small, "Closed-box loudspeaker systems-part 1: analysis," *Journal of the Audio Engineering Society*, vol. 20, no. 10, pp. 798–808, 1972.

[15] Eric Brandão, *Apostila da dsiciplina de Eletroacústica I*, Engenharia Acústica, Universidade Federal de Santa Maria, 2014.

[16] Greg W. Swift, *Thermoacoustics*, Acoustical Society of America, 2002.

[17] Joseph A. D'Appolito, *Testing loudspeakers*, Audio Amateur Press, 1998.

[18] COMSOL Multiphysics, "www.comsol.com," Feb. 2016.

[19] APC International Ltd. - Piezo calculator, "www.americanpiezo.com/knowledge-center/apc-piezo-calc.html," Feb. 2016.

Sociedade de Engenharia de Áudio
Artigo de Congresso

Apresentado no 14º Congresso de Engenharia de Áudio
20ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2016, São Paulo, SP

Este artigo foi reproduzido do original final entregue pelo autor, sem edições, correções ou considerações feitas pelo comitê técnico. A AES Brasil não se responsabiliza pelo conteúdo. Outros artigos podem ser adquiridos através da Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA, www.aes.org. Informações sobre a seção Brasileira podem ser obtidas em www.aesbrasil.org. Todos os direitos são reservados. Não é permitida a reprodução total ou parcial deste artigo sem autorização expressa da AES Brasil.

GuitarAMI: Desenvolvimento, Implementação e Performance de um Instrumento Musical Aumentado

Eduardo Aparecido Lopes Meneses,[1] José Eduardo Fornari Novo Júnior[1] e Marcelo Mortensen Wanderley[2]

[1] Universidade Estadual de Campinas (UNICAMP), Instituto de Artes, Núcleo Interdisciplinar de Comunicação Sonora (NICS)
Campinas, SP, 13083-872, Brasil

[2] McGill University, Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT)
Montreal, Quebec, H3A 0G4, Canada

eduardo@edumeneses.com, tutifornari@gmail.com, marcelo.wanderley@mcgill.ca

RESUMO

Este artigo relata a construção e o desenvolvimento de um instrumento musical aumentado intitulado GuitarAMI cujo início da pesquisa foi descrito em [1]. O GuitarAMI foi construído utilizando um violão e explorando algumas possibilidades de modificação de características intrínsecas e comumente consideradas restritivas no instrumento através da execução de algoritmos de manipulação sonora em conjunto com processadores de efeitos. Apresentamos aqui três protótipos do GuitarAMI que foram construídos e avaliados em sessões de improvisação livre executadas pelo grupo musical B.E.A.T. e por estudantes do um curso de música e tecnologia ministrado no *Programa Guri*.

0 INTRODUÇÃO

A construção de instrumentos musicais aumentados ou AMIs[1] têm se mostrado um campo promissor de desenvolvimento técnico e artístico, com literatura consideravelmente ampla e incluindo várias publicações nacionais e internacionais [2].

Projetar AMIs que utilizem instrumentos musicais tradicionais em sua construção envolve a escolha de instrumentos que permitam ao intérprete a exploração de sua *spare bandwidth* [3], ou seja, a disponibilidade individual de cada instrumentista em executar movimentos além dos já usualmente realizados durante a execução do instrumento e diretamente ligados à produção sonora, também classificados de gestos livres [4], bem como quaisquer outros movimentos que possam ser convertidos em dados de controle e utilizados para comandar processos de síntese e manipulação sonora.

[1] Sigla de *Augmented Musical Instruments*

A criação de AMIs utilizando guitarras elétricas tem algumas facilidades como por exemplo os captadores, já presentes no instrumento, além do uso regular de processos de manipulação sonora através dos pedais de efeitos. Alguns exemplos de AMIs construídos com a guitarra elétrica incluem a *Talking Guitar* [5], a *Mobile Wireless Augmented Guitar* [6], a *Augmentalist* [7] e a *Multimodal Guitar* [8].

O violão também oferece algumas opções para captura de áudio, porém a quantidade de AMIs utilizando violão é consideravelmente menor. Exemplos interessantes incluem a *SmartGuitar*[2], onde os desenvolvedores usam atuadores para modificar o som acústico do instrumento e a *Carolan*, um violão que utiliza códigos digitais dentro dos padrões decorativos do instrumento para criar uma *pegada digital*[3] que pode ser seguida na Internet [9].

Este artigo relata a construção de um AMI usando o violão e explora algumas possibilidades de modificação sonora. Serão descritos na seção 1 as diferenças e particularidades deste AMI construído em relação aos dispositivos já citados.

1 GUITARAMI

Um dos principais objetivos no AMI desenvolvido durante este trabalho, intitulado GuitarAMI, foi a construção de um instrumento musical aumentado que explora algumas possibilidades de modificação das características intrínsecas usualmente consideradas restritivas no violão através da utilização de sensores e algoritmos de manipulação sonora em conjunto com processadores de efeitos normalmente aplicados em performances de guitarra elétrica.

Para o GuitarAMI foram considerados não apenas a *spare bandwidth* [1] possível para os violonistas, mas também as características físicas do instrumento utilizado, o mapeamento entre os dados gestuais captados pelo dispositivo e a sua conversão em dados de controle para o algoritmo responsável pelo processamento sonoro.

Como muitos instrumentos de cordas pinçadas, o violão tem uma grande variedade timbrística, porém outras características como a pouca sustentação das notas musicais emitidas ou a impossibilidade de controle da intensidade sonora após o ataque são geralmente consideradas severas restrições impostas pela estrutura física do instrumento.

Uma das possibilidades encontradas foi a utilização de algoritmos para modificar as restrições descritas acima. Para esta tarefa, criamos um *patch* programado com a linguagem *Pure Data* (PD) que recebe o som captado do violão e executa um algoritmo programado para sustentar um elemento sonoro registado pelo computador ou controlar o envelope dinâmico deste mesmo elemento em tempo real [10].

[2] http://instrum.ircam.fr/smartinstruments/.
[3] Digital footprint.

A construção do dispositivo foi organizada de acordo com as seguintes fases:

- Classificação e seleção dos movimentos utilizados;
- Aquisição de dados gestuais destes movimentos;
- Construção do protótipo do GuitarAMI;
- Avaliação do protótipo em performance;
- Atualização ou reconstrução dos protótipos seguido de nova avaliação.

Para cada fase da criação e avaliação do GuitarAMI um protótipo foi construído e testado em performances, além de um curso ministrado no Programa Guri[4], conforme descrito na seção 1.4. A avaliação realizada pelos instrumentistas e pelos alunos guiou o desenvolvimento de três protótipos construídos durante a pesquisa, conforme descrito na seção 1.1.

1.1 Seleção e Classificação Gestual

Há várias definições de gesto na literatura acadêmica e muitas delas são complementares entre si em certos contextos [4, p. 74]. Neste artigo usamos a definição de gesto instrumental proposta por Claude Cadoz e Marcelo Wanderley em [4].

Os gestos instrumentais, de acordo com Cadoz e Wanderley, podem ser classificados em efetivos, acompanhantes ou figurativos. Ao contrário dos gestos instrumentais também devemos considerar o gesto livre (*geste à nu*) que, de acordo com os autores, não se qualifica como um gesto instrumental porque não há interação física entre músico e instrumento. Gestos livres podem ser utilizados, no entanto, para controle dos parâmetros de síntese e manipulação sonora através de sensores e câmaras.

Com esta classificação em mente planejou-se a construção do GuitarAMI utilizando possibilidades gestuais não conflitantes com as ações necessárias para a prática instrumental. Foram escolhidos como movimentos passíveis de captura a possibilidade que o instrumentista tem de mover o instrumento durante a performance, distância entre seu corpo e o instrumento ou a posição espacial da mão do violão.

1.2 Aquisição Gestual

Para coletar dados de movimentação do instrumento dois sensores foram utilizados no violão: O acelerômetro ADXL345 e o sensor ultrassônico HC-SR04. Ambos foram montados num invólucro preso à faixa do instrumento musical de forma não invasiva.

[4] O Programa Guri é um programa de educação musical começou em 2008 a partir de uma parceria entre a Secretaria do Estado de são Paulo Estado de Cultura e Santa Marcelina - Organização social da Cultura (http://www.gurisantamarcelina.org.br/) [11].

O acelerômetro escolhido é capaz de medir aceleração em 3 eixos e possui características que permitem a sua utilização para a aquisição de vários movimentos simultaneamente. Devido às características físicas do sensor e o ângulo de montagem do invólucro no instrumento, é necessária a sua calibração de acordo com a posição do sensor e a postura do instrumentista após a montagem no violão.

O sensor ultrassônico foi utilizado para medir a distância entre o instrumentista e o instrumento ou para recolher dados gerados por gestos livres.

Também foram utilizados interruptores momentâneos SPST montados em um pedal para explorar gestos livres acionados pelos pés do instrumentista. Estes interruptores têm funções de controle secundárias, mas podem eventualmente controlar os parâmetros de produção sonora.

Estes mesmos interruptores também são utilizados para ativar ou desativar algoritmos específicos descritos na seção 1.5 e oferecem possibilidades para o controle de algoritmos acrescentados pelo intérprete ou compositor, como descrito na seção 2.

1.3 Construção do GuitarAMI

Como ponto de partida na construção do GuitarAMi foi utilizado o modelo de instrumento musical digital (ou DMI[5]) criado por Mark Marshall a partir da análise dos modelos propostos por Bert Bongers, David Birnbaum, Marcelo Wanderley e Perry Cook [12]. Partindo deste modelo foram realizados ajustes para a criação de um AMI utilizando o violão.

Com toda a informação recolhida foi possível criar um diagrama de construção para o primeiro protótipo do GuitarAMI e definir os componentes necessários para a tarefa.

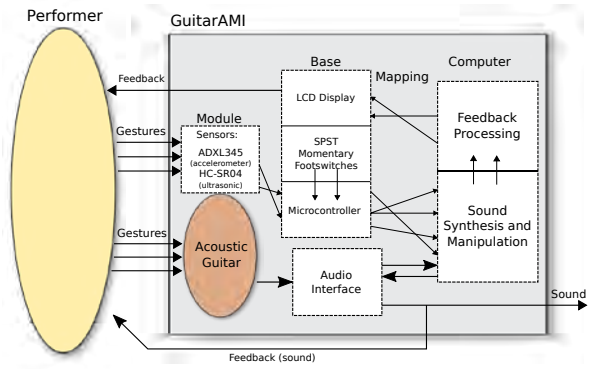


Figura 1: Diagrama de construção do GuitarAMI (em inglês), criado à partir do modelo de DMI, segundo Mark Marshall.

O sistema de aquisição de dados gestual é composto por um Arduino Leonardo[6] que recebe informações dos sensores fixados no violão. Um *display* LCD foi adicionado para fornecer *feedback* visual ao instrumentista durante a performance.

Um dos objetivos na construção do GuitarAMI desde o início desta pesquisa foi construir um dispositivo que pudesse ser utilizado com a maioria dos violões de forma não invasiva. Considerando que cada instrumento é único e o instrumentista estabelece uma forte ligação afetiva com seu próprio instrumento [13], é importante que o GuitarAMI possa ser montado no instrumento de preferência do músico.

A escolha dos componentes do GuitarAMI também considerou a facilidade de programação, bem como uma extensa documentação e suporte [14]. Optamos, portanto, por utilizar o Arduino e o Pure Data (PD) no dispositivo.

1.4 Protótipos do GuitarAMI

1.4.1 Primeiro protótipo

De acordo com o diagrama de construção mostrado na Figura 1 e as informações relatadas na seção 1.3, foi possível construir o primeiro protótipo do GuitarAMI para uso em apresentações musicais, entre agosto de 2014 e junho de 2015, conforme relatado em [1]. Como visto na Figura 1, o GuitarAMI é composto por quatro elementos além do violão: base, computador, interface de áudio e módulo.

O módulo do GuitarAMI, ligado por fio à base, é responsável pela aquisição de movimentos previamente selecionados, conforme descrito na seção 1.1 e enviados para a base A base do GuitarAMI é responsável pelo primeiro processamento destes dados e seu envio ao computador. A interface de áudio envia o som do violão captado para o computador, o qual é responsável pela execução de algoritmos programados em PD, além de mapeamento de dados gestuais para estes algoritmos.

Este protótipo foi utilizado durante o 1º Colóquio Franco-Brasileiro de Análise e Criação Musicais com Suporte Computacional [1, 11] ocorrido em agosto de 2014 no NICS[7] e em um *workshop* para estudantes de música do *Programa Guri* que ocorreu entre o segundo semestre de 2014 e o primeiro semestre de 2015, relatado em [15].

1.4.2 Segundo protótipo

Com as informações de dados e *feedback* fornecidos pelos instrumentistas durante a utilização do primeiro protótipo do GuitarAMI, foi possível viabilizar a construção o segundo protótipo do dispositivo.

Um problema relatado durante a utilização do primeiro protótipo foi a conexão com fio entre a base e módulo do GuitarAMI, dificultando movimentação do violão para controle de processos de síntese e manipulação sonora nos algoritmos. A ligação por fio também alterava desintencionalmente a posição do módulo do

[5] Sigla de Digital Musical Instrument.
[6] https://www.arduino.cc/.
[7] Núcleo Interdisciplinar de Comunicação Sonora da Unicamp.

Figura 2: Primeiro protótipo do GuitarAMI, construído com um Arduino Leonardo, um sensor ultrassônico e acelerômetro [1]. O segundo protótipo é fisicamente idêntico ao primeiro, com exceção do módulo do GuitarAMI que foi modificado para funcionamento sem fio.

GuitarAMI, fixado na faixa do violão, causando problemas de calibração no acelerômetro durante a performance. Este problema foi corrigido com a adição de um segundo Arduino no módulo do GuitarAMI. Um Arduino Nano e um invólucro para baterias recarregáveis também foram instalados, transformando o módulo do GuitarAMI em um dispositivo sem fio.

Algumas mudanças também foram realizadas no modo como as informações eram apresentadas no display LCD, acrescentando informações sobre os algoritmos executados pelo computador durante cada estágio da performance.

O segundo protótipo foi utilizado em performances artísticas e num curso ministrado em uma das unidades de ensino de Programa Guri na cidade de Osasco, durante o segundo semestre de 2015. Este curso teve 11 encontros de uma hora de duração e contou com 15 alunos que realizavam atividades de pesquisa e prática sobre estética, técnicas e procedimentos desenvolvidos ao longo do século XX.

Em uma performance do trio B.E.A.T.[8], o GuitarAMI foi utilizado na execução de um *patch* de PD para envio de mensagens MIDI para o *software* Integra Live[9] utilizados em algoritmos de síntese sonora controlados dinamicamente pelo GuitarAMI. Esta performance pode ser assistida no *link*: https://youtu.be/e2QuWLH87pg na sua versão completa ou em https://youtu.be/dUd-1i0h104 para uma versão curta.

No curso ministrado no *Programa Guri - Polo Osasco*, estudantes de música foram capazes de utilizar a GuitarAMI para acionar e controlar eventos de som previamente amostrados. A demonstração rápida do segundo protótipo pode ser vista no endereço https://youtu.be/0mrw3Kwb7UU.

1.4.3 Terceiro protótipo

Os problemas relatados durante a utilização do segundo protótipo do GuitarAMI podemos destacar problemas na montagem da GuitarAMI em performances e a dificuldade em controlar algoritmos simultaneamente executados pelo PD e processos de manipulação sonora escolhidos pelo instrumentista e executadas pelo Integra Live.

O terceiro protótipo foi montado sobre um único *case* para facilitar a montagem do dispositivo, juntamente com a utilização de um processador multi-efeitos e um *hub USB*, que permitiu uma conexão rápida entre a base do GuitarAMI, a interface de áudio e o computador, como pode ser observado na figura 3.

Figura 3: Terceiro protótipo do GuitarAMI, com conexão sem fio entre base e module, além da adição de um pedal multiefeitos.

No terceiro protótipo do GuitarAMI o processamento de áudio foi dividido entre um processador multi-efeitos e os algoritmos executados pelo computador. Enquanto o processador multiefeitos é responsável por manipulações de som tradicionais, como distorção, *delay*, *reverb*, *whah whah*, equalizador, entre outros; o computador que executa o *patch* (algoritmo) em PD realiza o mapeamento gestual e controle de algoritmos predefinidos (descritos na seção 1.5), além de oferecer a possibilidade de utilização de novos algoritmos programados instrumentistas ou compositores.

1.5 Características sonoras do violão e seu processamento

Como parte integrante do GuitarAMI tem-se um *patch* programando em PD, incluindo o mapeamento

[8] O B.E.A.T. (*Brazilian Electronic Aleatorium Trio*) é um grupo formado pelos músicos Walmir Gil (trompete), Alê Damasceno (bateria) e Edu Meneses (violão) com a proposta de explorar possibilidades de livre improvisação utilizando recursos eletrônicos e instrumentos acústicos.

[9] Integra Live é um programa de computador que utiliza a linguagem de programação PD para o processamento de áudio interativo, mantido pelo Integra Lab no Conservatório de Birmingham, no Reino Unido (http://www.integralive.org/).

gestual e dois algoritmos de manipulação sonora programados para alterar algumas características intrínsecas no violão.

Para possibilitar ao violão a emissão de um som que possa ser sustentado indefinidamente utilizamos a Transformada Rápida de Fourier (FFT) para processar uma janela no domínio do tempo de sinal de áudio e o resultado desta análise é sintetizado num novo som através da operação inversa da FFT [16]. Este algoritmo é comumente chamado *Spectral Freeze* e sua implementação no GuitarAMI funciona como um *patch* programado em PD intitulado *Freeze.pd* [1].

Os parâmetros disponíveis para controle desta síntese sonora são: acionamento de captura de áudio instantâneo, controle de tempo de duração do áudio sintetizado e controle de volume (intensidade sonora) deste mesmo som. O instrumentista controla esses parâmetros com dois interruptores disponíveis na base do GuitarAMI e configurados para iniciar e parar a síntese sonora, bem como receber os dados do acelerômetro enviados pelo módulo GuitarAMI e mapeá-los para a função de controle da intensidade sonora.

É desejável que o controle de intensidade sonora possa ser utilizado independentemente da síntese FFT, permitindo manter um som sustentado indefinidamente simultaneamente a outros sons de intensidade sonora controlada livremente pelo instrumentista.

Para tornar possível ao instrumentista controlar a intensidade sonora independentemente da utilização do *patch Freeze.pd*, programamos um *patch* especificamente responsável por modelar o envelope dinâmico do som emitido pelo violão em tempo real, chamado *Envshaper.pd* [10].

Esta configuração apresenta ao executante um instrumento pronto para ser utilizado, além de oferecer a possibilidade de execução de novos algoritmos para GuitarAMI e mapeamento dos dados gestuais recebidos.

2 RECURSOS IMPROVISACIONAIS

A principal configuração de GuitarAMI oferece funcionalidade suficiente para, após um período de adaptação, fornecer ferramentas para utilização em performance de improvisação. Os instrumentistas que utilizaram o GuitarAMI relataram a necessidade de um período de adaptação para atingir uma resposta aceitável entre uma idéia improvisatória contendo elementos eletroacústicos e sua execução ao instrumento. Como os instrumentistas não estão habituados a movimentar o instrumento para longe de si mesmos ou modificar a inclinação do braço do violão durante a performance (movimentando a mão do instrumento) com o intuito de controlar determinados parâmetros da geração sonora, foi necessário um período de adaptação para que eles pudessem usar esses gestos com proficiência.

É importante frisar a necessidade de configurar os algoritmos no PD assim como o processamento de áudio no processador multi-efeitos, organizando-os de modo a permitir o acesso rápido para o instrumentista. Esta configuração é pessoal e varia de acordo com o desejo do instrumentista.

Esta característica está diretamente relacionada com a suposta dificuldade de se estabelecer um *modus operandi* único do instrumentista [17]. Enquanto as possibilidades de utilização de recursos de tecnologia acessíveis está cada vez mais ampla, tanto técnica e financeiramente, somos levados a um cenário artístico onde performances musicais são extremamente pessoais, ao contrário da abordagem tradicional, onde o artista trabalha com regras fixas que podem ser flexibilizadas no processo de criação e performance.

Ao mesmo tempo, os instrumentistas desejam ser tecnicamente desafiados pelos instrumentos e, consequentemente, por AMIs [13]. Fornecendo um modo predefinido para GuitarAMI é possível instigar o instrumentista a dominar as configurações padrão do GuitarAMI e, simultaneamente, proporcionar a possibilidade de personalizar o instrumento.

A funcionalidade de adicionar novos algoritmos para o GuitarAMI abre várias possibilidades de composição. Estas possibilidades ainda não foram exploradas na prática, mas a hipótese é que o compositor possa também explorar os dados recebidos dos sensores do GuitarAMI em suas composições.

3 DISCUSSÕES

Durante performances utilizando o GuitarAMI os instrumentistas foram capazes de mudar o envelope dinâmico ou sustentar indefinidamente sons emitidos pelo instrumento, enquanto eles estão dentro da faixa sonora que pode ser captada pela interface de áudio utilizada.

Timing é de fundamental importância na utilização do *patch Freeze.pd*. A qualidade do som sintetizado pelo *patch* depende do momento em que o instrumentista captura a amostra sonora. Ao realizar a captura em um momento muito próximo ao ataque, é esperado que o o som sintetizado reproduza vários transientes do som capturado, acrescentando assim ruído ao som sustentado. Por isso, o momento ideal para gravar a janela de captura para que o som sintetizado possua o mínimo de transientes acrescentados ao som original parece ser alguns milisegundos após o ataque, quando grande parte do ruído transiente (estocástico) perde a energia e o algoritmo sintetiza um som com maior presença de componentes periódicas (determinísticas) do som do violão [18]. Estes mesmos transientes podem ser utilizados criativamente pelo instrumentista para síntese de sons mais complexos que os sons originalmente adquiridos pelo sistema de captação utilizado no GuitarAMI.

A síntese sonora realizada pelo *patch Freeze.pd* torna-se mais distante do som original à medida que a complexidade sonora aumenta. Acordes e sons complexos ricos em harmônicos podem gerar um áudio sintetizado com características muito diferentes em comparação à fonte sonora original. Este recurso pode ser

utilizado de uma forma expressiva durante uma performance, no entanto este não é o objetivo original do *patch*, que foi projetado para produzir um momento de som que é morfologicamente equivalente ao som original, porém com duração indeterminada.

O controle de intensidade do som sintetizado no *patch Freeze.pd* exige posicionamento preciso do violão. Os parâmetros de sensibilidade do acelerômetro devem ser calibrados de acordo com o instrumentista, antes de iniciar a performance, para fornecer uma resposta consistente com a intenção estética deste.

O *patch env-shaper.pd*, apesar de sua simplicidade, requer alguma prática do instrumentista para de produzir o resultado almejado durante a performance. É possível recriar em tempo real sons usualmente utilizados em música acusmática e que anteriormente só eram possíveis de serem capturados com a inversão do envelope sonoro, no domínio do tempo, realizada apenas em tempo diferido.

4 CONCLUSÕES E TRABALHOS FUTUROS

Este artigo apresentou a construção e a utilização de três protótipos de um instrumento musical aumentado intitulado GuitarAMI, que fornece ao instrumentista a capacidade de controlar algumas das características sonoras intrínsecas consideradas restritivas no violão, bem como utilizar processos de áudio normalmente usados na guitarra elétrica e criar novos processos de síntese e manipulação sonora através de *patches* programados em Pure Data. Estas são características importantes que tornam este AMI uma interessante ferramenta para performance e improvisação.

O GuitarAMI apresenta uma série de possibilidades de desenvolvimentos e futuras expansões. Uma possibilidade é a definição de diferentes parametrizações dos algoritmos utilizados nos *patches*. O *patch Freeze.pd* possui diversos parâmetros passíveis de alteração do resultado audível. Mais estudos serão necessários para obter uma síntese de sonoridades psicoacusticamente mais próximas às do som do violão.

A criação de um *patch* projetado especificamente para interligar os parâmetros de controle dos *patches* do GuitarAMI a qualquer outro algoritmo desenvolvido por um instrumentista ou compositor pode contribuir para ampliar as possibilidades de utilização desta plataforma em estudos e desenvolvimentos futuros.

5 AGRADECIMENTOS

Os autores agradecem à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo apoio financeiro concedido a esta pesquisa e ao NICS (Núcleo Interdisciplinar de Comunicação Sonora) da Unicamp (Universidade Estadual de Campinas) por fornecer a infraestrutura onde foi desenvolvida esta pesquisa.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Eduardo Aparecido Lopes Meneses, José Fornari, and Marcelo M. Wanderley, "Projeto e construção de uma interface gestual para improvisação livre," in *Anais do 13º Congresso de Engenharia de Áudio da AES Brasil*. AES – Audio Engineering Society – Brazil Section, maio 2015, pp. 75–81.

[2] Eduardo R. Miranda and Marcelo M. Wanderley, *New Digital Musical Instruments: Control and Interaction Beyond the Keyboard*, Computer Music and Digital Audio Series. A-R Editions, Inc., 2006.

[3] P Cook, "Principles for designing computer music controllers," *Conference on New Interfaces for Musical Expression (NIME-01)*, pp. 1–4, 2001.

[4] Claude Cadoz and Marcelo M. Wanderley, "Gesture-music," in *Trends in Gestural Control of Music*, Marcelo M. Wanderley and Marc Battier, Eds., pp. 71–94. Editions IRCAM – Centre Pompidou, Paris, 2000.

[5] Liam Donovan and Andrew McPherson, "The talking guitar: Headstock tracking and mapping strategies," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, London, United Kingdom, 2014, pp. 351–354, Goldsmiths, University of London.

[6] Nicolas Bouillot, Mike Wozniewski, Zack Settel, and Jeremy R. Cooperstock, "A mobile wireless augmented guitar," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, Antonio Camurri, Gualtiero Volpe, and Stefania Serafin, Eds., Genoa, Italy, 2008, pp. 189–192.

[7] Dan Newton and Mark T. Marshall, "Examining how musicians create augmented musical instruments," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, Oslo, Norway, 2011, pp. 155–160.

[8] Loïc Reboursière, Christian Frisson, Otso Lähdeoja, John A. Mills, Cécile Picard-Limpens, and Todor Todoroff, "Multimodal guitar : A toolbox for augmented guitar performances," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, Sydney, Australia, 2010, pp. 415–418.

[9] Steve Benford, Adrian Hazzard, Alan Chamberlain, and Liming Xu, "Augmenting a guitar with its digital footprint," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, Edgar Berdahl and Jesse Allison, Eds., Baton Rouge, Louisiana, USA, May 31 – June 3 2015, pp. 303–306, Louisiana State University.

[10] Eduardo Aparecido Lopes Meneses and José Fornari, "Guitarami: um instrumento musical aumentado que transpõe restrições intrínsecas do violão," in *Proceedings of the 15th Brazilian Symposium on Computer Music*, Campinas/SP, 2015, Sociedade Brasileira de Computação - SBC.

[11] Eduardo Aparecido Lopes Meneses, José Fornari, and Marcelo M. Wanderley, "A study with hyperinstruments in free musical improvisation," in *Anais do XI Simpósio Internacional de Cognição e Artes Musicais*, Pirenópolis/GO, maio 2015, Associação Brasileira de Cognição e Artes Musicais (ABCM).

[12] Mark T. Marshall, *Physical Interface Design for Digital Musical Instruments*, Ph.D. thesis, McGill University, Montreal, 2008.

[13] Thor Magnusson and Enrike H. Mendieta, "The acoustic, the digital and the body : A survey on musical instruments," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, New York City, NY, United States, 2007, pp. 94–99.

[14] PD-Graz Group, *Bang*, Wolke Verlag, 2006.

[15] Eduardo Aparecido Lopes Meneses and José Fornari, "Educação musical através da improvisação livre com recursos computacionais: contribuições e desafios," in *Anais do XXV Congresso da Associação Nacional de Pesquisa e Pós-Graduação em Música*, Vitória/ES, 2015, Associação Nacional de Pesquisa e Pós-Graduação em Música – ANPPOM.

[16] Gabriel Vigliensoni Martin, "Final project report: Real-time Sound Freezing and Spectral Shifting," Tech. Rep., 1986.

[17] José Fornari, "Arte , Afeto e Artefato," *DAPesquisa*, vol. vol. 7, pp. 418–426, 2011.

[18] Xavier Serra, *A System for Sound Analysis/Transformation/Synthesis based on a Deterministic plus Stochastic Decomposition*, Ph.D. thesis, Stanford University, 1989.

Sociedade de Engenharia de Áudio
Artigo de Congresso

Apresentado no 14º Congresso de Engenharia de Áudio
20ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2016, São Paulo, SP

Este artigo foi reproduzido do original final entregue pelo autor, sem edições, correções ou considerações feitas pelo comitê técnico. A AES Brasil não se responsabiliza pelo conteúdo. Outros artigos podem ser adquiridos através da Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA, www.aes.org. Informações sobre a seção Brasileira podem ser obtidas em www.aesbrasil.org. Todos os direitos são reservados. Não é permitida a reprodução total ou parcial deste artigo sem autorização expressa da AES Brasil.

Controle automático de *loudness* em conteúdo de formato curto para radiodifusão

Leandro da Silva Pires,[1,2] Maurílio Nunes Vieira[3] e Hani Camille Yehia[3]

[1] Programa de Pós-Graduação em Engenharia Elétrica - Universidade Federal de Minas Gerais
Av. Antônio Carlos 6627, 31270-901, Belo Horizonte, MG, Brasil

[2] Gerência Regional da Agência Nacional de Telecomunicações em Minas Gerais
Rua Maranhão 166, 30150-330, Belo Horizonte, MG, Brasil

[3] CEFALA - Centro de Estudos da Fala, Acústica, Linguagem e Música
UFMG-DELT: Av. Antônio Carlos 6627, 31270-901, Belo Horizonte, MG, Brasil

leandropires@ufmg.br, maurilionunesv@ufmg.br, hani@cefala.org

RESUMO

Nestes primeiros anos de vigência da Recomendação ITU-R BS.1770-2 (2011) para o controle da intensidade subjetiva de áudio (*loudness*) na radiodifusão, verificou-se que seus descritores básicos são insuficientes para conferir conformidade a conteúdos de formato curto tais como comerciais, anúncios e inserções ao vivo. Este trabalho propõe uma solução automática de controle de *loudness* na cadeia de sinal do ambiente de entrega de conteúdo (*playout*) à luz do Suplemento nº 1 da Recomendação EBU R.128 (2014). A proposta consiste tanto na detecção de blocos comerciais por métodos de inteligência computacional baseados em redução de dimensionalidade e em classificação de margem larga via Máquinas de Vetor de Suporte (*Support Vector Machines* - SVMs), quanto no controle de faixa dinâmica do sinal com o uso de técnicas de processamento baseadas em integrações de *loudness* próprias para segmentos curtos e em medições de pico verdadeiro.

0 INTRODUÇÃO

O controle da intensidade subjetiva de áudio, ou *loudness*, talvez seja um dos tópicos mais relevantes em áudio para radiodifusão nos últimos anos. Gravações musicais passadas parecem significativamente mais suaves que as modernas e os blocos de propaganda são percebidos bem mais fortes que os de programação. A compressão excessiva de faixa dinâmica, feita com o propósito de que uma determinada peça de áudio pareça mais forte do que suas concorrentes, tem incomodado progressivamente a audiência. A elevação injustificável de volume nos intervalos comerciais da radiodifusão

com tecnologia digital é considerada uma infração à legislação brasileira [1]. Porém somente com os algoritmos de medição de *loudness* na programação de radiodifusão [2] [3] foi possível medi-la objetivamente. Estas recomendações formam a base do regulamento de padronização do volume de áudio nos intervalos comerciais da radiodifusão [4].

1 APRESENTAÇÃO DO PROBLEMA

1.1 Compressão de Faixa Dinâmica

A compressão 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 (fitas, películas e rádios AM/FM) [5], é um efeito frequentemente usado para fazer com que a trilha de áudio soe mais forte 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 quantizador de um dispositivo de reprodução.

1.2 Ferramenta de Produção

Com o advento da 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 [6]. Graças a 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*) [7].

1.3 Mercado Publicitário

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 [8]. Na radiodifusão aberta, por exemplo, é comum o telespectador assistir 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.

1.4 Motivação

De modo a mitigar o problema exposto anteriormente, estes algoritmos perceptivos de medida e normalização foram embarcados aos poucos em soluções de processamento de áudio voltadas à Qualidade de Experiência (QoE) na radiodifusão digital. Dito isso, (i) é possível depreender que os *presets* de configuração de um processador de áudio para atuação nos blocos de programação devem ser distintos dos aplicados ao processamento dos blocos de propaganda e vinhetas, (ii) não é seguro confiar integralmente numa normalização prévia do conteúdo devido ao caráter dinâmico da programação, a exemplo das chamadas ao vivo, das matérias produzidas de forma expressa e de outras inserções nas quais não há tempo hábil para o refino no tratamento do áudio e (iii) a inviabilidade de um gatilho manual para o chaveamento da atuação do processador aponta para a necessidade de métodos de detecção de blocos comerciais baseados em algoritmos de inteligência computacional [9].

2 OBJETIVO

Conforme o acompanhamento da adoção prática dos parâmetros de *loudness* constantes das Recomendações [2] e [3] (e da legislação brasileira), no que tange ao conteúdo de formato curto, os próprios organismos de padronização verificaram que [10]:

- O parâmetro "Faixa de *Loudness*" [11] 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" [12] para controle de peças altamente comprimidas.

O objetivo deste trabalho é propor um esquema de controle de conteúdo de formato curto diretamente na cadeia de sinal do *playout* com base nestas novas recomendações que, embora ainda não abarcadas pela regulamentação brasileira, têm impacto na QoE. O trabalho vislumbra 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.

3 DESCRIÇÃO DAS TÉCNICAS EMPREGADAS

3.1 Análise de Componentes Principais

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 [13]. 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. Vamos à sua definição matemática [13] [14].

Dada uma variável aleatória multidimensional:

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} \tag{1}$$

Sua média é dada por:

$$\mathbf{m}_x = E[\mathbf{X}] = \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_N] \end{bmatrix} \tag{2}$$

E sua matriz de covariância é dada por:

$$\begin{aligned} \Sigma_x &= E\left[(\mathbf{X}-\mathbf{m}_x)^T(\mathbf{X}-\mathbf{m}_x)\right] \\ &= E\left[(\mathbf{X}_o)^T(\mathbf{X}_o)\right] \\ &= \begin{bmatrix} E[X_{o1}^2] & E[X_{o1}X_{o2}] & \cdots & E[X_{o1}X_{oN}] \\ E[X_{o2}X_{o1}] & E[X_{o2}^2] & \cdots & E[X_{o2}X_{oN}] \\ \vdots & \vdots & \ddots & \vdots \\ E[X_{oN}X_{o1}] & E[X_{oN}X_{o2}] & \dots & E[X_{oN}^2] \end{bmatrix} \end{aligned} \tag{3}$$

Σ_x é uma matriz positiva definida que pode ser decomposta em valores singulares (*Singular Value Decomposition - SVD*).

$$\Sigma_x = \boldsymbol{USU}^T \tag{4}$$

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

$$\boldsymbol{UU}^T = \boldsymbol{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \tag{5}$$

Agora se definirmos:

$$\mathbf{Y} = \boldsymbol{U}^T(\mathbf{X}-\mathbf{m}_x) \tag{6}$$

Pela Equação (3), teremos:

$$\begin{aligned} \Sigma_y &= E\left[\boldsymbol{U}^T(\mathbf{X}-\mathbf{m}_x)(\mathbf{X}-\mathbf{m}_x)^T\boldsymbol{U}\right] \\ &= \boldsymbol{U}^T E\left[(\mathbf{X}-\mathbf{m}_x)(\mathbf{X}-\mathbf{m}_x)^T\right]\boldsymbol{U} \\ &= \boldsymbol{U}^T\Sigma_x\boldsymbol{U} \\ &= \boldsymbol{U}^T\boldsymbol{USU}^T\boldsymbol{U} \\ &= \boldsymbol{ISI} \\ &= \boldsymbol{S} \end{aligned} \tag{7}$$

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

$$\mathbf{X} = \boldsymbol{U}\mathbf{Y} + \mathbf{m}_x \tag{8}$$

Assim sendo, as Equações (6) e (8) definem a representação de $\mathbf{X}$ em componentes principais e sua inversa.

3.2 Máquinas de Vetor de Suporte

Paralelamente à regressão logística, aos sistemas de inferência nebulosa e às redes neuronais artificiais, a máquina de aprendizado denominada Máquina de Vetor de Suporte (*Support Vector Machine* - SVM) é também um algoritmo poderoso e cada vez mais popular tanto na indústria quando no meio acadêmico [14] [15]. A função objetivo usada para treinamento do SVM como classificador binário é dada por:

$$\begin{aligned} \min_{\mathbf{w}} C\left[\sum_{i=1}^{m} y^{(i)}\text{cost}_1(\mathbf{w}^T x^{(i)}) + (1-y^{(i)})\text{cost}_0(\mathbf{w}^T x^{(i)})\right] \\ + \frac{1}{2}\sum_{j=1}^{n}\mathbf{w}_j^2 \end{aligned} \tag{9}$$

onde, para cada i-ésimo exemplo do conjunto de treinamento, as funções cost(z) são aproximações lineares por partes das funções de custo do algoritmo de regressão logística. Cumpre notar que $\text{cost}_1(\mathbf{w}^T x^{(i)})$ e $\text{cost}_0(\mathbf{w}^T x^{(i)})$ consistem nas funções de custo do SVM compostas com as funções discriminantes lineares para as classes $y = 1$ e $y = 0$, respectivamente.

A expressão fora do somatório dos exemplos é o chamado *termo de regularização* [16], 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.

Com os custos acima sendo zero ficamos, a partir da Equação (9), com o problema reduzido de otimização sujeito a restrições:

$$\begin{aligned} \min_{\mathbf{w}} C\cdot 0 + \frac{1}{2}\sum_{j=1}^{n}\mathbf{w}_j^2 \textit{ sujeito a:} \\ \mathbf{w}^T x^{(i)} \geq 1 \textit{ se } y^{(i)} = 1 \textit{ e} \\ \mathbf{w}^T x^{(i)} \leq -1 \textit{ se } y^{(i)} = 0 \end{aligned} \tag{10}$$

Note que o produto interno $\mathbf{w}^T\mathbf{x}$ pode ser escrito como a magnitude da projeção de $\mathbf{x}$ que multiplica a

norma do vetor **w**. Assim sendo, a Equação (10) pode ser reescrita da forma:

$$\min_{\mathbf{w}} C \cdot 0 + \frac{1}{2}\sum_{j=1}^{n} \mathbf{w}_j^2 \textit{ sujeito a :}$$
$$p^{(i)} \cdot ||\mathbf{w}|| \geq 1 \textit{ se } y^{(i)} = 1 \textit{ e}$$
$$p^{(i)} \cdot ||\mathbf{w}|| \leq -1 \textit{ se } y^{(i)} = 0 \qquad (11)$$

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

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 *kernels* que medem a similaridade entre um exemplo $x^{(i)}$ e um marcador $l^{(i)}$ posicionado no espaço de características. Cada exemplo $x^{(i)}$ será então descrito por um vetor de novas características composto pelos graus de similaridade entre $x^{(i)}$ e cada um dos marcadores existentes. Usar SVM com *kernels* consiste na execução das etapas abaixo relacionadas [17]:

- Dados $(x^{(1)}, y^{(1)})$, $(x^{(2)}, y^{(2)})$, ..., $(x^{(m)}, y^{(m)})$, escolha $l^{(1)} = x^{(1)}$, $l^{(2)} = x^{(2)}$, ..., $l^{(m)} = x^{(m)}$
- Para cada exemplo $x^{(i)}$:

$$f = \begin{bmatrix} f_1 = \text{similaridade}(x^{(i)}, l^{(1)}) \\ f_2 = \text{similaridade}(x^{(i)}, l^{(2)}) \\ \vdots \\ f_m = \text{similaridade}(x^{(i)}, l^{(m)}) \end{bmatrix} \qquad (12)$$

  no qual a função de similaridade pode ser, por exemplo, uma função de base radial da distância euclidiana entre o exemplo $x^{(i)}$ e o marcador:

$$f_j = \exp\left(-\frac{\| x^{(i)} - l^{(j)} \|^2}{2\sigma^2}\right) \qquad (13)$$

- Uma vez obtido o novo vetor de características $\boldsymbol{f}$, executa-se a predição das classes na nova função discriminante linear: $y = 1$ se $\mathbf{w}^T f \geq 0$, por exemplo.
- Sendo os parâmetros **w** obtidos do treinamento do modelo com a minimização da função objetivo da Equação (9).

Note que para o uso descrito, há dois parâmetros de ajuste identificáveis: o parâmetro de regularização C e a variância do *kernel* gaussiano σ^2. 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 variâncias estreitas e largas por estarem associadas a variações bruscas e suaves das características $\boldsymbol{f}$, respectivamente.

3.3 Algoritmo de medida de *Loudness* multicanal para a radiodifusão

O algoritmo de medida de *loudness* multicanal apresentado originalmente em [2] é composto de quatro estágios: ponderação em frequência pela curva K, cálculo da média quadrática do sinal em cada canal, soma ponderada da energia do sinal nos canais e o cálculo do valor de *loudness* como ilustrado na Figura 1. A unidade de medida é LKFS, que pode ser traduzida como "*Loudness* ponderado pela curva K, referente ao fundo de escala digital (0 dBFS)". Medidas relativas são descritas com a Unidade de Loudness (*Loudness Unit*), para as quais 1 LU = 1 dB.

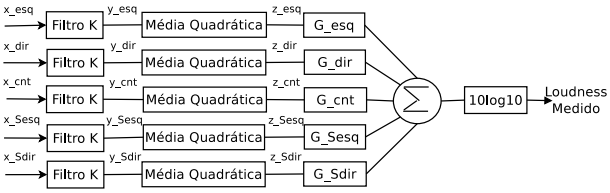


Figura 1: Diagrama em blocos do algoritmo de *loudness* multicanal constante da primeira versão da Recomendação ITU-R BS.1770

O bloco de média quadrática é o chamado *nível sonoro contínuo equivalente*, ou L_{eq} que corresponde à energia média do sinal durante o intervalo de medida. Matematicamente, L_{eq} é definido como:

$$L_{eq}(W) = 10\log_{10}\left(\frac{1}{T}\sum_{m=0}^{T-1}\frac{x_W^2[m]}{x_{Ref}^2[m]}\right) \text{ dB}$$
$$= 20\log_{10}\sqrt{\frac{1}{T}\sum_{m=0}^{T-1}\frac{x_W^2[m]}{x_{Ref}^2[m]}} \text{ dB} \qquad (14)$$

onde T é a duração do áudio, $x_W[m]$ é o sinal ponderado em frequência amostrado nos instantes m, e $x_{Ref}[m]$ é o sinal de referência (a segunda formulação é uma versão RMS da medida). Após diferentes avaliações de curvas típicas de ponderação em frequência, em [18] foi introduzida uma nova curva de ponderação, denominada *curva de ponderação B revisada em baixa frequência* (RLB), ilustrada na Figura 2. Esta é precedida de um pré-filtro para contabilizar os efeitos acústicos da cabeça [2], cuja curva de resposta é ilustrada na Figura 3. A combinação destas duas respostas em frequência compõem a curva K, aplicada ao sinal por filtragem IIR no domínio do tempo.

Considerando o fundo de escala digital como referência (0 dBFS), para o i-ésimo canal representado

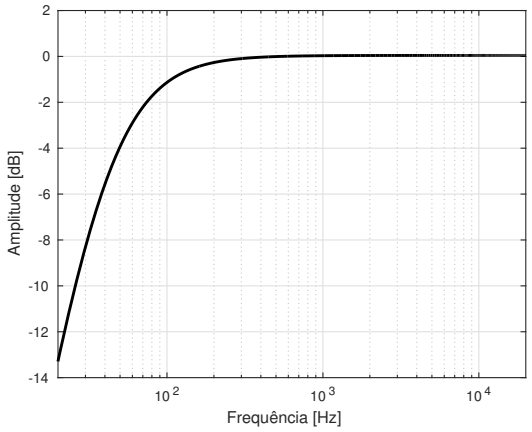


Figura 2: Curva de ponderação B revisada em baixa frequência (RLB) tal como proposta em [18].

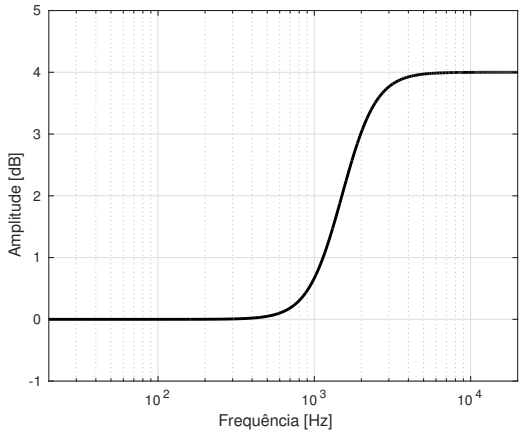


Figura 3: Pré-filtro utilizado para representação dos efeitos acústicos na cabeça [2].

no caminho de sinal da Figura 1, temos:

$$z_i = L_{eq}(K)\lfloor_{\text{Linear}} = \frac{1}{T}\sum_{m=0}^{T-1} y_i^2[m] = \frac{1}{T}\sum_{m=0}^{T-1} x_{iK}^2[m] \qquad (15)$$

onde $x_{iK}[m]$ é o sinal de entrada ponderado pela curva K, e $i \in I = \{L, R, C, L_S, R_S\}$, o conjunto de canais de entrada conforme configuração 5.1[1] padronizada em [20].

Os valores z_i são então ponderados pelos pesos G_i correspondentes aos ângulos de chegada dos sinais oriundos de cada canal, objetivando contabilizar a percepção mais intensa de ondas acústicas que incidam por trás do ouvinte. Caixas acústicas dispostas num azimute θ tal que $\frac{\pi}{3} \leq |\theta| \leq \frac{2\pi}{3}$, a exemplo das caixas dos canais *surround*, têm pesos $G = \sqrt{2}$ $(+1,5$ dB) e, para os demais azimutes, $G = 1$ (0 dB). Caso o ângulo de elevação ϕ seja tal que $|\phi| > \frac{\pi}{6}$, todos os ganhos G são unitários.

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

$$L_K = -0,691 + 10\log 10 \sum_i G_i \cdot z_i \text{ LKFS}, \qquad (16)$$

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 *phon* ou 64 *sone* [21].

O documento [2] atualmente se encontra em sua quarta versão. Na segunda edição foi introduzido o conceito de *loudness* segmentado (*gated loudness*) proposto em [3], que consiste na segmentação das integrações em blocos 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. A medida de nível de pico verdadeiro foi incluída na terceira revisão, e a versão mais recente acomodou sistemas de áudio com mais de cinco canais no cálculo do algoritmo.

3.4 *Loudness* Momentâneo e *Loudness* de Curta Duração

Quando da especificação dos medidores de *loudness* baseados no algoritmo em [2], os descritores *Loudness* Momentâneo e *Loudness* de Curta Duração foram sugeridos em [12] e [22] para caracterizar a percepção imediata num instrumento – como num medidor de volume (VU) – e a percepção da audiência quanto à variações de intensidade sonora. Ambos são calculados a partir da Equação (16), substituindo os limites de somatório da Equação (15) para 400 milissegundos e 3 segundos, respectivamente. No entanto, para as medidas de *Loudness* Momentâneo, o trabalho em [23] avaliou esta e outras alternativas de balística e sugeriu um filtro digital com resposta ao impulso de duração infinita (IIR) de primeira ordem com constante de tempo $\tau = 400$ ms como melhor opção, posteriormente incluída na Recomendação ITU-R BS.1771, que contém sugestão de coeficientes desde que o sinal seja dizimado a 320 amostras por segundo [24]. Para fins gerais de implementação, sua equação diferença é:

$$y[n] = (1-\alpha)x[n] + \alpha y[n-1], \qquad (17)$$

cujo coeficiente α é dado por:

$$\alpha = \exp\left(\frac{-1}{f_s \tau}\right), \qquad (18)$$

[1] 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 [19].

onde f_s é a frequência de amostragem e τ é a constante de tempo. Para $f_s = 48000$ Hz e $\tau = 400$ ms, a resposta em frequência do filtro IIR é ilustrada na Figura 4.

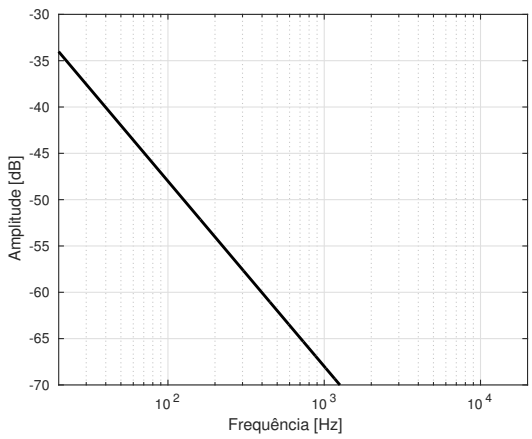


Figura 4: Resposta em frequência do filtro IIR de balística do medidor de *Loudness* Momentâneo para $\tau = 400$ ms e $f_s = 48000$ Hz.

Nota-se que este é um filtro passa-baixas com decaimento (*rolloff*) de -20 dB/década (ou -6 dB/oitava) que, se inserido no caminho do medidor, resulta num "galvanômetro" de ataque rápido e decaimento suave, cujo comportamento se assemelha aos antigos medidores analógicos (Ex.: VU e PPM) tornando-o mais confortável para o operador de áudio. Por esse motivo, esta foi a balística dos medidores de *loudness* desenvolvidos neste trabalho.

3.5 Medição de Pico Verdadeiro

A necessidade de abrir mão do paradigma de se controlar a saturação pelos picos das amostras nos medidores digitais foi apontada em [25], onde observou-se que não há 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. 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 equidistantes da ocorrência do pico verdadeiro de uma senóide oscilando na máxima frequência normalizada. Seja 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 :

$$\begin{aligned} \text{Erro}\lfloor_{\text{MAX}} &= 20log\left(cos\left(\frac{2\pi f_n f_s}{2Lf_s}\right)\right) \\ &= 20log\left(cos\left(\frac{\pi f_n}{L}\right)\right), \end{aligned} \tag{19}$$

onde o número 2 no denominador indica que o teto de erro é, pelo critério de Nyquist, metade da frequência de amostragem. A Equação (19) foi testada na terceira versão de [2] para diferentes valores de L e f_n. Na ocasião, verificou-se que para $L = 4$, $f_n = 0,45$ esta resultaria em Erro$\lfloor_{\text{MAX}} = 0,554$ dB e para $f_n = 0,5$ resultaria em Erro$\lfloor_{\text{MAX}} = 0,688$ 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 discricionariedade do projetista.

A implementação do medidor de pico verdadeiro neste trabalho se orienta pelo diagrama em blocos da Figura 5:

Figura 5: Diagrama em blocos do medidor de pico verdadeiro conforme especificações no Anexo 2 de [2].

O valor de atenuação/ganho de 12,04 dB advém da relação sinal-ruído do quantizador de (B + 1) bits [26]. 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:

$$\begin{aligned} \text{SNR} &= 10\log_{10}\left(\frac{\sigma_x^2}{\sigma_n^2}\right) \\ &= 6,02B + 10,8 - 20\log_{10}\left(\frac{FS}{\sigma_x}\right) \end{aligned} \tag{20}$$

Se o sinal possui uma boa excursão de faixa dinâmica, o termo $-20log_{10}\left(\frac{FS}{\sigma_x}\right)$ da Equação (20) torna-se bem pequeno. Nestas condições, pode-se dizer que a relação 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. Assim sendo, os blocos das pontas do diagrama não são obrigatórios para implementações em *software* usando aritmética de ponto flutuante, mas deverão ser observados para implementações em DSP[2].

3.6 Expansão de Faixa Dinâmica

Sinais com dinâmica altamente comprimida não precisam ser veiculados com uma intensidade sonora percebida como sendo mais forte que a do restante da

[2]O medidor é um aprimoramento do trabalho em [27], no qual são discutidos aspectos de implementação e erros de medida

programação. A expansão de faixa dinâmica é uma das formas de se controlar saltos de percepção de intensidade [28]. 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 [29]. Um sinal de áudio (de valores reais) $x[n]$ pode ser expresso como:

$$x[n] = x_r[n] + jx_i[n], \tag{21}$$

onde $x_r[n]$ e $x_i[n]$ também são sequências reais. Este sinal $x[n]$ é denominado *sinal analítico* e sua transformada de Fourier iguala zero na metade inferior do círculo unitário ($-\pi \leq \frac{2k\pi}{N} \leq 0$) [26]. Sejam $X_r[k]$ e $X_i[k]$ as transformadas discretas de Fourier de $x_r[n]$ e $x_i[n]$, então:

$$X[k] = X_r[k] + jX_i[k], \tag{22}$$

sendo

$$X_r[k] = \frac{1}{2}\left[X[k] + X^*[k]\right] \tag{23}$$

e

$$jX_i[k] = \frac{1}{2}\left[X[k] - X^*[k]\right] \tag{24}$$

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:

$$X_i[k] = \begin{cases} -jX_r[k], & 0 \leq \frac{2k\pi}{N} \leq \pi, \\ jX_r[k], & -\pi \leq \frac{2k\pi}{N} \leq 0. \end{cases} \tag{25}$$

ou

$$X_i[k] = H[k]X_r[k], \tag{26}$$

onde

$$H[k] = \begin{cases} -j, & 0 \leq \frac{2k\pi}{N} \leq \pi, \\ j, & -\pi \leq \frac{2k\pi}{N} \leq 0. \end{cases} \tag{27}$$

A resposta em frequência $H[k]$ possui magnitude unitária, um ângulo de fase de $-\frac{\pi}{2}$ rad para $0 \leq \frac{2k\pi}{N} \leq \pi$ e um ângulo de fase de $\frac{\pi}{2}$ rad para $-\pi \leq \frac{2k\pi}{N} \leq 0$. Este sistema linear é denominado *deslocador de fase em 90 graus* ou simplesmente *transformador de Hilbert*. Sua resposta ao impulso a partir da Equação (27) é dada por:

$$h[n] = \frac{1}{N}\sum_{k=0}^{\frac{N}{2}-1} je^{j(\frac{2\pi}{N})kn} - \frac{1}{N}\sum_{k=\frac{N}{2}}^{N-1} je^{j(\frac{2\pi}{N})kn} \tag{28}$$

ou

$$h[n] = \begin{cases} \frac{2}{\pi}\frac{sen^2(\frac{\pi n}{2})}{n}, & n \neq 0, \\ 0, & n = 0. \end{cases} \tag{29}$$

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

$$x_i[n] = \sum_{m=-\infty}^{\infty} h[n-m]x_r[m] \tag{30}$$

e

$$x_r[n] = -\sum_{m=-\infty}^{\infty} h[n-m]x_i[m] \tag{31}$$

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 (21). A magnitude do sinal analítico $|x[n]|$ é denominada *amplitude instantânea* e corresponde à envoltória do sinal de entrada $x_r[n]$.

4 METODOLOGIA

4.1 Classificação

Para o problema de detecção de segmentos com características de formato curto, como comerciais e vinhetas, as técnicas descritas na seção anterior foram aplicadas como se segue:

- Decompõe-se a matriz de covariância do conjunto de dados em valores singulares para determinar o máximo de dimensões redutíveis do problema.
- Escolhe-se o menor valor de k 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:

$$\frac{\sum_{i=1}^{k} S_{ii}}{\sum_{i=1}^{m} S_{ii}} \geq 0.99 \tag{32}$$

- Aplica-se análise de componentes principais no pré-processamento dos dados para reduzir o espaço de características em $m-k$ dimensões.
- Para a etapa de classificação, treina-se máquinas de vetor de suporte com $kernels$ distintos e com diferentes ajustes de C e σ, para então testar os modelos treinados no conjunto de validação e registrar o percentual de acertos (acurácia).

4.2 Processamento

Para a etapa subsequente, as técnicas apresentadas nas seções anteriores foram empregadas na construção de um decisor baseado nos critérios de níveis máximos em [10]:

- pico verdadeiro $= -1$ dBTP
- *loudness* de curta duração $= -18$ LKFS
- *loudness* momentâneo $= -15$ LKFS

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 subbloco de 19200 amostras com 50% de sobreposição, uma medida de *loudness* de curta duração e outra de pico verdadeiro para todo o bloco. Se algum dos critérios anteriores

for violado, o bloco tem sua faixa dinâmica expandida via atenuação das amostras com nível inferior ao limiar de 20 dB abaixo (20 dB *down*) do pico de magnitude do sinal analítico oriundo do transformador de Hilbert. Para os casos de pico verdadeiro superiores a -1 dBTP, os valores acima do limiar são minimamente atenuados para evitar ceifamento. O bloco "descomprimido" volta ao *playout* e o bloco subsequente é capturado, classificado, medido e expandido se necessário. O processo é realizado continuamente durante a transmissão.

5 TESTES

5.1 Classificador

Para os experimentos usou-se a base de dados "*TV News Channel Commercial Detection Dataset*" disponível em [9]. O conjunto de dados contém 6 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 Cruzamentos na Origem (ZCR), centróides espectrais, decaimento espectral e fluxo espectral [30].

A energia do sinal num tempo curto, ou num único quadro (STE), é calculada como sendo o quadrado das amplitudes das amostras contidas na janela limitante do quadro centrado no instante de análise $\hat{n}$:

$$E_{\hat{n}} = \sum_{m=0}^{N-1} (x[m]w[\hat{n}-m])^2 = \sum_{m=0}^{N-1} x^2[m]w^2[\hat{n}-m], \tag{33}$$

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:

$$Z_{\hat{n}} = \sum_{m=0}^{N-1} \frac{1}{2} \left| sgn\{x[m]\} - \text{sgn}\{x[m-1]\} \right| w[\hat{n}-m], \tag{34}$$

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 concentrada 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:

$$\sum_{k=0}^{m-1} X_i[k] = P \sum_{k=0}^{N-1} X_i[k], \tag{35}$$

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

$$\text{Centróide} = \frac{\sum_{k=0}^{N-1} k|X[k]|}{\sum_{k=0}^{N-1} |X[k]|} \tag{36}$$

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:

$$Fl_{(i,i-1)} = \sum_{k=0}^{N-1} \left(\frac{X_i[k]}{\sum_{l=0}^{N-1} X_i[l]} - \frac{X_{i-1}[k]}{\sum_{l=0}^{N-1} X_{i-1}[l]} \right)^2 \tag{37}$$

A fala apresenta fluxos maiores do que a música em razão de um maior número de transitórios e alterações entre fala vozeada e surda. A medida também é usada em detecção de timbre e identificação de *onsets*.

5.2 Processador

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. O *loudness* integrado dos blocos provenientes das trilhas do TS oscilou entre $-25,1$ e $-24,1$ LKFS, enquanto que o dos blocos distorcidos oscilou entre $-18,3$ e $-17,2$ LKFS, valores estes mais próximos dos critérios de controle de conteúdo de formato curto pretendidos.

6 RESULTADOS

6.1 Classificação

Os resultados dos testes sinalizam o sucesso da estratégia, como pode ser visto nos percentuais de acerto da classificação dos dados dos conjuntos de validação para SVMs treinados com diferentes ajustes de regularização do modelo e de variância da função de similaridade (C, σ) nas Tabelas 1, 2, 3 e 4. O percentual de acertos (ou acurácia) é medido baseado no número de verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos, da forma:

$$Acc(\%) = 100 \cdot \left(\frac{VP + VN}{VP + FP + VN + FN} \right) \tag{38}$$

Tabela 1: BBC

C	σ	(%)
0,01	0,01	47,65
0,01	0,10	56,52
0,01	1,00	77,17
0,01	3,00	78,24
0,03	1,00	78,48
0,10	1,00	79,53
0,30	1,00	80,26
1,00	0,10	91,65
3,00	0,10	92,28
30,0	0,30	92,63

Tabela 2: CNNIBN

C	σ	(%)
0,01	0,01	65,52
0,01	0,03	76,91
0,01	1,00	80,82
0,03	1,00	81,51
0,10	1,00	82,38
0,30	0,30	94,66
10,0	0,30	99,92
30,0	0,30	100,0
–	–	–
–	–	–

Tabela 3: NDTV

C	σ	(%)
0,01	0,01	73,46
0,01	0,03	73,66
0,01	0,10	74,96
0,01	1,00	84,00
0,10	1,00	84,50
1,00	1,00	84,90
3,00	1,00	85,18

Tabela 4: CNN

C	σ	(%)
0,01	0,01	36,10
0,01	0,03	83,02
0,01	0,10	98,00
1,00	0,10	98,61
3,00	0,30	99,81
10,0	0,30	99,98
–	–	–

6.2 Processamento

Como esperado, não houve atuação do controlador de *loudness* para o primeiro grupo de blocos com conteúdo normalizado a -23 LKFS. Já para as contrapartes distorcidas, houve expansão dinâmica nos segmentos com *loudness* de curta duração superiores a -18 LKFS e em alguns poucos segmentos contendo subblocos com *loudness* momentâneo superiores a -15 LKFS. Um limiar de expansão configurado a 20 dB abaixo do nível de pico da envoltória resultou em atenuações de *loudness* de curta duração da ordem de 1 a 3 LU, como ilustrado na Figura 6.

Figura 6: Controle de *loudness* num conteúdo de 60 segundos de duração. Houve expansão dinâmica nos 30 primeiros segundos e nos 15 últimos, nos quais as medidas de *loudness* de curta duração (*LoudnessIn*) foram superiores a -18 LKFS. As medidas nos blocos expandidos (*LoudnessOut*) giraram em torno de -20 LKFS.

7 DISCUSSÃO

Muito embora detectores de blocos comerciais que fazem uso de parâmetros audiovisuais produzam regiões de decisão mais complexas [9], um dos desafios deste trabalho residiu em descobrir se seria possível alcançar bons 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. As tabelas de 1 a 4 ilustram bons resultados se considerarmos os poucos descritores utilizados. Trabalhos futuros deverão testar outros descritores de baixa redundância como por exemplo, o rastreio da frequência fundamental e coeficientes cepstrais em na escala Mel de frequências (MFCCs).

Cabe ressaltar que o detector de conteúdo de formato curto aqui proposto foi testado somente no conjunto de dados descrito na Seção 5.1, pois o material bruto utilizado nos testes de processamento era insuficiente para um treinamento satisfatório 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 e o filtro de média exponencial para o cálculo do *loudness* momentâneo não produz retardo algum. Traballhos futuros deverão avaliar a implementação desta solução em *hardware*, atentando para alterações do esquema para operação em aritmética de ponto fixo, a exemplo da modificação alertada na Seção 3.5.

A atuação do expansor 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. Conforme sugerido na Seção 2, embora o método aqui proposto destine-se aos casos de peças de programação/propaganda não normalizadas, não há melhor cenário do que peças originalmente mixadas com uma dinâmica saudável.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Presidência da República, "Lei nº 10.222 de 09/05/2001," Maio 2001.

[2] União Internacional de Telecomunicações, "Recomendação bs.1770-2," Março 2011.

[3] União Europeia de Radiodifusão, "Recomendação r.128," Junho 2011.

[4] Ministério das Comunicações, "Portaria nº 354 de 11/07/2012," Julho 2012.

[5] Barry A Blesser, "Audio dynamic range compression for minimum perceived distortion," *Audio and Electroacoustics, IEEE Transactions on*, vol. 17, no. 1, pp. 22–32, Março 1969.

[6] Udo Zölzer, *DAFX: digital audio effects*, John Wiley & Sons, Chichester, UK, 3rd edition, Junho 2011.

[7] Earl Vickers, "The loudness war: Background, speculation, and recommendations," *Audio Engineering Society Convention 129*, pp. 1–27, Novembro 2010.

[8] Larry Percy and John R Rossiter, "A theory-based approach to pretesting advertising," *WD Wells*, pp. 267–281, 1997.

[9] Apoorv Vyas, Raghvendra Kannao, Vineet Bhargava, and Prithwijit Guha, "Commercial block detection in broadcast news videos," in *Proc. of 2014 Indian Conference on Computer Vision Graphics and Image Processing*. ACM, Dezembro 2014, p. 63.

[10] União Europeia de Radiodifusão, "Suplemento nº 1 da recomendação r.128," Novembro 2014.

[11] E Skovenborg, "Loudness Range (LRA) - Design and Evaluation," *Audio Engineering Society Convention*, vol. 132, pp. 1 – 12, Abril 2012.

[12] E Skovenborg and SH Nielsen, "Real-time Visualisations of Loudness Along Different Time Scales," *10th International Conference on Digital Audio Effects (DAFx)*, pp. 1–6, Setembro 2007.

[13] Isabelle Guyon, Steve Gunn, Masoud Nikravesh, and Lofti A Zadeh, *Feature extraction: foundations and applications*, vol. 207, Springer, 1st edition, Agosto 2008.

[14] Richard O Duda, Peter E Hart, and David G Stork, *Pattern classification*, John Wiley & Sons, Nova Iorque, 2nd edition, Agosto 2012.

[15] Marti A. Hearst, Susan T Dumais, Edgar Osman, John Platt, and Bernhard Scholkopf, "Support vector machines," *Intelligent Systems and their Applications, IEEE*, vol. 13, no. 4, pp. 18–28, 1998.

[16] A de P Braga, ACPLF Carvalho, and Teresa Bernarda Ludermir, *Redes neurais artificiais: teoria e aplicações*, Livros Técnicos e Científicos, Rio de Janeiro, 2nd edition, 2000.

[17] Ingo Steinwart and Andreas Christmann, *Support vector machines*, Springer Science & Business Media, Nova Iorque, 1st edition, 2008.

[18] Scott G Norcross and Gilbert A Soulodre, "Objective measures of loudness," in *Audio Engineering Society Convention 115*, Outubro 2003.

[19] Scott G Norcross and Michel C Lavoie, "Investigations on the inclusion of the lfe channel in the itu-r bs. 1770-1 loudness algorithm," in *Audio Engineering Society Convention 127*. Audio Engineering Society, Outubro 2009.

[20] União Internacional de Telecomunicações, "Recomendação bs.775-3," Agosto 2012.

[21] Hugo Fastl and Eberhard Zwicker, *Psychoacoustics: Facts and models*, vol. 22, Springer Science & Business Media, Nova Iorque, 3rd edition, 2007.

[22] União Europeia de Radiodifusão, "Ebu – tech 3341," Janeiro 2011.

[23] Scott Norcross, Felix Poulin, and Michel Lavoie, "Evaluation of live meter ballistics for loudness control," in *Audio Engineering Society Convention 130*. Audio Engineering Society, Maio 2011.

[24] União Internacional de Telecomunicações, "Recomendação bs.1771-3," Agosto 2012.

[25] Thomas Lund, "Stop Counting Samples," *Audio Engineering Society Convention*, vol. 121, pp. 1 – 11, Outubro 2006.

[26] Alan V Oppenheim, Ronald W Schafer, John R Buck, et al., *Discrete-time signal processing*, vol. 2, Prentice-hall Englewood Cliffs, Nova Jérsei, 2nd edition, 1989.

[27] Leandro da Silva Pires, "Medição de Picos Verdadeiros de Sinais de Áudio," in *Anais do Simpósio de Processamento de Sinais da UNICAMP*, Campinas, Setembro 2014.

[28] Guy W McNally, "Dynamic range control of digital audio signals," *Journal of the Audio Engineering Society*, vol. 32, no. 5, pp. 316–327, 1984.

[29] EA Feilat, "Detection of voltage envelope using prony analysis-hilbert transform method," *Power Delivery, IEEE Transactions on*, vol. 21, no. 4, pp. 2091–2093, Outubro 2006.

[30] LR Rabiner and RW Schafer, "Digital speech processing," *The Froehlich/Kent Encyclopedia of Telecommunications 6 (2011): 237*, vol. 258, 2011.

Sociedade de Engenharia de Áudio
Artigo de Congresso

Apresentado no 14º Congresso de Engenharia de Áudio
20ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2016, São Paulo, SP

Este artigo foi reproduzido do original final entregue pelo autor, sem edições, correções ou considerações feitas pelo comitê técnico. A AES Brasil não se responsabiliza pelo conteúdo. Outros artigos podem ser adquiridos através da Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA, www.aes.org. Informações sobre a seção Brasileira podem ser obtidas em www.aesbrasil.org. Todos os direitos são reservados. Não é permitida a reprodução total ou parcial deste artigo sem autorização expressa da AES Brasil.

Identificação de doenças de laringe a partir de sinais de vocalização e redes neurais artificiais

Bruno Ribeiro Sodré,[1] Marcelo de Oliveira Rosa,[1] Marco José da Silva[1] e Ana Paula Dassie-Leite[2]

[1] Universidade Tecnológica Federal do Paraná (UTFPR),
Pós-Graduação em Engenharia Elétrica e Informática Industrial
Curitiba, Paraná, 80620-210, Brasil

[2] Universidade Estadual do Centro-Oeste (UNICENTRO),
Departamento de Fonoaudiologia
Irati, Paraná, 84500-000, Brasil

engdesignbl@gmail.com, mrosa@utfpr.edu.br, mdasilva@utfpr.edu.br, pauladassie@hotmail.com

RESUMO

Neste artigo usa-se redes neurais artificiais para classificar sinais de vocalização (sintetizados por medidas acústicas que representam tais sinais) em 23 condições da laringe. Verificou-se que as 19 medidas acústicas oriundas de programas de computador de uso clínico podem ser reduzidas a apenas 6 através de análise de componentes principais. Os resultados mostraram que a acurácia e sensibilidade da classificação dos sinais em condição normal e patológica (2 classes) podem atingir 99.5% e 91.2%. Já a discriminação individual das patologias produziu resultados inferiores (acurácia e sensibilidade até 50.2% e 6.4%, respectivamente). Entretanto, mostrou-se que algumas patologias são mais fáceis de ser identificadas do que outras, justificando a similaridade acústica de algumas doenças da laringe.

0 INTRODUÇÃO

Diagnósticos médicos são realizados a partir de avaliações da anatomia e do funcionamento dos órgãos de interesse. No caso da voz, profissionais de saúde (laringologistas, por exemplo) empregam laringoscópios (acoplados ou não a dispositivos de gravação de vídeo) para exame da laringe, mais especificamente, das cordas vocais.

O processo de avaliação da laringe por vezes é difícil (exige que o clínico visualize a laringe), dependente da especialização e conhecimento do laringologista, e pequenas variações (por exemplo, redução de um tumor de laringe) podem não ser quantificado adequadamente. Em resumo, o diagnóstico em si é quali-

tativo.

Assumindo que a própria vocalização do paciente contém informações sobre as condições de sua laringe, determinou-se um conjunto de parâmetros numéricos estimados a partir do sinal de voz, que representam suas características espectro-temporais de modo sintético. Assim, o sinal de voz é traduzido nas chamadas medidas acústicas.

Teoricamente um sinal de voz de tratos vocal e glotal sem qualquer problema anatomo-funcional deve ser perfeitamente periódico, com alta SNR. Logo, desvios desses patamares teóricos implicariam em anormalidades que poderiam auxiliar na indicação do problema que afeta a laringe do locutor/paciente.

Perturbações na periodicidade da vocalização são capturadas por medidas acústicas de *jitter* e *shimmer*. *Jitter* quantifica variações da frequência fundamental de vibração das cordas vocais ao longo do tempo. Em vozes patológicas, tais variações tendem a serem maiores do que em vozes normais. Já *shimmer* quantifica variações da amplitude do sinal de fala, assumindo que vocalizações normais têm a amplitude de seus sinais praticamente constante.

Para calculá-las, primeiramente a curva de frequência fundamental (f_0) é estimada (geralmente por algoritmos baseados em autocorrelação ou correlação cruzada com algum aprimoramento heurístico [1, 2, 3]).

Com base nessa curva temporal, medidas de *jitter* como jitter médio [4] ou variação de f_0 em semi-tons [5] foram definidas. Para reduzir o efeito de lentas variações nessa curva, foram definidas medidas como o RAP [6] e PPQ [7], que são médias móveis centradas de tamanhos 3 e 11, respectivamente.

Medidas de *shimmer* quantificam a variação de amplitude do sinal de voz ao longo do tempo, como APQ [8, 9], que é uma média móvel centrada de tamanho 5 da curva de amplitude do sinal. Para gerar tal curva (de onde essas medidas são extraídas), os algoritmos buscam amplitudes máximas ou mínimas, ciclo-a-ciclo (usando a curva de f_0 como referência).

Tanto *jitter* e *shimmer* relacionam-se com níveis perceptuais de aspereza e rouquidão.

Para avaliar a presença de "ruídos" na voz, medidas que avaliam a relação entre a quantidade de energia dita harmônica e sua contraparte não-harmônica foram criadas, como HNR [10]. Tais avaliações podem ser calculadas no domínio do tempo [1] ou no domínio da frequência, e se relacionam com níveis perceptuais de rouquidão.

Outra medida acústica para tal avaliação perceptual é a energia de ruído normalizada (NNE) [11], que usa filtro do tipo pente, adaptativo, para isolar e quantificar a energia não-harmônica. Em uma voz normal, tal energia seria mínima.

Recentemente, medidas baseadas em comportamento dinâmico do sistema de produção de fala (análise de recorrência [12] e análise caótica [13]) e em coeficientes de análise por wavelets ([14]) mostraram-se promissoras para avaliação das condições da laringe (apesar de não estarem presentes em programas de computador de uso clínico).

Apesar do uso inicial de medidas acústicas individuais para identificação de doenças da laringe, métodos estatísticos multidimensionais permitiram aumento da taxa de acertos, apesar do limitado conjunto de patologias a identificar [12, 13, 14].

Este trabalho avaliou de forma exaustiva combinações diversas dessas medidas acústicas em um reconhecedor de padrões baseado em redes neurais artificiais para determinar: que patologias são mais fáceis e mais difíceis de serem identificadas, que medidas acústicas são úteis ou descartáveis (individualmente ou em grupos), e que combinações de medidas acústicas são clinicamente úteis. Estas avaliações foram feitas a partir de informações constantes em matrizes de confusão [15], focando particularmente falsos negativos e falsos positivos, além de acurácia, visto que são essas as informações pertinentes para o uso clínico desta abordagem (algo pouco discutido nos trabalhos analisados). As medidas acústicas usadas são encontradas em programas de computador para análise de voz usados em clínicas médicas.

Assim, apresenta-se uma descrição da base de dados usada para o estudo, com comentários sobre as similaridades acústicas de algumas das patologias analisadas. Posteriormente as medidas acústicas são descritas, bem como sua relação com aspectos aerodinâmicos e vibratórios da laringe. Na sequência as redes neurais artificiais usadas são apresentadas sucintamente (focando em seu emprego neste trabalho, para auxílio na reprodutibilidade das análises). Finalmente os resultados obtidos são apresentados e discutidos, encerrando-se com uma conclusão.

1 MATERIAIS E MÉTODOS

1.1 Sinais e doenças da laringe

Para este estudo, o banco de sinais de fala *Massachusetts Eye and Ear Infirmary Voice and Speech Laboratory MEEEI4337* foi usado para permitir que os resultados aqui apresentados sejam comparados com os de outros autores. De sua totalidade, foram analisadas apenas as vocalizações do fonema /a/, sendo 59 sinais oriundos de locutores normais e 898 de locutores disfônicos. A Tabela 1 quantifica o número de sinais por patologias da laringe usados (22 no total), além dos sinais de locutores normais.

Algumas considerações importantes: Nota-se que o número de sinais varia significativamente entre as patologias. Isso afeta os reconhecedores de padrões (como será mostrado posteriormente).

Há similaridade inter-patologias: por exemplo, cistos, pólipos e nódulos impedem o fechamento completo da glote pelas cordas vocais durante a fase de fecha-

Tabela 1: Número de sinais por patologia da laringe - incluindo caso sem patologia

	Patologia	*Número de locutores*
1	Normal (sem patologia)	59
2	Constrição antero-posterior	164
3	Curvatura das cordas vocais	11
4	Disfonia de conversão	12
5	Cistos	9
6	Refluxo gástrico	47
7	Granuloma	6
8	Hemi-laringectomia	4
9	Hiperfuncionalidade	258
10	Leucoplasia	26
11	Papiloma	4
12	Paralisia de corda vocal	60
13	Paresia	18
14	Presbifonia	13
15	Edema de Reinke	25
16	Presença de cicatrizes	16
17	Disfonia espasmódica	18
18	Compressão ventricular	108
19	Fadiga vocal	3
20	Edema nas pregas vocais	44
21	Pólipos vocais	19
22	Nódulos vocais	19
23	Tremores vocais	14
	Total	957

mento glotal. Isso significa que a percepção acústica da voz de pacientes com tais patologias é similar, e, naturalmente, afeta a performance dos reconhecedores.

1.2 Medidas acústicas

As medidas acústicas usadas são encontradas em programas de computador voltados para análise acústica da voz. Elas são listadas na Tabela 2. Notar-se-á que algumas medidas acústicas não se relacionam diretamente com disfonias: por exemplo, $\overline{f_0}$ e $\overline{T_0}$ (valores médios da frequência e período fundamentais do sinal de voz) estão diretamente ligadas ao gênero do locutor (apesar de situações cotidianas de homens com alto f_0 e mulheres com baixo f_0). Seu uso se justifica para avaliar sua relação com vozes disfônicas.

O termo ciclo-a-ciclo indica que alguns medidas acústicas exigiram cálculos ciclo-a-ciclo ao longo do sinal de voz para gerar uma série de valores que receberam um tratamento estatístico (média, valor máximo, etc.).

Adicionalmente, as medidas acústicas foram calculadas ao longo do sinal de voz (vogal sustentada /a/ no caso), ou seja, capturaram variações que a laringe do locutor pudesse produzir durante a vocalização.

Essas medidas acústicas foram calculadas a partir de sinais de voz do fonema sustentado /a/, pronunciado por locutores americanos (homens e mulheres). Os sinais foram amostrados a 50kHz e 16bits/amostras. Os cálculos foram realizados sobre o trecho vocálico do

Tabela 2: Descritivo de medidas acústicas usadas neste trabalho

Índice	*Medida*	*Descritivo*
1	$\overline{f_0}$	Valor médio das f_0 ciclo-a-ciclo
2	$\overline{T_0}$	Valor médio dos T_0 ciclo-a-ciclo
3	f_{hi}	Máxima f_0 ciclo-a-ciclo
4	f_{lo}	Mínimo f_0 ciclo-a-ciclo
5	STD	Desvio padrão das f_0 ciclo-a-ciclo
6	PFR	Intervalo de f_0 ciclo-a-ciclo, em expressa em semi-tons
7	TSAM	Período de vocalização do sinal de voz
8	JITA	*Jitter* absoluto calculado a partir dos T_0 ciclo-a-ciclo [4]
9	JITT	Percentual de *jitter*, ou *jitter* relativo [16]
10	RAP	Pertubação média relativa dos T_0 ciclo-a-ciclo [6]
11	PPQ	Quociente de perturbação de T_0 ciclo-a-ciclo [7]
12	VF0	Desvio padrão dos T_0 ciclo-a-ciclo
13	SHDB	*Shimmer*, em decibéis
14	SHIM	Percentual de *shimmer*
15	APQ	Quociente de perturbação em amplitude [8]
16	VAM	Desvio padrão das amplitudes ciclo-a-ciclo
17	NHR	Razão entre energias harmônica e não-harmônica [10]
18	VTI	Índice de turbulência da voz, ou nível de energia relativa de ruído em alta-frequência (até $5{,}8\text{kHz}$)
19	SPI	Índice fonético suave, determinado pela razão de energias harmônicas entre as faixas espectrais de 70-1600Hz e $1{,}6$-$4{,}5\text{kHz}$

sinal (os trechos de silêncio foram descartados manualmente).

1.3 Redes neurais artificiais (RNAs)

Redes com *perceptrons* em multicamadas [17] atuaram como reconhecedores de padrões neste estudo. Apenas uma camada intermediária foi empregada, na qual o número de seus neurônios variou de 1 a 40.

A entrada das RNAs foi composta de todas as medidas acústicas disponíveis no banco de vozes, uma medida por neurônio dessa camada. Adicionalmente aplicou-se a análise das componentes principais (ACP) para redução de dimensão desses vetores (de 19 para 6, segundo o critério de Guttman-Kaiser [18, 19]). Por esse critério, foram mantidas as componentes principais cujos autovalores da matriz de covariância das 19 medidas acústicas foram superiores ao valor médio dos 19 autovalores. Aqui, apenas os 6 primeiros componentes principais tinham autovalores superiores a tal média.

Para a camada de saída, duas análises foram feitas:

- Primeiramente cada neurônio da camada de saída

representou uma patologia ou tipo de voz (no total de $T = 23$). Logo, na codificação do vetor de saída, o valor "1" foi atribuído a um dos neurônios de saída se o locutor daquela voz possuia a patologia que o neurônio representava. O valor zero indicou o contrário;

- Além dessa identificação individual dos tipos de voz, considerou-se também a possibilidade da RNA identificar se um locutor possui ou não uma disfonia. Logo, a camada de saída conteve apenas 1 neurônio, sendo codificado com o valor zero para um locutor sem qualquer patologia (normal) e "1" para o caso contrário.

Assim, dado um sinal de voz (reduzido a um conjunto de medidas acústicas), a intenção foi que os neurônios de saída RNA indicassem cada patologia que o locutor desse sinal possuia.

Para se avaliar o potencial (individual e em conjunto) das medidas acústicas na identificação das doenças da laringe (incluindo indicativo de locutor sem doença), gerou-se $P = \left(\sum_{n=1}^{19} C_{19}^{n}\right)$ combinações das medidas acústicas para os $N = 957$ sinais de voz, resultando em P RNAs. Para o caso envolvendo ACP (com 6 dimensões), $P = \left(\sum_{n=1}^{6} C_{6}^{n}\right)$.

Todas as RNAs foram treinadas individualmente por retropropagação de erros de estimação usando algoritmo de Levenberg-Marquardt. O conjunto de dados dessas RNAs foi dividido (aleatoriamente, por RNA) em subconjuntos de treinamento, teste e validação a razão de $60/20/20$. Essas fases de treinamento, teste e validação foram repetidas 10 vezes e cada uma gerou uma matriz de confusão $\mathbf{M_i}$, $(1 \leq i \leq 10)$. A matriz de confusão final para as 10 repetições é $\mathbf{M} = \sum_{i=1}^{10} \mathbf{M_i}$ (os percentuais de acurácia, sensitividade, precisão e especificidade foram calculados sobre $\mathbf{M}$). Logo, o total de análises foi $10 \times P$).

Os seguintes parâmetros de treinamento foram usados em todos os casos:

- Max. épocas para treinamento = 1000;
- Parâmetro de amortecimento (μ) inicial = 0.001;
- Fator de atualização de μ = 10;
- Máximo $\mu = 10^{10}$;
- Função de transferência do tipo tangente hiperbólico sigmoidal para neurônios da camada intermediária;
- Função de transferência do tipo linear para neurônios da camada de saída;

2 RESULTADOS

2.1 Normal vs patológico

Primeiramente verificou-se quais as configurações de entradas produziram os melhores e piores resultados, baseado em pior taxa de verdadeiros positivos ou negativos (sensibilidade ou especificidade), performance de identificação (precisão) e média desses percentuais (incluindo acurácia, para definir uma performance geral de uma RNA) [15]. Essas são avaliações diretas de uma matriz de confusão. Neste estudo, os casos positivo e negativo foram respectivamente locutor com e sem disfonia.

Tabela 3: Melhores e piores resultados para acurácia e pior taxa de sensibilidade ou especificidade na identificação de locutores com ou sem disfonia (valores em percentuais). Os valores das medidas são índices na Tabela 1.

Ordem	*Medidas*	*Acurácia*	*Medidas*	*Sens./Espec.*
1^a	1, 3, 4, 5, 7, 8, 9, 11, 12, 14, 15	99.5	1, 3, 4, 5, 7, 8, 9, 11, 12, 14, 15	91.2
2^a	1, 3, 4, 5, 7, 8, 9, 10, 12, 13, 16, 19	99.5	1, 3, 4, 5, 7, 8, 9, 10, 12, 13, 16, 19	91.2
3^a	1, 7	99.4	1, 4, 5, 7, 8, 9, 17, 18	91.2
4^a	3, 7	99.4	4, 7, 10, 12, 17	91.0
...	...	...	...	...
524285^a	3, 5, 6, 11, 12, 17	86.3	5, 6, 8, 9, 11, 16, 18	0.0
524286^a	1, 2, 4, 9, 11, 13, 14, 17, 19	86.2	5, 8, 9, 10, 12, 16	0.0
524287^a	3, 6, 8, 13, 19	86.1	6, 11, 16, 17, 18	0.0

Tabela 4: Melhores e piores resultados para precisão e média geral das métricas da matriz de confusão na identificação de locutores com ou sem disfonia (valores em percentuais). Os valores das medidas são índices na Tabela 1.

Ordem	*Medidas*	*Precisão*	*Medidas*	*Média geral*
1^a	1, 4, 7, 8, 9, 10, 11, 12, 13, 16, 18	99.4	1, 3, 4, 5, 7, 8, 9, 11, 12, 14, 15	95.4
2^a	1, 3, 4, 5, 7, 8, 9, 10, 12, 13, 16, 19	99.4	1, 3, 4, 5, 7, 8, 9, 10, 12, 13, 16, 19	95.4
3^a	1, 4, 5, 7, 8, 9, 17, 18	99.4	1, 4, 5, 7, 8, 9, 17, 18	95.3
4^a	4, 7, 10, 12, 17	99.4	4, 7, 10, 12, 17	95.2
...	...	...	...	...
524285^a	5, 8, 9, 11, 12, 16, 18	0.0	5, 8, 11, 12	23.5
524286^a	5, 8, 9, 10, 12, 16	0.0	5, 8, 9, 10, 12, 16	23.5
524287^a	6, 11, 16, 17, 18	0.0	6, 11, 16, 17, 18	23.4

O uso de piores sensibilidade ou especificidade garantiu que uma alta taxa de verdadeiros positivos não mascarasse uma baixa taxa de verdadeiros negativos (e vice-versa).

Dada a quantidade de resultados, são apresentados apenas situações extremas que caracterizam a dispersão

dos resultados produzidos pelas RNAs. As Tabelas 3 e 4 mostram que apesar de uma acurácia (taxa de acertos) superior a 86.1%, houve casos significativos de falsos positivos e falsos negativos (0.0% de sensibilidade, ou especificidade, ou precisão). Tal situação foi raramente discutida em outros trabalhos.

Tabela 5: Percentual de participação de cada medida acústica nos melhores resultados de classificação segundo critério de média geral da matriz de confusão.

Medida acústica	*100 melhores*	*200 melhores*	*500 melhores*
$\overline{f_0}$	33	31	29
$\overline{T_0}$	18	21	19
f_{hi}	31	26	30
f_{lo}	71	83	85
STD	32	29	30
PFR	18	21	23
TSAM	100	100	100
JITA	15	18	23
JITT	7	15	24
RAP	12	17	22
PPQ	9	19	22
VF0	20	20	25
SHDB	17	20	23
SHIM	13	17	24
APQ	16	15	19
VAM	12	15	18
NHR	22	27	31
VTI	11	15	17
SPI	11	16	18

Além disso, claramente houve uma diversidade de conjuntos de medidas acústicas que classificaram bem as vozes em normais ou patológicas. Isso indica que várias dessas medidas acústicas eram matematicamente relacionadas. Para identificar a melhor medida acústica, determinou-se quantas vezes (contagem) uma medida acústica fez parte dos 100, 200 e 500 melhores resultados de classificação normal-patológica a partir da métrica média geral.

Os resultados da Tabela 5 mostram claramente que a medida acústica TSAM foi preponderante para a classificação de locutores com ou sem disfonia, seguida da medida acústica f_{lo}. A medida acústica NHR (razão entre níveis de energia harmônica e não harmônica), STD (desvio padrão da curva de frequência fundamental) e f_{hi} apresentaram-se em, aproximadamente, 30% das melhores RNAs.

Quando aplicou-se o ACP à versão normalizada das medidas acústicas (tal normalização foi feita subtraindo cada medida acústica de sua média, e dividindo o resultado pelo desvio padrão da medida acústica) e selecionou-se apenas 6 das 19 medidas rotacionadas (ACP envolve rotação de espaço para obtenção de medidas descorrelacionadas linearmente entre si) reforçou-se, pelas Tabelas 6 e 7, que as medidas acústicas apresentadas apresentaram muita redundância, pois tais resultados foram equivalentes aos apresentados pelas RNAs usando as medidas acústicas diretamente.

Na avaliação com medidas normalizadas e rotacionadas, não se verificou relação entre o número de neurônios na camada intermediária e as métricas de matriz de confusão: O aumento de neurônios nessa camada não melhorou os resultados de classificação.

As medidas rotacionadas 1, 4, e 6 participaram em 100% dos 30 melhores resultados de classificação, a partir da métrica de matriz de confusão média geral.

Tabela 6: Melhores e piores resultados para acurácia e pior taxa de sensibilidade ou especificidade na identificação de locutores com ou sem disfonia (valores em percentuais) para medidas normalizadas e rotacionadas via ACP (valores das medidas são ordinais de acordo com autovalores de ACP, e valores entre colchetes são número de neurônios na camada intermediária das RNAs).

Ordem	*Medidas*	*Acurácia*	*Medidas*	*Sens./Espec.*
1^a	1, 2, 3, 4, 6 [13]	99.4	1, 2, 3, 4, 6 [13]	90.3
2^a	1, 2, 4, 6 [15]	99.4	1, 2, 4, 6 [15]	90.3
3^a	1, 2, 3, 4, 5, 6 [9]	99.4	1, 2, 3, 4, 6 [1]	90.3
4^a	1, 2, 4, 6 [17]	99.4	1, 2, 3, 4, 5, 6 [9]	90.0
...	...	...	...	...
2518^a	5 [15]	86.3	5 [19]	0.0
2519^a	3, 4 [23]	89.5	5 [31]	0.0
2520^a	2, 5 [33]	85.6	5 [15]	0.0

Tabela 7: Melhores e piores resultados para precisão e média geral das métricas da matriz de confusão na identificação de locutores com ou sem disfonia (valores em percentuais) para medidas normalizadas e rotacionadas via ACP (valores das medidas são ordinais de acordo com autovalores de ACP, e valores entre colchetes são número de neurônios na camada intermediária das RNAs).

Ordem	*Medidas*	*Precisão*	*Medidas*	*Média geral*
1^a	1, 2, 3, 4, 6 [13]	99.4	1, 2, 3, 4, 6 [13]	94.9
2^a	1, 2, 4, 6 [15]	99.4	1, 2, 4, 6 [15]	94.9
3^a	1, 2, 3, 4, 5, 6 [9]	99.3	1, 2, 3, 4, 5, 6 [9]	94.7
4^a	1, 2, 4, 6 [17]	99.3	1, 2, 4, 6 [17]	94.7
...	...	...	...	...
2518^a	5 [19]	0.0	5 [19]	23.4
2519^a	5 [31]	0.0	5 [31]	23.4
2520^a	5 [15]	0.0	5 [15]	23.4

2.2 Identificação de disfonias

Considerando os resultados obtidos na seção anterior para a redução de dimensão (19 para 6) das entradas das RNAs via normalização de medidas acústicas e ACP, analisou-se a capacidade de RNAs discriminarem patologias da laringe. Os resultados apresentados foram obtidos com RNAs cuja camada intermediária possuia apenas 20 neurônios. O aumento desses neurônios apenas reduziu a dispersão dos resultados (diferença entre máxima e mínima acurácia).

Primeiramente, nota-se a redução significativa na capacidade de discriminação de doenças da laringe (Tabelas 8 e 9). Nota-se que mesmo com o uso de todas as medidas rotacionadas, obteve-se apenas 50.4%.

Cabe uma explicação sobre as métricas sensibilidade e especificidade mostradas na Tabela 9. Elas representam os piores resultados obtidos para cada uma dos tipos de voz. Isso significa que algumas patologias, por exemplo, apresentaram percentual muito baixo de verdadeiros positivos (baixa sensibilidade), mesmo que outras tenham apresentado alta sensibilidade. Outras patologias apresentaram percentual elevado de verdadeiros negativos (alta especificidade). Esses piores resultados reforçam a dificuldade de discriminar patologias que são acusticamente similares (conforme mencionado anteriormente).

Tabela 8: Melhores e piores resultados para acurácia e precisão na discriminação das doenças da laringe para medidas normalizadas e rotacionadas via ACP (Valores entre colchetes são número de neurônios na camada intermediária das RNAs).

Ordem	*Medidas*	*Acurácia*	*Medidas*	*Precisão*
1^a	1, 2, 3, 4, 5, 6	50.4	1, 2, 3, 4, 5, 6	14.8
2^a	1, 3, 4, 5, 6	45.2	1, 2, 3, 4, 6	11.5
3^a	1, 2, 4, 5, 6	41.8	1, 3, 4, 5, 6	11.3
4^a	2, 3, 4, 5, 6	41.5	2, 3, 4, 5	11.0
...	...	...	...	...
61^a	1	8.6	1	0.0
62^a	6	8.2	6	0.0
63^a	2	7.2	2	0.0

Assim, mesmo usando todas as medidas rotacionadas, obteve-se apenas 6.4% de sensibilidade (pior caso de sensibilidade para os 23 tipos de vozes). O elevado valor de especificidade indica que há uma tendência da RNA com todas essas medidas rotacionadas acertar verdadeiros positivos (94.9%). Isso será discutido posteriormente.

A Tabela 10 é uma expansão dos resultados sumarizados nas Tabelas 8 e 9. Nela as patologias (e tipo de voz normal) com melhores e piores sensibilidades e especificidades na discriminação via RNAS são listadas.

Reforça-se o fato de que locutores normais foram melhor identificáveis em relação a todas as patologias da laringe (maiores sensibilidade e especificidades médias). Esse tipo de voz foi seguido de "papiloma", "fadiga vocal", "hemilaringectomia" e "nódulos vocais". Parte dessa qualidade nos resultados pode ser atribuída à baixa relação quantidade/doença (excetuando-se o caso do "nódulo vocal").

Tabela 9: Melhores e piores resultados para sensibilidade e especificidade na discriminação das doenças da laringe para medidas normalizadas e rotacionadas via ACP (Valores entre colchetes são número de neurônios na camada intermediária das RNAs).

Ordem	*Medidas*	*Sensib.*	*Medidas*	*Especif.*
1^a	1, 2, 3, 4, 5, 6	6.4	1, 2, 3, 4, 5, 6	94.9
2^a	1, 3, 4, 5, 6	5.1	2, 3, 4, 5, 6	94.9
3^a	2, 3, 5, 6	4.1	1, 3, 4, 5, 6	94.7
4^a	1, 2, 4, 5, 6	3.9	1, 2, 4, 5, 6	94.7
...	...	...	...	...
61^a	1	0.0	1, 2	82.6
62^a	6	0.0	5	81.9
63^a	2	0.0	6	80.1

Em sentido oposto, os piores resultados foram obtidos para "constrição ânterior-posterior", "compressão ventricular" e "refluxo gástrico". Isso significa que tais tipos de voz foram de difícil discriminação, pois as RNAs, nesses casos, apresentaram a tendência de identificar tais sinais como outras patologias.

3 DISCUSSÕES

A partir dos sinais utilizados e das medidas acústicas extraídas destes (que são encontradas em programas de computador para uso clínico), as RNAs conseguem determinar se uma voz é proveniente de uma laringe normal ou de uma acometida por alguma doença. Tal resultado é similar aos obtidos por [20] (que usa o mesmo banco de sinais), [13] e [14]. Enquanto esses autores empregaram medidas acústicas mais sofisticadas, o presente trabalho produziu resultados similares quando empregou medidas acústicas encontradas em programas de computador de uso clínico.

Na discriminação dos tipos de voz, as RNAs "identificaram" uma característica intrínseca a diversas patologias: similaridade acústica. Isso fica claro no reduzido percentual de sensibilidade do classificador. A tendência do classificador baseado em RNAs foi produzir grande quantidade de verdadeiros negativos (alta especificidade) inter-patologias.

Usando o mesmo banco de sinais com diferentes medidas acústicas e classificador, [21] trabalhou apenas com conjuntos separados de 4 e 5 doenças da laringe, ao invés dos 23 tipos de voz do presente trabalho. Somente quando tais autores redefiniram as doenças, assumindo que algumas vozes podem ser atribuídas a uma ou mais patologias (ou tipos de voz) foi que os autores obtiveram resultados de sensibilidade superiores aos aqui apresentados.

Tabela 10: Relação entre métricas sensibilidade e especificidade e os tipos de vozes (locutores com e sem disfonia) nos resultados de discriminação de doenças da laringe. Tipo de voz é um índice para as doenças descritas na Tabela 10.

Tipo	*Sensibilidade*		*Especificidade*	
de voz	*Média*	σ	*Média*	σ
1	67.68	20.38	97.38	3.61
2	2.63	2.74	98.64	2.19
3	20.00	14.02	95.53	2.88
4	20.44	14.56	97.02	2.20
5	31.41	17.83	95.80	2.56
6	6.63	5.00	97.52	2.44
7	48.25	18.27	93.37	4.63
8	50.97	24.06	96.00	3.63
9	3.12	3.71	98.18	3.34
10	13.76	8.79	96.90	1.79
11	58.55	22.05	96.36	3.34
12	8.84	5.99	98.00	1.45
13	14.97	12.82	97.12	2.08
14	14.81	10.67	97.15	1.54
15	28.23	13.69	94.30	2.88
16	38.80	16.49	96.80	3.42
17	14.46	10.89	95.25	3.28
18	4.14	4.16	98.10	2.86
19	56.15	23.62	97.64	3.81
20	10.62	8.11	97.24	2.00
21	15.47	9.84	96.29	2.28
22	44.65	17.20	94.33	3.44
23	17.61	10.82	97.66	1.56

Em [12], as análises também restringiram o universo de patologias (Edema, paralisia e nódulos), usando técnica de análise de recorrência. Para esses casos, quando comparados individualmente à voz normal, significativa taxa de acertos (incluindo métricas de matriz de confusão) foi obtida. Aqui evidenciou-se que a doença "nódulo vocal" realmente é discriminável em relação às demais disfonias, mas em uma abordagem classificatória que envolve todas as outras disfonias.

Decomposição por análise wavelets e índices de previsibilidade para seleção de medidas acústicas adequadas à discriminação de doenças de laringe [14] também produziram resultados melhores do que os apresentados aqui. Entretanto, as avaliações restringiram-se a vozes normais e aquelas oriundas de nódulos e edemas de Reinke.

ACP mostrou que muitas das medidas acústicas são relacionadas entre si. O critério usado para a redução (Guttman-Kaiser) definiu que as 19 medidas acústicas podiam ser representadas por apenas 6 medidas (ortogonais). Considerando-se que as medidas acústicas usadas restringiram-se a características como perturbações temporais de f_0 e de amplitude, e níveis de ruído, esta redução parece coerente. Os resultados atestaram que tal redução é possível. Logo, seria necessário definir medidas acústicas que possuam outras características que não essas três para melhora dos resultados de discriminação de tipos de voz. Isso é corroborado pelos resultados obtidos por outros autores em suas tentativas empregando técnicas distintas das usadas aqui.

4 CONCLUSÕES

Este trabalho mostrou como medidas acústicas encontradas em programas de computador para uso clínico podem ser usadas, juntamente com um classificador baseado em RNAs, para classificação de vozes em normais ou oriundas de laringes com alguma patologia da laringe, e para discriminação de doenças da laringe. Os resultados foram apresentados a partir de análise das matrizes de confusão resultantes dos testes dos classificadores, e sempre focaram os piores percentuais de acurácia, sensibilidade, especificidade e precisão.

Combinando-se exaustivamente as medidas acústicas, obteve-se uma acurácia de 99.5% (com elevados sensibilidade, especificidade e precisão). Assim, reforçados por outros estudos, é possivel afirmar que a classificação de uma voz como proveniente de uma laringe normal ou com alguma disfonia é possível. Tais resultados apoiaram-se fundamentalmente nas medidas acústicas temporais, envolvendo tempo de fonação e *jitter*.

A discriminação patológica é possível para algumas patologias, visto que o patamar superior de acertos foi de 50.4%, com algumas patologias apresentando sensibilidade de no mínimo 6.4% para tal patamar de acurácia. Individualmente há tipos de voz mais favoráveis a discriminação (evidência reforçada pelos estudos analisados), sendo que outras influenciam negativamente os resultados de sensibilidade e acurácia do identificador neural de doenças da laringe.

Atualmente há grande ênfase no chamado *deep learning*. Tal abordagem só seria possível com a disponibilidade de um grande banco de vozes. O estudo voltou-se para RNAs com apenas 1 camada intermediária de neurônios, sem recorrência, para clara definição de que medidas acústicas são mais significativas, e que patologias são mais fáceis de serem identificadas, o que seria difícil de determinar com abordagens do tipo *deep learning* ou uso de redes recorrentes.

Conclui-se que a discriminação automática de doenças da laringe a partir dos sinais de voz deve ser feita se os tipos (classes) de vozes que se deseja identificar sejam redefinidos. Ao invés de se identificar pólipos, cistos e nódulos, por exemplo, talvez seja necessário qualificá-los como um única classe, como alguns estudos mostraram. Tal reclassificação prévia deve ser discutida para padronização dos resultados dos estudos.

Clinicamente os resultados mostram que um método de triagem automatizado poderia ser construído a partir das técnicas matemáticas apresentadas aqui, permitindo a identificação rápida de que uma pessoa é portadora de alguma disfonia. Após tal identificação,

o paciente seria encaminhado a um especialista para o diagnóstico e tratamento dessa disfonia. Mais ainda, a solução poderia ser encapsulada em pequenos dispositivos portáteis (inclusive *smartphones*, devidamente calibrados).

REFERÊNCIAS BIBLIOGRÁFICAS

[1] P. Boersma, "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," *IFA Proceedings*, vol. 17, pp. 97–110, 1993.

[2] P. Boersma, "Should jitter be measured by peak-picking or by waveform matching?," *Folia Phoniatrica et Logopaedica*, vol. 61, pp. 305–308, 2009.

[3] A. de Cheveigne and H. Kawahara, "Yin, a fundamental frequency estimator for speech and music," *J. Acoust. Soc. Am.*, vol. 111, pp. 1917–1930, 2002.

[4] Y. Horii, "Fundamental frequency perturbation observed in sustained phonation," *J. Speech Hear. Res.*, vol. 22, pp. 5–19, 1979.

[5] I. Honjo and N. Isshiki, "Laryngoscopic and voice characteristics of ages persons," *Arch. Otolaryngol.*, vol. 106, pp. 149–150, 1980.

[6] Y. Koike, "Application of some acoustic measures for the evaluation of laryngeal dysfunction," *Studia Phonologica*, vol. 7, pp. 17–23, 1973.

[7] H. Takahashi and Y. Koike, "Some perceptual dimensions and acoustical correlates of pathologic voices," *Acta Otolaryngol. Suppl.*, vol. 338, pp. 1–24, 1976.

[8] S. Davis, "Acoustic characteristics of normal and pathological voices," in *Speech and language: advances in basic research and pratice*, N. Lass, Ed., vol. 1, pp. 271–335. Academic Press, New York, NY, 1979.

[9] Y. Koike, H. Takahashi, and T. Calcaterra, "Acoustic measures for detecting laryngeal pathology," *Ann. Otol. Rhinol. Laryngol.*, vol. 84, pp. 117–124, 1977.

[10] E. Yumoto, W. J. Gould, and T. Baer, "Harmonics-to-noise ratio as an index of the degree of hoarseness," *J. Acoust. Soc. Am.*, vol. 71, pp. 1544–1550, 1982.

[11] H. Kasuya, S. Ogawa, K. Mashima, and S. Ebihara, "Normalised noise energy as an acoustic measure to evaluate pathologic voice," *J. Acoust. Soc. Am.*, vol. 80, pp. 1329–1334, 1986.

[12] W. Costa, S. Costa, V. Vieira, C. Dourado, and T. Araújo, "Análise da dinâmica do sistema de produção vocal na presença de patologias da laringe," in *XI Simpósio Brasileiro de Automação Inteligente*, Fortaleza, Brasil, 2013.

[13] P. Henriquez, J. Alonso, M. Ferrer, and C. Travieso, "Characterization of healthy and pathological voice through measures based on nonlinear dynamics," *IEEE Trans. Audio Speech Lang. Process.*, vol. 17, pp. 1186–1195, 2009.

[14] P. Scalassara, L. Santos, and C. Maciel, "Voice pathology detection with predictable component analysis and wavelet decomposition model," in *IEEE Information Theory Workshop*, Paraty, Brazil, 2011.

[15] R. Kohavi and F. Provost, "Glossary of terms," *Machine Learning*, vol. 30, pp. 271–274, 1998.

[16] T. Murry and E. Doherty, "Selected acoustic characteristics of pathologic and normal speakers," *J. Speech Hear. Res.*, vol. 23, pp. 361–369, 1980.

[17] S. Haykin, *Neural networks and learning machines*, Prentice-Hall, Upper Saddle River, New Jersey, 2008.

[18] D. Jackson, "Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches," *Ecology*, vol. 74, pp. 2204–2214, 1993.

[19] J. Jackson, *A user's guide to principal components*, John Wiley & Sons, Hoboken, New Jersey, 2003.

[20] A. Dibazar, T. Berger, and S. Narayana, "Feature analysis for automatic detection of pathological speech," in *Proceedings of the 23rd IEEE EMBS Annual International Conference*, Istambul, Turkey, 2001, pp. 182–183.

[21] A. Dibazar, T. Berger, and S. Narayana, "Pathological voice assessment," in *Proceedings of the 28th IEEE EMBS Annual International Conference*, New York, USA, 2006, pp. 1669–1673.

Sociedade de Engenharia de Áudio
Artigo de Congresso

Apresentado no 14º Congresso de Engenharia de Áudio
20ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2016, São Paulo, SP

Este artigo foi reproduzido do original final entregue pelo autor, sem edições, correções ou considerações feitas pelo comitê técnico. A AES Brasil não se responsabiliza pelo conteúdo. Outros artigos podem ser adquiridos através da Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA, www.aes.org. Informações sobre a seção Brasileira podem ser obtidas em www.aesbrasil.org. Todos os direitos são reservados. Não é permitida a reprodução total ou parcial deste artigo sem autorização expressa da AES Brasil.

Extração de Descritores Sonoros a partir da Síntese Aditiva Fractal

Thiago Rossi Roque[1] e Rafael Santos Mendes[1]

[1] Universidade Estadual de Campinas, FEEC, DCA
Campinas, São Paulo, 13083-970, Brasil

ra072515@fee.unicamp.br, rafael@dca.fee.unicamp.br

RESUMO

A busca por novas formas de síntese sonora tem se aproveitado dos recentes avanços da *Music Information Retrieval* (MIR). Uma grande contribuição da MIR para a síntese sonora reside nos descritores sonoros, os quais possibilitaram uma quantização de aspectos, muitas vezes abstratos, do que é chamado de timbre. Recentes esforços buscam a utilização dos descritores sonoros como uma forma de manipulação e alteração controlada de aspectos timbrais. Este trabalho pretende dar continuidade a um projeto em curso na busca de novas técnicas de extração de descritores sonoros focadas na possibilidade de ressíntese e manipulações timbrais. Será apresentada aqui a síntese aditiva fractal como base para extração de dois descritores sonoros, o centroide espectral harmônico e o expoente Hurst de banda harmônica.

0 INTRODUÇÃO

O final do século XX marcou o início da era da informação, onde a tecnologia digital se desenvolveu de forma a englobar e modificar praticamente todas as áreas da ciência. O meio musical não foi exceção.

No final da década de 1990, a partir da fusão entre diversas ciências envolvidas com o som, a manipulação de informação e com processamentos de sinais, nasceu a *Music Information Retrieval* (MIR) [1]. Esta ciência consiste no desenvolvimento de técnicas de organização e classificação de conteúdos sonoros. Para tal, a MIR propôs a elaboração de um conjunto de ferramentas capazes de extrair os chamados descritores sonoros que, baseados em estudos psicoacústicos, possibilitam a quantificação de aspectos geralmente qualitativos de um determinado sinal. Ou seja, através dos descritores sonoros têm-se a capacidade de descrever certas características de um sinal de forma objetiva. Técnicas para a extração de diversos descritores sonoros foram desenvolvidos ao longo dos últimos anos e uma vasta biblioteca foi criada por pesquisadores do IRCAM [2].

Inicialmente os descritores sonoros foram criados

unicamente com o objetivo de extrair informações dos sinais de áudio para facilitar sua classificação. Porém, recentemente, pesquisadores enxergaram nesta ferramenta uma forma de possibilitar modificações nos sinais musicais baseadas diretamente em aspectos perceptuais. Esta possibilidade criaria novas formas de um usuário lidar com o objeto sonoro, possibilitando-o tratar e manipular suas características em níveis mais altos de abstração.

Umas das áreas que mais se beneficiaria sobre esta nova possibilidade seria a síntese sonora. Por lidar diretamente com aspectos altamente abstratos do som, como o timbre, a síntese sonora poderia utilizar os descritores sonoros como base para manipulação de parâmetros diretamente perceptuais, dando ao usuário uma maior capacidade de abstração e previsibilidade do resultado sonoro.

Este trabalho tem como objetivo dar continuidade a pesquisas dos autores [3] [4] acerca de novas formas de extração de descritores sonoros baseadas em técnicas de análise/síntese, favorecendo a possibilidade de manipulação de parâmeros timbrais. Serão apresentados os resultados parciais desta pesquisa ainda em curso, envolvendo a extração dos descritores centroide espectral harmônico e expoente Hurst de banda harmônica através da síntese aditiva fractal (FAS).

A FAS é uma recente técnica de codificação de sinais tonais que tem como princípio a análise independente de cada componente harmônica, separando seus conteúdos determinísticos e estocásticos. Seu funcionamento se dá a partir da premissa de que todo sinal tonal apresenta uma componente estocástica manifestada através do perfil das bandas laterais das componentes harmônicas. Detalhes sobre o funcionamento da FAS serão expostos na seção a seguir.

1 SÍNTESE ADITIVA FRACTAL

Concebida por Pietro Polotti e Gianpaolo Evangelista [5], a síntese aditiva fractal compreende uma técnica de codificação de sinais tonais inspirada nos trabalhos de Gregory W. Wornell a respeito da representação de sinais fractais a partir da transformada *wavelet* [6]. Polotti e Evangelista observaram que sinais pseudo-periódicos possuíam em torno de seus harmônicos uma banda lateral com perfil cujo formato se aproxima da função $1/|f - f_k|$, sendo f_k a frequência do k-ésimo harmônico, devido a microflutuações frequenciais presentes em sinais tonais naturais [7]. Dada a ligação entre o espectro de sinais fractais e o perfil de decaimento $1/f$, Polotti e Evangelista adaptaram o conceito criado por Wornell desenvolvendo a transformada *wavelet* de banda harmônica (HBWT) [8] com o objetivo de analisar o perfil harmônico de sinais de áudio. A HBWT tem como objetivo realizar um sistema de análise/síntese de sinais tonais através da separação entre o conteúdo harmônico determinístico, representado pela harmonicidade pura como soma de senoides, e o conteúdo estocástico, representado pelas bandas laterais e seu decaimento $1/f$.

A ideia de codificar sinais tonais em conteúdos estocásticos e determinísticos não é nova. No final da década de 1990 Xavier Serra apresentou a *Spectral Modeling Synthesis* (SMS) [9]. Na SMS o conteúdo determinístico representado pelos harmônicos do sinal musical é codificado por uma soma de cossenoides puras. Já o conteúdo estocástico do sinal é obtido pela diferença entre o modelo determinístico e o sinal original. Dessa forma sinais tonais podem ser representados com grande semelhança pela soma de suas componentes determinísticas $A_k(t)cos(\Theta_k(t))$ e estocásticas $e(t)$, conforme descrito na equação (1), sendo $A_k(t)$ e $\Theta_k(t)$ respectivamente a amplitude e fase instantâneas da k-ésima parcial harmônica.

$$s(t) = \sum_{h=1}^{H} A_h(t)\cos[\Theta_h(t)] + e(t) \qquad (1)$$

Pela sua simplicidade e resultados suficientemente satisfatórios, a SMS foi amplamente difundida e utilizada. Porém por não levar em conta o caráter pseudo-periódico dos sistemas vibrantes naturais, a SMS deixa de codificar o aspecto orgânico do som, deixando-o com as características artificiais de senoides perfeitas.

1.1 Transformada *Wavelet* de Banda Harmônica

É justamente com o objetivo de superar esta fragilidade da técnica desenvolvida por Serra, que Polotti e Evangelista desenvolveram a HBWT e a FAS. Com o objetivo de codificar o decaimento de cada banda lateral de cada harmônico, a HBWT se fundamenta em um processo que se inicia pela separação das bandas laterais através de um banco de filtros para então codificá-las através da transformada *wavelet*. Uma implementação eficiente de um banco de filtros com as características necessárias para a HBWT é dada através da transformada discreta cosseno modulado (MDCT). Um exemplo de como esta transformada é utilizada na HBWT é dado na figura 1.

Uma das vantagens ao utilizar a MDCT na realização deste banco de filtros é a amostragem crítica. Ao trabalhar com a dizimação do sinal em cada canal do banco de filtros, além de remover informações redundantes, a MDCT tem a característica de demodular um sinal passa-faixa transformando-o em banda base [10]. Este processo de demodulação através da reamostragem se torna essencial à HBWT para que a transformada *wavelet* possa analisar as bandas laterais dos harmônicos. Esta demodulação está esquematizada na figura 2.

Como cada banda lateral já separada e devidamente demodulada, se inicia a etapa de análise das bandas. Wornell [6] prova que um processo estocástico $x(t)$ de média temporal espectral $\frac{\sigma_x^2}{|\omega|^\gamma}$ que possua variância finita não nula, pode ser sintetizado através de uma base

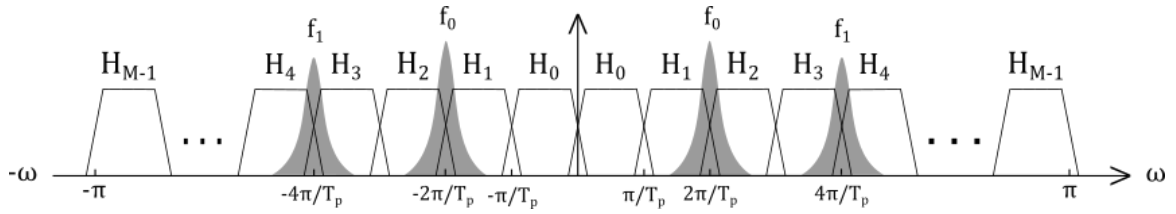


Figura 1: Separação das bandas laterais através da MDCT.

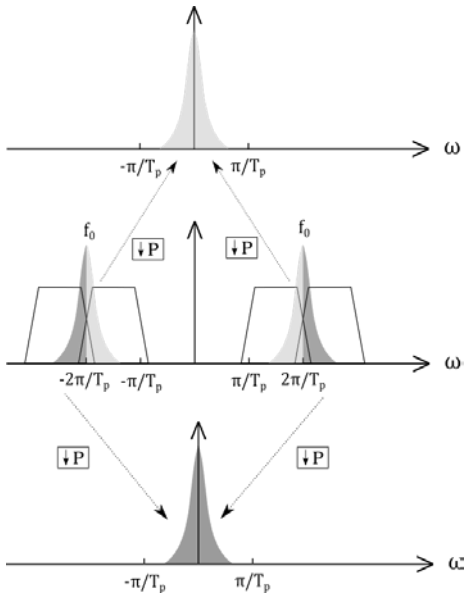


Figura 2: Processo de demodulação das bandas laterais.

wavelet ortonormal $\psi_{n,m}(t)$ e coeficientes $x_n(m)$ mutuamente descorrelacionados de média nula e variância $\sigma^2 2^{-\gamma m}$, conforme descrito pela equação (2). Ou seja, demonstra a semelhança entre a decomposição resultante da transformada *wavelet* discreta e processos estocásticos com perfil espectral $1/f$.

$$x(t) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} x_n(m)\psi_{n,m}(t) \tag{2}$$

Polotti e Evangelista [8] expandiram a dedução (2) para sinais discretos $x(k)$ ciclo estacionários em sentido amplo (WSCS), com período 2^N através da equação (3), sendo $\varphi(k)$ a função escala relativa a função *wavelet* $\psi(k)$, em suas versões discretas.

$$\begin{aligned} x(k) = &\sum_{n=1}^{N} \sum_{m=-\infty}^{\infty} b_n(m)\psi_{n,m}(k) \\ &+ \sum_{m=-\infty}^{\infty} a_N(m)\varphi_{N,m}(k) \end{aligned} \tag{3}$$

A energia espectral média de $x(k)$, apresentado na equação (3), é dada por (4), onde $\Psi(\omega)$ e $\Phi(\omega)$ são as transformadas de Fourier das funções *wavelet* ψ e escala φ respectivamente, γ é o expoente que determina o decaimento do perfil $1/f$ através de $1/f^\gamma$ e σ^2 é a variância do sinal e controla sua energia média.

$$\bar{S}_N(\omega) = \sigma^2 \sum_{n=1}^{N} 2^{n\gamma} \frac{|\Psi_{n,0}(\omega)|^2}{2^n} + 2^{N\gamma} \frac{|\Phi_{N,0}(\omega)|^2}{2^N} \tag{4}$$

A partir de valores apropriados em σ^2 e $2^{n\gamma}$, a representação gráfica de $\bar{S}_N(\omega)$ pode tomar a forma da figura 3, onde cada faixa representa um canal de decomposição da transformada *wavelet* e a resultante da soma dos canais se aproxima do decaimento $1/f$, conforme desejado. Com base na teoria *wavelet* pode se deduzir a relação entre o canal de mais baixa frequência da figura 3 e os coeficientes de aproximação resultantes da função φ_N. Estes coeficientes são utilizados para representar o conteúdo determinístico do sinal, enquanto os canais restantes estão ligados aos os coeficientes de análise decorrentes da função ψ_n e codificarão o conteúdo estocástico presente nas bandas laterais dos harmônicos.

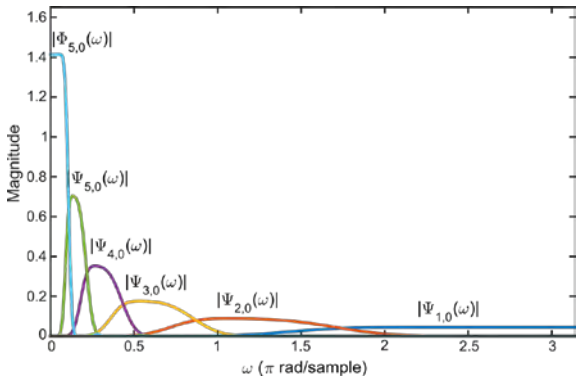


Figura 3: Espectro de $\bar{S}_N(\omega)$ com perfil $1/f$.

Ao unir as duas etapas apresentadas, a HBWT assume a forma representada na figura 4. O sinal $x(t)$ é inserido no banco de filtros MDCT de P canais, onde cada canal irá separar uma única banda lateral de cada harmônico. Após a filtragem o sinal sofre uma dizimação por P responsável por demodular o sinal filtrado transformando-o em banda base. O sinal demodulado é então analisado através da transformada *wavelet*, que resultará nos coeficientes de análise $b_{n,p}$ e nos coeficientes de aproximação $a_{N,p}$, separando assim os conteúdos determinísticos e estocásticos do sinal.

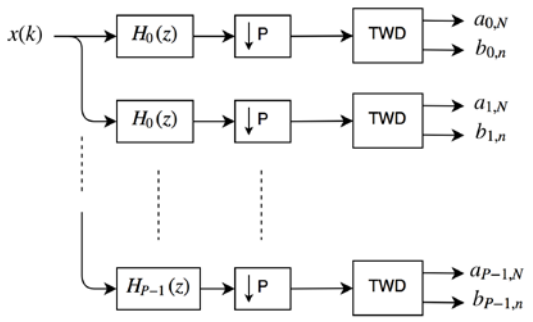


Figura 4: Diagrama da codificação HBWT.

1.2 Codificando os conteúdos determinísticos e estocásticos

Em [11], Polotti e Evangelista propuseram uma forma de parametrizar os coeficientes da HBWT de forma a codificar os coeficientes, reduzindo assim a quantidade de informação necessária para armazenamento do sinal, e ao mesmo tempo simplificando os parâmetros de controle, algo que abriria as possibilidades de manipulação das características sonoras.

Inspirado novamente por Wornell, Polotti notou que os coeficientes de análise $b_{n,p}$, por serem de baixa correlação, não precisariam ser necessariamente armazenados e utilizados, pois um sinal descorrelacionado seria capaz de sintetizar o mesmo sinal [6]. Dado que este sinal descorrelacionado poderia ser facilmente obtido através da geração de um ruído branco, somente é necessário o armazenamento dos coeficientes de aproximação $a_{N,p}$ e o perfil de decaimento de cada banda lateral.

O perfil de decaimento da banda lateral é obtido através de uma regressão linear do logaritmo da variância dos coeficientes $b_{n,p}$, conforme dado pela equação (5)

$$log_2(\mathrm{Var}(b_{n,p})) = \gamma_p n + \text{constante} \tag{5}$$

De forma a melhorar a qualidade sonora da FAS, Polotti propôs sintetizar os coeficientes $b_{n,p}$ através de um sinal de mesma correlação que os coeficientes originais ao invés de um ruído de correlação nula. Tal melhoria é feita através da aplicação de uma codificação linear preditiva (LPC) dos coeficientes $b_{n,p}$, a fim de gerar filtros autorregressivos que possam "colorir" ruídos brancos alterando sua autocorrelação nula em uma próxima do sinal original. Esta alteração proporciona uma considerável melhora na qualidade da FAS a um baixo custo de processamento.

Dado que as bandas laterais direita e esquerda de um mesmo harmônico são codificadas em canais diferentes, é necessário criar um único parâmetro para controlar o conteúdo determinístico de cada harmônico. Para tal é criado um novo coeficiente $c_{N,p}$ pela união dos coeficientes de aproximação das bandas direita e esquerda, $a_{N,2k-1}$ e $a_{N,2k}$, conforme apresentado na equação (6).

$$c_{N,p} = a_{N,2p-1} + ja_{N,2p}. \tag{6}$$

A partir das técnicas acima expostas, a síntese aditiva fractal codifica os coeficientes da HBWT conforme esquematizado na figura 5, onde a componente determinística é codificada em $c_{N,p} = A_{c(k,m)}e^{j\phi_{c(k,m)}}$ e a componente estocástica através dos coeficientes LPC e γ_p, para cada componente harmônica do sinal. Desta forma, a quantidade de informação necessária para representar um sinal tonal comum é reduzida consideravelmente e em um formato prático de manipular.

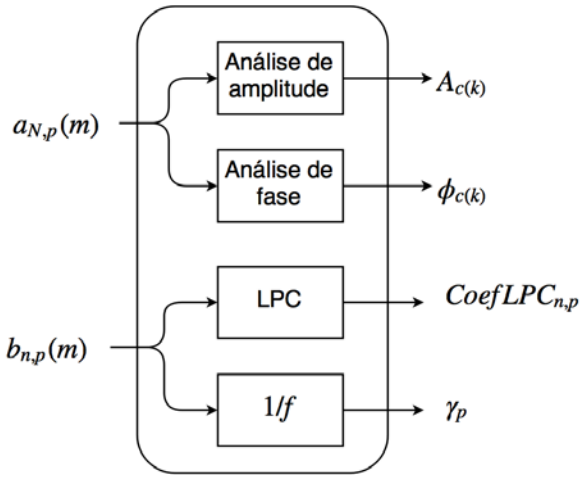


Figura 5: Codificação FAS dos coeficientes HBWT.

Na figura 6 é esquematizado o processo de síntese dos coeficientes HBWT a partir dos parâmetros da FAS.

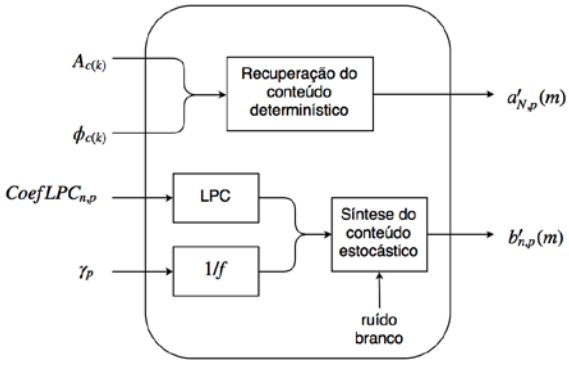


Figura 6: Síntese dos coeficiente HBWT a partir dos parâmetros FAS.

2 DESCRITORES SONOROS

Segundo Jamie Bullock, "um descritor de áudio é qualquer aspecto qualitativa ou quantitativamente mensurável do som" [12]. Desde a criação do conceito de descritores, diversos tipos de descritores sonoros foram

propostos para diferentes áreas de pesquisa, como por exemplo reconhecimento de fala, classificação de instrumentos musicais, classificação de gêneros musicais e psicoacústica [13]. Em 2003 o projeto CUIDADO, liderado pelo IRCAM, realizou um extenso trabalho acerca dos descritores sonoros com o objetivo de compilar as diversas técnicas desenvolvidas organizando-as por categorias, além de expandir a padronização do uso dos descritores feita pelo Moving Picture Experts Group (MPEG) através do MPEG-7.

Durante o projeto CUIDADO foram definidos 72 descritores sonoros. Este conjunto de descritores pode ser dividido em 6 tipos: descritores temporais, descritores de energia, descritores espectrais, descritores harmônicos, descritores perceptuais e descritores variados. Para cada tipo de descritor existem 2 seções: descritores globais e descritores instantâneos. Descritores globais são extraídos do som por inteiro, de seu início ao seu fim. Já descritores instantâneos estão relacionados a segmentos do sinal e são extraídos a partir de janelas temporais de curta duração.

Conforme foi exposto, as técnicas envolvendo os descritores sonoros foram desenvolvidas inicialmente com o objetivo de extração de parâmetros apenas. Porém recentemente algumas pesquisas visaram utilizar tais conceitos com o objetivo de realizar síntese sonora [14], [15], [16] e [17].

Conforme a pesquisa em que este trabalho está inserido, pretende-se apresentar neste trabalho os resultados parciais com relação a extração de dois descritores sonoros a partir dos parâmetros da FAS. Neste primeiro momento da pesquisa optou-se pelos descritores "centroide espectral harmônico", "expoente Hurst de banda harmônica" e "razão entre energia de harmônicos ímpares e pares". Pretende-se em breve expandir a lista dos descritores obtidos pela FAS e realizar não somente a extração destes descritores como sua modulação, obtendo assim uma técnica de síntese capaz de modificações timbrais a partir dos descritores sonoros.

Uma breve explanação sobre os descritores utilizados será dada a seguir.

2.1 Centroide espectral harmônico

O centroide espectral harmônico (CEH) pode ser interpretado como o centro de massa do espectro, dado que seu cálculo se baseia na média frequencial de suas componentes harmônicas ponderadas por suas amplitudes. Estudos sugerem uma alta correlação entre o valor do centroide espectral e a sensação de brilho de um sinal [18]. Após adaptado para seu cálculo para a FAS, a extração desse descritor se dá através da equação (7)

$$CEH = \frac{\sum_{k=1}^{K} A_c(k,n) f(k)}{\sum_{k=1}^{K} A_c(k,n)} \tag{7}$$

2.2 Expoente Hurst de banda harmônica médio

Originalmente o expoente Hurst foi desenvolvido como uma medida do grau de auto-similaridade em processos estocásticos. O perfil de decaimento $1/f$, observado nas bandas laterais dos harmônicos, pode ser considerado um caso especial de processos estocásticos auto-similares não estacionários, portanto, podem ser modelados por movimentos Brownianos fracionário (fBm).

Em "Análise de Instrumentos Musicais através do Expoente Hurst de Banda Harmônica - Estudo Comparativo da Quena e de outros Instrumentos de Sopro" [19], Aldo André Díaz Salazar propõe a utilização do expoente Hurst para a caracterização de instrumentos musicais pela análise do perfil do decaimento das bandas laterais dos harmônicos através da HBWT.

Ao apresentar o cálculo do expoente Hurst através da HBWT, Díaz propôs um novo descritor nomeado de expoente Hurst de banda harmônica (EHBH), com o objetivo de caracterização de instrumentos musicais. Inspirado pelo estudos de Díaz, é proposto aqui o expoente Hurst de banda harmônica médio (EHBHM), um descritor global que possa fornecer um único parâmetro para todo o som.

O cálculo do expoente Hurst de banda harmônica, H_p, é feito através do parâmetro γ_p, apresentado na equação (5), e da equação (8) [19].

$$H_p = \frac{\gamma_p - 1}{2} \tag{8}$$

A partir do expoente Hurst de banda harmônica é definido o expoente Hurst de banda harmônica médio, $\bar{H}$, pela utilização do valor médio $\bar{\gamma}$ de γ_p ao longo de p, dado pela equação (9).

$$\bar{H} = \frac{\bar{\gamma} - 1}{2} \tag{9}$$

Através de experimentos, constatou-se que harmônicos com baixa energia possuem tendências a apresentar valores de γ próximos a zero na medida que sua energia se aproxima da energia do ruído de fundo. Tal efeito tende a trazer o valor médio de γ para baixo dos valores esperados, causando uma distorção na medida do descritor $\bar{H}$. Este problema foi solucionado pela inclusão de um expoente em γ_p no cálculo de $\bar{\gamma}$, reduzindo o peso dos γ relacionados a harmônicos de baixa energia, conforme dado pela equação (10). O valor $1,2$, do expoente de γ_p, foi definido empiricamente.

$$\bar{\gamma} = \frac{1}{P} \sum_{p=0}^{P} \gamma_p^{1,2}, \tag{10}$$

2.3 Razão entre energia de harmônicos ímpares e pares

Conforme o nome sugere, a razão entre energia de harmônicos ímpares e pares (RHIP) fornece uma medida da proporção dos harmônicos pares para os harmônicos ímpares no sinal. Este descritor é importante para uma parametrização do perfil espectral de um som, assim como para classificação de certos instrumentos musicais, como por exemplo instrumentos da família dos aerofones de tubo fechado. Uma diferença notável na razão entre a energia de harmônicos ímpares e pares também é observada em sistemas não-lineares com funções de transferência ímpar, como é o caso dos amplificadores de guitarras elétricas que, quando operados em classe AB e levados a saturação, geram distorções muitas vezes apreciadas e utilizadas como recurso estético e expressivo.

O cálculo do descritor RHIP é apresentado pela equação (11)

$$RHIP = \frac{\sum_{k=1:2:K} \mathrm{Var}(A_c(k,n))^2}{\sum_{k=2:2:K} \mathrm{Var}(A_c(k,n))^2} \tag{11}$$

3 METODOLOGIA

A implementação dos algoritmos da FAS e da extração dos descritores foi feita na ferramenta de cálculo matemático *MATLAB*, onde obteve-se uma estrutura de análise/síntese conforme a figura 7.

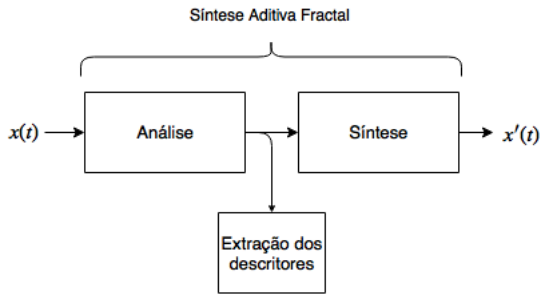


Figura 7: Algoritmo implementado.

Para validar o cálculo destes dois descritores através da FAS, foram feitos experimentos com amostras de instrumentos musicais reais. Estas amostras foram obtidas na *University of Iowa Musical Instrument Samples*, uma vasta biblioteca digital de amostras sonoras de instrumentos musicais [20].

Inicialmente foi validada a implementação do processo de análise/síntese da FAS através da comparação entre o sinal original e o sinal sintetizado. Somente após a confirmação deste processo foram feitos experimentos acerca da extração dos descritores.

A validação da extração do CEH foi feita seguindo a mesma metodologia de trabalhos anteriores [4], onde os valores obtidos pela técnica aqui apresentada são comparados com resultados gerados através de técnicas consagradas fundamentadas na transformada de Fourier. Para a extração deste descritor através da transformada de Fourier foi utilizada a ferramenta on-line criada pelo *Institut für Telekommunikationssysteme* da *Technische Universität Berlin* [21].

Similarmente, para o RHIP, os valores obtidos através da FAS foram confrontados com os calculados através da transformada de Fourier seguindo a formulação proposta por Peeters [2].

Para o caso do expoente Hurst de banda harmônica, dado o fato deste descritor ser muito recente, ainda não há uma técnica já consagrada. Os valores obtidos neste trabalho serão comparados com os obtidos por Díaz.

Para todos os descritores, foram utilizadas amostras de quatro instrumentos: um oboé (B3, *pianissimo*), um violoncelo (A3, *fortissimo*), um trompete (D3, *fortissimo*) e um clarinete (G3, *fortissimo*). Em todos os casos a FAS foi implementada para análise dos primeiros 40 harmônicos.

4 RESULTADOS

Após a implementação do algoritmo da FAS, foram feitos experimentos acerca de sua validação como técnica de análise/síntese. Os resultados obtidos são expostos a seguir. Nas figuras 8 e 9 estão representados respectivamente o formato de onda e o espectro da amostra original utilizada (amostra de uma nota A3 de um oboé). Já as figuras 10 e 11 foram obtidas após o processo de análise e síntese da FAS.

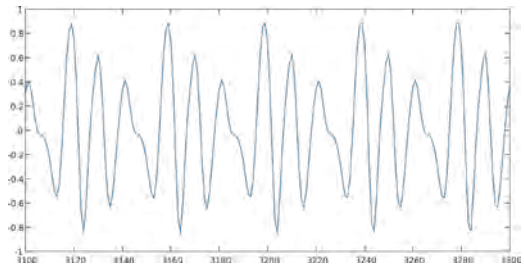


Figura 8: Trecho da forma de onda da amostra original.

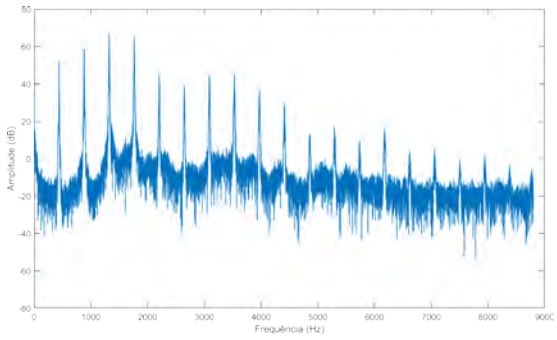


Figura 9: Espectro original da amostra de um Oboé.

Através destas imagens nota-se considerável semelhança entre o sinal original e o sinal sintetizado.

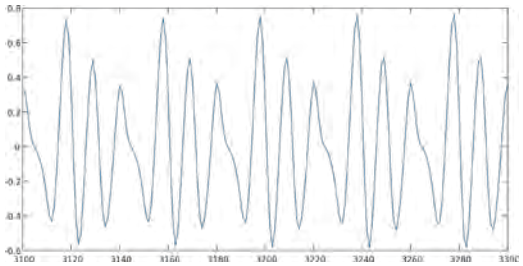


Figura 10: Trecho da forma de onda da amostra ressintetizada.

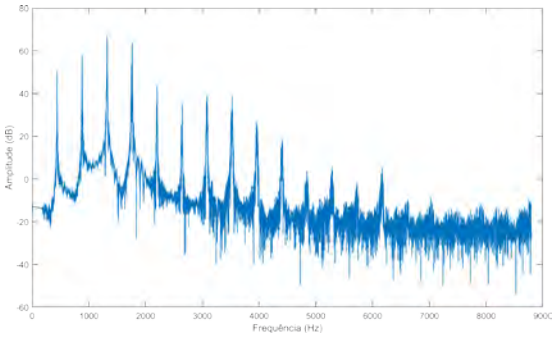


Figura 11: Espectro ressintetizado da amostra de um Oboé.

Testes auditivos apresentaram uma diferença levemente notável, porém sem grande prejuízo às características tonais do instrumento.

A tabela 1 explicita os resultados obtidos para o centroide espectral harmônico, tanto a partir da FAS quanto de Fourier.

Instrumento	FAS	Fourier
Oboé	1096 Hz	1013 Hz
Violoncelo	1130 Hz	1007 Hz
Trompete	1961 Hz	1844 Hz
Clarinete	960 Hz	726 Hz

Tabela 1: Comparativo de resultados para o centroide espectral harmônico.

Apesar dos valores obtidos na FAS serem mais altos (cerca de 100 Hz acima), comparados com os de Fourier, nota-se uma clara correlação entre os valores, afirmando seu funcionamento.

Para o expoente Hurst de banda harmônica médio, os resultados obtidos estão na tabela 2

Instrumento	EHBHM
Oboé	0,4425
Violoncelo	0,7528
Trompete	1,0756
Clarinete	0,4204

Tabela 2: Comparativo de resultados para o centroide espectral harmônico.

A partir dos trabalhos de Díaz, pôde-se estimar valores médios de H em torno de 0,2 para a quena, 0,4 para a escaleta e 0,9 para a flauta doce. Ao comparar os valores obtidos nesse trabalho com os observados na literatura, pôde-se notar considerável variação de valores. Era de se esperar que valores do EHBHM fossem próximos para instrumentos com características similares, como a quena e a flauta doce. Porém alguma semelhança pode ser observada entre o oboé e a clarineta.

Conforme descrito na seção 2.2, as bandas laterais dos harmônicos são modelados como processos fBm. Portanto, de acordo com Díaz e Mendes [22], é esperado que o valor de EHBH se encontre dentro do intervalo entre 0 e 1, sendo valores mais altos relacionados com sinais puramente harmônicos e valores mais baixos relacionados com sinais de alto conteúdo estocástico. Considerando uma certa margem, para o caso do trompete, nota-se que os valores aqui obtidos se encontraram dentro do intervalo esperado, confirmando a teoria adotada. Porém mais estudos são necessários para uma melhor compreensão deste descritor.

A tabela 3 apresenta o comparativo dos resultados obtidos para o descritor RHIP.

Instrumento	FAS	Fourier
Oboé	0.8685	0.8805
Violoncelo	1.1330	1.1143
Trompete	0.9766	0.9524
Clarinete	0.5677	0.6477

Tabela 3: Comparativo de resultados para a razão de energia entre os harmônicos pares e ímpares.

Através da análise da tabela 3, pode-se notar uma grande semelhança entre os resultados das duas técnicas apresentadas.

5 CONCLUSÃO

Através da análise dos resultados obtidos pode-se confirmar o funcionamento do algoritmo aqui proposto acerca da síntese aditiva fractal e da possibilidade de extração de descritores sonoros através desta técnica. Nota-se a necessidade de um maior aprofundamento nestas técnicas, bem como uma maior quantidade de experimentos para sua validação.

Com relação ao expoente Hurst de banda harmônica, nota-se uma necessidade de aprofundamento em seus estudos, principalmente com relação ao comportamento deste parâmetro em uma maior gama de instrumentos musicais e em variadas notas.

Como próximos passos pretende-se estudar o significado sensorial do expoente Hurst de banda harmônica, incluir novos descritores e desenvolver uma forma de modulação dos descritores a partir dos parâmetros FAS.

AGRADECIMENTOSS

Os autores agradecem à CAPES por financiarem esta pesquisa.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Michael Fingerhut, "Music information retrieval, or how to search for (and maybe find) music and do away with incipits," *IAML-IASA Congress, Oslo*, August 8–13 2004.

[2] Geoffroy Peeters, "A large set of audio features for sound description (similarity and classification) in the cuidado project," Tech. Rep., IRCAM, 2004.

[3] Thiago Rossi Roque and Rafael Santos Mendes, "Extração de centroide espectral através da tranformada wavelet packet," in *Proceedings of the SPS 2013*, 2013.

[4] Thiago Rossi Roque and Rafael Santos Mendes, "Extração de descritores sonoros timbrísticos a partir da transformada wavelet packet," in *Proceedings of the 12th AES Brazil Conference, São Paulo, Brazil, May 13-15, 2014*, 2014, pp. 39–46.

[5] Pietro Polotti, *Fractal Additive Synthesis: Spectral Modeling of Sound for Low Rate Coding of Quality Audio*, Ph.D. thesis, École Polytechnique Fédérale de Lausanne, 2003.

[6] Gregory W. Wornell, "Wavelet-based representations for the 1/f family of fractal processes," *Proceedings of the IEEE*, vol. 81, no. 10, pp. 1428–1450, Oct 1993.

[7] Pietro Polotti and Gianpaolo Evangelista, "Multiresolution sinusoidal/stochastic model for voiced-sounds," in *Proceedings of the COST G-6 conference on digital audio effects (DAFX-01), Limerick, Ireland*, December 6-8 2001, pp. 120–124.

[8] Pietro Polotti and Gianpaolo Evangelista, "Analysis and synthesis of pseudo-periodic1/f-like noise by means of wavelets with applications to digital audio," *EURASIP J. Appl. Signal Process.*, vol. 2001, no. 1, pp. 1–14, january 2001.

[9] Xavier Serra, *A System for Sound Analysis/Transformation/Synthesis based on a Deterministic plus Stochastic Decomposition*, Ph.D. thesis, Stanford University, 1989.

[10] Richard G. Lyons, *Understanding Digital Signal Processing (3d Edition)*, chapter 13, pp. 781–783, Prentice Hall PTR, 2010.

[11] Pietro Polotti and Gianpaolo Evangelista, "Fractal additive synthesis via harmonic-band wavelets," *Comput. Music J.*, vol. 25, no. 3, pp. 22–37, Sept. 2001.

[12] Jamie Bullock and UCEB Conservatoire, "Libxtract: A lightweight library for audio feature extraction," in *Proceedings of the International Computer Music Conference*. Citeseer, 2007, vol. 43.

[13] Geoffroy Peeters, "Automatic classification of large musical instrument databases using hierarchical classifiers with inertia ratio maximization," in *Audio Engineering Society Convention 115*. Audio Engineering Society, 2003.

[14] Matt Hoffman and Perry R. Cook, "Feature-based synthesis: Mapping acoustic and perceptual features onto synthesis parameters," in *in Proceedings of the International Computer Music Conference, New Orleans*, 2006.

[15] Xavier Serra and Jordi Bonada, "Sound transformations based on the sms high level attributes," in *Proceedings of the Digital Audio Effects Workshop (DAFX98)*, 1998.

[16] Diemo Schwarz, "Corpus-based concatenative synthesis," *Signal Processing Magazine, IEEE*, vol. 24, no. 2, pp. 92–104, 2007.

[17] Tae Hong Park, Jonathan Biguenet, Zhiye Li, Conner Richardson, and Travis Scharr, "Feature modulation synthesis (fms)," in *Proc. ICMC*, 2007, vol. 2007.

[18] Emery Schubert, Joe Wolfe, and Alex Tarnopolsky, "Spectral centroid and timbre in complex, multiple instrumental textures," in *Proc. of the 8th International Conference on Music Perception & Cognition (ICMPC)*, 2004.

[19] Aldo André Díaz Salazar, "Análise de instrumentos musicais através do expoente hurst de banda harmônica - estudo comparativo da quena e de outros instrumentos de sopro," M.S. thesis, Faculdade de Engenharia Elétrica, UNICAMP, Campinas, Brasil, 2015.

[20] "Electronic music studios, University of Iowa," http://theremin.music.uiowa.edu/MIS.piano.html, Online; acessado em 28/10/2015.

[21] Amjad Samour, Hyoung-Gook Kim, Juan José Burred, and Martin Haller, "Mpeg-7 audio analyzer low level descriptors extractor," http://mpeg7lld.nue.tu-berlin.de/, 12 2003.

[22] Aldo Díaz and Rafael Mendes, "Analysis/synthesis of the andean quena via harmonic band wavelet transform," in *Proceedings of the 18th International Conference on Digital Audio Effects (DAFx-15), November 30 — December 3*, Trondheim, Norway, 2015, pp. 433–437.

Sociedade de Engenharia de Áudio
Artigo de Congresso

Apresentado no 14º Congresso de Engenharia de Áudio
20ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2016, São Paulo, SP

Este artigo foi reproduzido do original final entregue pelo autor, sem edições, correções ou considerações feitas pelo comitê técnico. A AES Brasil não se responsabiliza pelo conteúdo. Outros artigos podem ser adquiridos através da Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA, www.aes.org. Informações sobre a seção Brasileira podem ser obtidas em www.aesbrasil.org. Todos os direitos são reservados. Não é permitida a reprodução total ou parcial deste artigo sem autorização expressa da AES Brasil.

An improved method for determining pitch relevance in polyphonic audio

Isabela F. Apolinário,[1] and Luiz W. P. Biscainho[1]

[1] Federal University of Rio de Janeiro, COPPE, Electrical Engineering Program
Rio de Janeiro, Brazil

isabela.apolinario@smt.ufrj.br, wagner@smt.ufrj.br

ABSTRACT

The main purpose of this article is to present an improved version of a recently introduced timbre independent salience function here named as DLMP, which has been developed in the context of audio analysis, specifically to estimate the relevance of the pitches found in polyphonic signals. Based on the results attainable by the DLMP method, some modifications are proposed. They are presented along with the original salience computation. The performances of the original and modified DLMP (MDLMP) are compared through a set of examples.

0 INTRODUCTION

In the context of audio signal analysis, methods for multiple-pitch tracking are very relevant and constitute a dynamic research topic. Applications such as automatic music transcription [1], sound source separation [2], and principal melody detection [3], [4] can profit from such an analysis, since their performances rely highly on accurate pitch detection. Usually, parametric information extracted from audio signals is used to estimate pitch tracks, as is done in sinusoidal modeling [5]. An alternative strategy is to compute a "salience function" to quantify the relevance of the most predominant pitches in the signal.

The classical approach for the computation of the salience function of a given frequency f_0 considers the amount of energy present in each harmonic partial (multiples of f_0). This computation may be written as [3]

$$\rho(f_0) = \frac{1}{n_{\mathrm{H}}} \sum_{i=1}^{n_{\mathrm{H}}} \log |X(if_0)|, \tag{1}$$

where $X(f)$ is the spectrum of a discrete signal $x(n)$ and n_{H} a pre-determined number of harmonic partials. It can be noticed that Equation (1) uses information about the amplitude of harmonic partials and, consequently, about the timbre of the sound source.

The method proposed in [6] uses the information of frequency localization. First, the peaks of the spectrum are estimated, and then the possibility that each one is

a valid partial of some existing fundamental frequency is evaluated. A theoretical sequence of deviations is defined, based on the distance between the observed peak frequencies and the notes in a 12-note equal tempered scale [6].

Figure 1 shows the frequency localization (continuous line) of the fundamental frequencies of such a scale for a 440-Hz reference frequency[1]. It is also possible to see (dashed line) the frequency corresponding to a C4 note ($f_{C4} = 261.6$ Hz and MIDI number 6) and its superior harmonic partials. One can notice that there is a subtle difference between the localization of the notes composing the scale and the harmonic partials of the C4 note: they are not perfectly aligned.

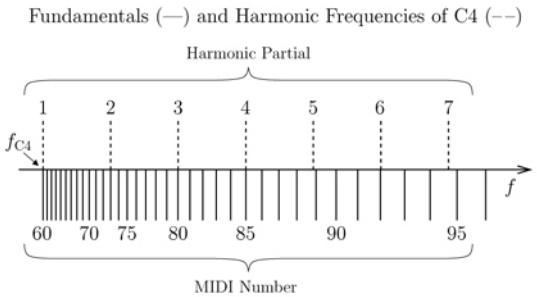


Figure 1: Frequency localization of the fundamentals of an equal tempered scale for a 440-Hz reference frequency (continuous line) and frequency localization of the harmonic partials for a 261.6-Hz fundamental (dashed line).

This article is organized as follows. Section 1 briefly explains the stages of the DLMP method. Section 2 details the computation of the DLMP salience function and each proposed modification. In a set of simulations, Section 3 presents some results obtained by the proposed MDLMP and compares them to those obtained by the original DLMP method. Finally, Section 4 draws some conclusions and future work.

1 OVERVIEW

Figure 2 shows the block diagram of the addressed DLMP salience computation method. The input signal $x(n)$ has L samples, i.e., $n = 0, 1, \ldots, L-1$. In this work, it will be assumed that $x(n)$ is a discrete audio signal. The first stage is to compute a time-frequency transform of $x(n)$, $X(k, j)$, where indexes $k = 0, 1, \ldots, K-1$ and $j = 0, 1, \ldots, J-1$ are the time and frequency axes, respectively. The number of samples used in the analysis window is N, and the hop size is τ. As in [6], the Short-Time Fourier Transform with $N = 4096$ samples (corresponding to 92.9 ms for a 44.1 kHz sampling frequency), $\tau = 2048$ samples (50% overlap), and a Hanning analysis window is adopted as the time-frequency transformation. It should be noted, however, that these values are not optimum (in the sense of the performance of the method) for all types of signals and can be changed accordingly.

[1] See details in Section 2.2.1.

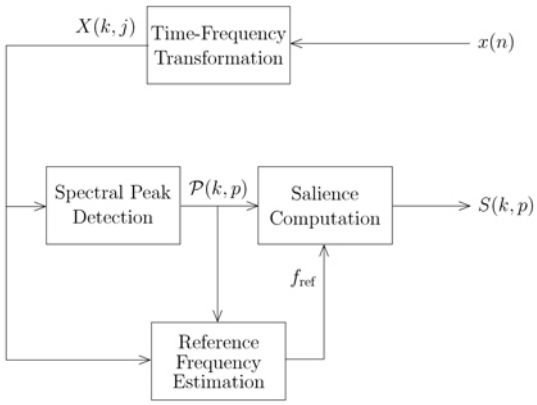


Figure 2: Block diagram of the DLMP salience computation method.

The next step is the spectral peak detection, resulting in the signal $\mathcal{P}(k, p)$, with $p = 0, 1, \ldots, P-1$. For each time frame k_0, $\mathcal{P}(k_0, p) = \{(f_0(k_0), a_0(k_0)), \ldots, (f_{P-1}(k_0), a_{P-1}(k_0))\}$, where $f_p(k)$ and $a_p(k)$ are respectively frequency and magnitude of the peak p in frame k. A value of $P = 80$ peaks per frame is chosen.

The obtained transform coefficients $X(k, j)$ and detected peaks $\mathcal{P}(k, p)$ are then utilized to estimate the reference frequency f_{ref}, essential to compute the frequencies of the notes that compose the equal tempered scale. The procedure is described in [7] and will not be detailed here.

Finally, to compute the salience function $S(k, p)$ for each frame k and peak p, the frequency locations of the peak p and each note from the equal tempered scale with reference frequency f_{ref} are compared. These frequency differences obey a certain law and can, therefore, be used as a theoretical measure to determine which of the detected peaks correspond to sound sources and which to harmonic partials.

2 DLMP AND MDLMP METHODS

In this section, the details of each step of the DLMP method will be explained followed by the proposed modifications.

2.1 Spectral Peak Detection

The strategy used for peak detection [8] searches for a peak as a local maximum: the magnitude of a peak is taken as the maximum value within a certain frequency range if higher than a predetermined threshold. The threshold value is initialized with the maximum spectrum magnitude and is decreased until the predefined maximum number of peaks ($P = 80$) is reached. To increase the frequency resolution of the detected peaks, a three-point parabolic interpolation among adjacent bins is performed [8].

The main proposed modifications are in the salience computation stage. However, a single modification was implemented in the spectral peak detection in order to improve the accuracy of the estimated peak frequencies.

Proposed Modification I: Removing noise floor by the SSE Method

The DLMP method is restricted to the procedure described above. However, this peak detection method takes into account the magnitude of the signal spectrum, which, due to audio signal's nature, usually decreases in frequency. Therefore, the information contained in higher partials is nearly not taken into consideration. To avoid this problem, a pre-processing step is added, where the main idea is to "flatten" the magnitude spectrum. The noise floor is initially estimated by means of the Stochastic Spectrum Estimator (SSE) method (detailed in [9]), and then subtracted from the original magnitude spectrum (in logarithmic scale). The peaks are estimated from the obtained flattened spectrum.

2.2 Salience Function Computation

2.2.1 Defining the Theoretical Deviation Sequence

As said before, it is necessary to define a theoretical deviation between some random frequency f_p and another belonging to the frequency grid of an equal tempered scale. It will be considered that each semitone is divided in 100 equal intervals in a geometric scale, each of which designate a cent, i.e, each octave is divided in 1200 cents [10].

Since the scale considered in this work is the equal tempered scale, each fundamental F_i belonging to this scale may be written as

$$F_i = f_{\text{ref}} 2^{\frac{1}{12}}, \tag{2}$$

where f_{ref} will be estimated, and i corresponds to the note MIDI number[2] without the offset of 69 (for simplicity, it is assumed that the A4 MIDI number is $i = 0$ instead of $i = 69$).

The audio signal $x(n)$ is assumed to be formed of sound sources with fundamental frequency F_j and harmonic partials $f_h^{F_j}$. Given a fundamental F_0, its harmonic partials are given by

$$f_h^{F_0} = hF_0. \tag{3}$$

The deviations in cents of each partial $f_h^{F_0}$, with $h = 2, 3, \ldots, n_{\text{H}}$, from the fundamentals of the equal tempered scale are defined as [6]

$$d_h^{F_0} = 100\left[12\log_2\left(\frac{f_h^{F_0}}{f_{\text{ref}}}\right) - \left\lfloor 12\log_2\left(\frac{f_h^{F_0}}{f_{\text{ref}}}\right)\right\rceil\right], \tag{4}$$

where $\lfloor\cdot\rceil$ is the rounding operator and f_{ref} is the A4 tuning frequency estimated from the signal. The sequence of all harmonic partials of F_0 is denoted by $\{f_h^{F_0}\}_h$, and the sequence of theoretical deviations by $\{d_h^{F_0}\}_h$.

By replacing Equations (2) and (3) in Equation (4), one obtains

$$d_h^{F_i} = 100[12\log_2(h) - \lfloor 12\log_2(h)\rceil]. \tag{5}$$

It can be noticed that the deviations sequence $\{d_h^{F_i}\}_h$ does not depend anymore on the fundamental F_i.

Proposed Modification II: Removing Note Quantization in Theoretical Sequence

In the original DLMP technique, as proposed in [6], the author uses Equation (5) as the reference sequence. However, in this work, the initial definition of theoretical deviations, given by Equation (4), will also be considered as a valid option. This modification is necessary to correctly model frequency variations in the existing fundamentals F_j; by doing this, one avoids to quantizes them only into fundamental frequencies F_i of the equal tempered scale.

2.2.2 Salience Computation

A sequence of observed deviations $\{\hat{d}_h^{f_p}\}_h$ should also be obtained for each peak p and frame k, as will be explained in the next section, where f_p is the frequency of peak p. It is important to recall that, according to Modification I, frequency f_p will be used instead of F_0 in the definition of the sequence of theoretical deviations $\{d_h^{F_0}\}_h$.

The salience $S(k, p)$ is computed for each peak p of each frame k. Here, for simplicity, the dependence on k will be omitted, resulting in the notation $S(p)$. The idea is to use the correlation between the theoretical and observed sequences ($\{d_h^{f_p}\}_h$ and $\{\hat{d}_h^{f_p}\}_h$, respectively) as a similarity method, ie.,

$$S(p) = a_p \mathrm{C}(\{d_h^{f_p}\}_h, \{\hat{d}_h^{f_p}\}_h), \tag{6}$$

where $\mathrm{C}(\cdot,\cdot)$ is a generic correlation measure, and a_p is the amplitude of peak p. Both sequences can be written in vector format as $\mathbf{d}^{f_p} = [d_2^{f_p}\ d_3^{f_p}\ \ldots\ d_{n_{\text{H}}}^{f_p}]^{\text{T}}$ and $\hat{\mathbf{d}}^{f_p} = [\hat{d}_2^{f_p}\ \hat{d}_3^{f_p}\ \ldots\ \hat{d}_{n_{\text{H}}}^{f_p}]^{\text{T}}$, respectively.

Proposed Modification III: Using Correlation Coefficient as Correlation Measure

The correlation method used originally in [6] is the inner product between the sequences. This choice, however, directly compares the absolute values of both sequences $\mathbf{d}^{f_p}$ and $\hat{\mathbf{d}}^{f_p}$. A better procedure would be to compare their relative variations; to this end, a pre-processing stage of normalization forces zero mean and unit standard deviation on both sequences, thus substituting their correlation coefficient for their correlation, defined as

$$r(p) = \frac{\langle \mathbf{d}^{f_p} - \boldsymbol{\mu}_{\text{teo}}, \hat{\mathbf{d}}^{f_p} - \boldsymbol{\mu}_{\text{obs}}\rangle}{\sqrt{\sigma_{\text{teo}}^2}\sqrt{\sigma_{\text{obs}}^2}}, \tag{7}$$

where $\boldsymbol{\mu}_{\text{teo}}$ and $\boldsymbol{\mu}_{\text{obs}}$ are vectors composed of $n_{\text{H}} - 1$ elements equal to the means μ_{teo} and μ_{obs} of vectors $\mathbf{d}^{f_p}$

[2] MIDI standard numbers notes from 0 (C0) to 127 (G10); A4 corresponds to 69.

and $\hat{\mathbf{d}}^{f_p}$, respectively; σ^2_{teo} and σ^2_{obs} are the variances of $\mathbf{d}^{f_p}$ and $\hat{\mathbf{d}}^{f_p}$, respectively; and $\langle \cdot, \cdot \rangle$ is the inner product operator. Now, the salience value is obtained from the correlation coefficient as $S(p) = a_p r(p)$.

2.2.3 Detecting Peaks that correspond to Harmonic Partials

The sequence of harmonic partials of peak p is denoted as $\{f_h^{f_p}\}_h$. Each $f_h^{f_p}$ value is given by Equation (3). The sequence $\{\hat{d}_h^{f_p}\}_h$ is composed of the observed deviations corresponding to each frequency of $\{f_h^{f_p}\}_h$. It is computed for each peak p of each frame k, i.e., all detected peaks are considered as possible candidates to fundamental frequency.

The preliminary step to decide whether a frequency f_p should be considered a fundamental is to analyze the remaining detected peaks p' (with corresponding frequencies $f_{p'}$) to determine which of them correspond to harmonic partials of f_p. This is done by using a Gaussian function G centered in the $h-$th harmonic partial of f_p, i.e., with mean $\mu_{h,p} = hf_p$, and evaluated at the other detected peaks' frequencies $f_{p'}$. The standard deviation of G is defined by $\sigma_{h,p} = \mu_{h,p}\left(2^{\frac{\kappa}{1200}} - 1\right)$, where $\kappa = 20$ was chosen experimentally in order to minimize the effect of the error introduced in the peak detection stage [6]. The idea is to allow a certain tolerance in the frequency position of the peaks.

Figure 3 shows an example for the case in which the first peak ($p = 0$), whose frequency f_0 matches the fundamental F_0, is analysed. The frequencies $f_{p'}$ of other peaks are given by f_1, f_2, f_3, f_4, and f_5. Here, only the five first harmonic partials will be analysed, i.e., $n_{\text{H}} = 6$. For this case, Gaussian functions with means and standard deviations given by $\mu_{h,0} = hf_0$ and $\sigma_{h,0} = \mu_{h,0}\left(2^{\frac{\kappa}{1200}} - 1\right)$, for $h = 2, 3, \ldots, 6$, respectively, are generated. Then, the values of G are evaluated at frequency positions $f_{p'}$, that is, $G(f_{p'}; \mu_{h,0}, \sigma_{h,0})$, with $p' = 1, 2, \ldots, 5$. In the figure, it is possible to see the Gaussian function for the case $h = 3$ and its value at $f_{p'}$, with $p' = 2$, i.e., $G(f_2; \mu_{3,0}, \sigma_{3,0})$. For the remaining values of p', $G(f_{p'}; \mu_{3,0}, \sigma_{3,0}) \to 0$. A vector of the form

$$\mathbf{G}_{\mu_{3,0},\sigma_{3,0}} \approx [0 \;\; G(f_2; \mu_{3,0}, \sigma_{3,0}) \;\; 0 \;\; 0 \;\; 0]^{\text{T}} \quad (8)$$

is, therefore, generated.

Proposed Modification IV: Forcing Impulse-Like Shape to Detect valid Peaks

In this work, we propose a different approach: forcing the equality sign in Equation (8), resulting in a vector $\hat{\mathbf{G}}_{\mu_{h,p},\sigma_{h,p}}$, whose only nonzero element is the maximum of $\mathbf{G}_{\mu_{h,p},\sigma_{h,p}}$. This is adopted because, in some cases, spurious peaks can be detected next to a valid harmonic partial that will not result in values next to zero when evaluated by the Gaussian function.

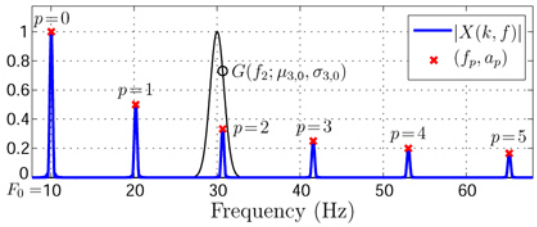


Figure 3: Peak selection by means of a Gaussian function G with mean and variance $\mu_{h,p}$ and $\sigma_{h,p}$, respectively, evaluated at frequencies $f_{p'}$ (here, $p' = 2$, $h = 3$, and $p = 0$).

2.2.4 Computing the Observed Deviation Sequence

The next step to compute the sequence $\{\hat{d}_h^{f_p}\}_h$ is to associate, for each peak p', deviations $d_{p'}$ as defined in Equation (4), i.e.,

$$d_{p'} = 100\left[12\log_2\left(\frac{f_{p'}}{f_{\text{ref}}}\right) - 12\log_2\left\lfloor\frac{f_{p'}}{f_{\text{ref}}}\right\rceil\right]. \quad (9)$$

In the example of Figure 3, vector $\mathbf{d}$ was generated by the deviations of peak $p' = 1, 2, \ldots, 5$, i.e., $\mathbf{d} = [d_{f_1} \;\; d_{f_2} \;\; \ldots \;\; d_{f_5}]^{\text{T}}$.

Lastly, the deviation $\hat{d}_h^{f_p}$ corresponding to the harmonic partial h of peak p (and whose frequency is given by $f_h^{f_p} = hf_p$) is computed as

$$\begin{aligned} \hat{d}_h^{f_p} &= \langle \mathbf{G}_{\mu_{h,p},\sigma_{h,p}}, \mathbf{d} \rangle \\ &= \sum_{p'=p+1}^{P-1} G(f_p; \mu_{h,p}, \sigma_{h,p}) d_{p'}. \end{aligned} \quad (10)$$

In practice, only those peaks p' located after peak p in frequency are analysed. The remaining ones satisfy $f_{p'} < f_p$, for $p' = 0, 1, \ldots, p-1$, and are not, therefore, plausible candidates for harmonic partials of peak p. Besides that, only peaks p whose frequencies $f_p \leq 5000\,\text{Hz}$ need to be considered, since this is the limit of usable pitches in audio signals.

Proposed Modification V: Adjusting Quantization Error in Observed Deviations

In some specific cases, when the module of the theoretical deviation approaches 50 cents, there is a possibility of error in the quantization step (rounding function of Equation (4)). The observed value of a deviation that should, at first, approach -50 cents could end up being, actually, close to 50 cents. An example of such occurrence can be seen in red in Figure 4. A little adjustment in the computation of the observed deviations was proposed to fix this flaw. A threshold, chosen empirically as $d_\lambda = 45$ cents, was adopted to be compared against the absolute value of the theoretical deviation found for each partial. For each h such that $|d_h^{f_p}| > d_\lambda$, two possibilities of observed deviation are computed: the rounding operation in Equation (4) is substituted by

floor, $\lfloor\cdot\rfloor$, and ceil, $\lceil\cdot\rceil$ operations, resulting in two identical observed deviation sequences, with the exception of point h (in the example of Figure 4, $h = 11$). The salience is, then, computed for both sequences and the higher value (dashed black curve) is kept.

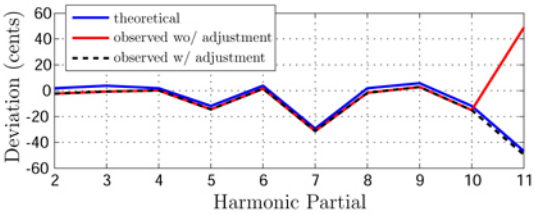


Figure 4: Computed deviations for the theoretical (in blue), observed without adjustment (in red), and observed with adjustment (dashed in black) sequences.

Proposed Modification VI: Adding a Post-Processing Stage

At last, a post-processing stage is added to remove ambiguity created by the multiples of the existent fundamental frequencies. The procedure is similar to that introduced in [3]. The idea is to analyze, for each peak p corresponding to the frequency f_p, whether the obtained salience in its submultiple frequencies f_p/q, for $q = 1, 2, \ldots, p$, is significant. For that, Gaussian functions with mean $\mu_{1/q,p} = f_p/q$ and standard deviation $\sigma_{1/q,p} = \mu_{1/q,p}\left(2^{\frac{\kappa}{1200}} - 1\right)$, with $\kappa = 20$, $G(f; \mu_{1/q,p}, \sigma_{1/q,p})$, will be employed again, but now their values at peaks p' whose frequencies $f_{p'} < f_p$, $G(f_{p'}; \mu 1/q, p, \sigma_{1/q,p})$, will be analysed. The salience in p, $S(p)$, will be replaced by

$$\bar{S}(p) = S(p) - \max_{q\in[1,\ldots,p]} \{S(p')G(f_{p'}; \mu_{1/q,p}, \sigma_{1/q,p})\}, \tag{11}$$

i.e., the maximum component of $S(p')$, with $f_{p'} < f_p$ and such that $f_{p'}$ be approximately a submultiple frequency of f_p[3], is subtracted from $S(p)$. Thus, the salience at harmonic frequencies of an existing fundamental is often forced to zero.

Figure 5 shows a salience computed before (in blue) and after (in dashed red) the post-processing stage for obtained peaks with frequency up to 3000 Hz. In this case, the salience of the peak corresponding to the fundamental frequency ($p = 1$) was not altered. However, the saliences of the peaks corresponding to its harmonic partials ($p = 2$, $p = 3$, and $p = 4$, for example) were forced to zero.

2.2.5 Inharmonic Model

No changes were proposed in the inharmonicity model adopted in [6]

$$f_h^{F_0}(\eta) = hF_0\sqrt{1 + \eta h^2}, \tag{12}$$

[3]When $\bar{S}(p) < 0$, $\bar{S}(p) = 0$ is assumed.

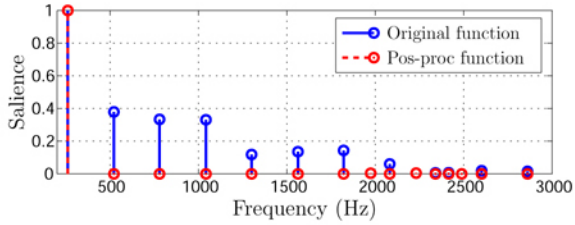


Figure 5: Computed normalized salience for obtained peaks up to 3000 Hz without (in blue) and with (in dashed red) pos-processing.

where F_0 is a given fundamental frequency and η is the inharmonicity coefficient, which is directly related to the physical properties of a string. Usually, η is not given and should be, therefore, estimated. The values of η may vary in the interval $\mathcal{N} = \{0\} \bigcup [10^{-5}, 10^{-3}]$ [11]. To estimate η, an exhaustive search is performed. A total of $N_\eta = 10$ geometrically spaced points in the interval $[10^{-5}, 10^{-3}]$ plus the value $\eta = 0$ were chosen to be analysed, resulting in $N_\eta = 10$ salience values, $S(p, \eta)$. The salience is, then, chosen to be

$$S(p, \eta^*) = \max_{\eta\in\mathcal{N}} S(p, \eta). \tag{13}$$

3 EVALUATION

The evaluation procedure, for a fair comparison, will be similar to the one presented in [6]. The idea is to construct a simple multiple-pitch estimation algorithm and evaluate the results when compared to a given ground truth [6].

At an early stage, a piano-roll representation, $\hat{\mathcal{R}}_{k,i}$, is generated from the obtained salience function, $S(k, p)$, with time-frames k MIDI Key Numbers i. If the algorithm detects a note i at frame k, $\hat{\mathcal{R}}_{k,i}$ is set to one, otherwise to zero. The procedure to compute the piano-roll representation is described in [6] and will not be repeated here. Some important parameters should, however, be highlighted; they are: size L of a sliding median filter, threshold T applied to the salience of a note i [4]. These values are set to $L = 6$ and $T = 0.2$, as in [6], throughout the experiments.

A ground truth piano-roll representation, $\mathcal{R}_{k,i}$, is computed from a text annotation (available in the database used, as will be described below) that reports onset time, offset time, and MIDI Key Number for each note played in a determined music signal. Both representations (estimated $\hat{\mathcal{R}}_{k,i}$ and reference $\mathcal{R}_{k,i}$) are, then, compared for each point (k, i). Three different types of measures were used to evaluate the results, namely Precision (P), Recall (R), and F-Measure (F),

[4]A sliding median filter with size L is applied to the representation in order to remove holes in the middle of notes and a fixed threshold is defined to determine whether a salience value is significant or not.

defined as

$$P = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}, \quad R = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \quad F = \frac{2PR}{P+R}, \tag{14}$$

where TP (true positive) is the total number of correctly identified notes, FN (false negative), of missed notes, and FP (false positive), of false notes. Since the F-measure parameter is a combination of both precision and recall parameters (more specifically, a harmonic mean), it will be the most relevant result analysed here.

The database used is called MIDI Aligned Piano Sounds (MAPS) and is described in [12]. It consists of a midi-annotated piano recordings, where the ground truth is first created as standard MIDI files and then generated in an automatic way by high quality synthesis software or a Disklavier. MAPS provides recordings with CD quality and has a total of 40 GB (65 hours) of audio recordings. In order to generalize the database, the pianos have been played in several different conditions, such as various rooms with different reverberation characteristics.

The database is divided in four subsets of signals, namely ISOL (isolated notes and monophonic excerpts), RAND (chords with random pitch notes), UCHO (usual chords from western music), and MUS (pieces of piano music). In this work, signals with different characteristics (from each of the composing sets) were chosen in order to better explore the performance of the modified method.

A set of eight experiments were proposed. The main goal is to illustrate the effects that each modification alone and the effect that all combined modifications cause in the results. The experiments are listed below:

1. DLMP method;
2. DLMP method with modification I;
3. DLMP method with modification II;
4. DLMP method with modification III;
5. DLMP method with modification IV;
6. DLMP method with modification V;
7. DLMP method with modification VI; and
8. MDLMP method.

For experiments one to seven, a music signal from the subset MUS of the database was chosen as test signal. For experiment eight, a total of 13 signals were used: one signal from subset ISOL (ascending and descending of tones); seven from RAND (random notes with MIDI number in the interval $[36, 95]$, intensity in the interval $[32, 96]$, and with seven different degrees of polyphony—from two to seven); four from UCHO (two major and two minor triads chords with intensity in the interval $[32, 96]$); and two from MUS (music signals, one of which was the chosen as test signal to perform experiments one to seven). Each of the experiments were carried out twice: the first time with the harmonic model and the second, with the inharmonic model.

The main parameters used throughout the experiments and their values are summarized below:

- number of samples $N = 4096$;
- hop size $\tau = 2048$;
- analysis window: Hanning;
- number of detected peaks per frame $P = 80$;
- number of considered harmonics $n_{\mathrm{H}} = 8$;
- sliding median window size $L = 6$; and
- salience threshold $T = 0.2$.

The results for the proposed experiments using the test signal from MUS subset can be seen in Tables 1 and 2. The first conclusion is that the results for the MDLMP were reasonably better (around 34%, 14.3%, and 23% for precision, recall, and F-measure, respectively) than the original DLMP method when using the harmonic model. When using the inharmonic model, the improvements were only moderate (around 24.5%, 5.5%, and 13.5% for precision, recall, and F-measure, respectively). Moreover, it is possible to notice that the improvement in the results caused by using inharmonicity are not as significant for the MDLMP as for the DLMP method, hence the small improvement presented in Table 2. It is also important to point out that the processing time, when using the inharmonic model, is considerably higher due to the exhaustive search performed to estimate parameter η.

When analysing the effects that each individual modification caused in the results, one can notice that, when using the harmonic model (Table 1), experiment four (DLMP with modification III) showed the best improvements (in terms of F-measure), while experiment three (DLMP with modification II) actually presented a drop when compared to the DLMP method. As a conclusion, one can say that changing the correlation measure from the inner product to the correlation coefficient has the most significant positive impact on the results, while removing note quantization can lead to poorer results. This last result, however, can be explained by the fact that the chosen test signal consists, essentially, of highly stationary and tuned notes, i.e, few variations around notes from the tempered scale. This way, the quantization step performed in Equation (5) helps to minimize small errors, hence producing a better result. Moreover, experiment six (DLMP with modification V) showed no changes at all. This is explained by the fact that no theoretical deviation computed by Equation. (5), with $h = 1, 2, \ldots, n_{\mathrm{H}}$ and $n_{\mathrm{H}} = 8$, is higher than $d_\lambda = 45$ cents, since modification II is not

applied in experiment six. Therefore, to notice some change, one should increase the number of considered harmonics n_{H} or apply modification II simultaneously, as done in experiment VIII (along with all other modifications).

Table 1: Results for test signal from the MUS subset using a harmonic model

Experiment	P	R	F
1	0.47	0.49	0.48
2	0.49	0.50	0.50
3	0.43	0.52	0.47
4	0.50	0.57	0.53
5	0.49	0.53	0.51
6	0.47	0.49	0.48
7	0.51	0.49	0.50
8	0.63	0.56	0.59

For the results using the inharmonic model, shown in Table 2, one can notice that it is not always beneficial to change from the harmonic to the inharmonic model, specially for experiments two and eight. In experiment two, the results were actually slightly worse than those obtained with the harmonic model and nearly 10% worse than the results from experiment one using the same model. This could happen because, when using the SSE method, higher harmonics are taken more often into consideration, specially for the inharmonic case. These higher partials could be favoring the estimation of false pitches, although multiple to the fundamental. This effect is minimized when combining modifications I and VI. It is also possible to see that, again, experiment four presented the best improvement when comparing the effects of each modification to the F-measure value.

Table 2: Results for test signal from the MUS subset using an inharmonic model

Experiment	P	R	F
1	0.49	0.55	0.52
2	0.46	0.49	0.47
3	0.48	0.57	0.52
4	0.50	0.62	0.55
5	0.49	0.55	0.52
6	0.50	0.55	0.52
7	0.56	0.53	0.54
8	0.61	0.58	0.59

One interesting aspect is that no modification alone is capable of reaching the best result, given by experiment eight, where all the modifications are applied at once. This means that the combination of them is what produces higher improvements, leaving room to further investigations about possible combinations of only some of the proposed modifications. This way, it may be possible to reach even better values of precision, recall, and F-measure.

Figure 6 shows the results for experiments with the DLMP (experiment one) and modified DLMP methods (experiment eight) for all 13 audio signals described above using the harmonic (top) and inharmonic (bottom) models. Dark colors represent results for experiment one and light colors for experiment eight. It is possible to see that, for the harmonic model, the results improved for almost all signals, except signals six and nine. These signals belong to the subsets RAND (sixth degree of polyphony) and UCHO (major triad), respectively. The decrease in F-measure for signal six was relatively slow, while for signal nine, it was a bit more significant, around 9%. For the inharmonic model, the results of the MDLMP model were better than the DLMP method for all signals. They also show, in general, an improvement when compared to the results obtained using the harmonic model. Further investigations should be performed in order analyze how the individual characteristics of each selected signal impacts the results for each method, DLMP and MDLMP.

At last, it should be pointed out that the proposed modifications result in no expressive changes when it comes to computational complexity. The stronger potential impacts could be: modification I, which performs spectrum equalization; and modification VI, which inserts a post-processing stage. Both of them slightly increase processing time. Moreover, this speed decrease is negligible when the inharmonic model is used, because the large computational cost of performing the exhaustive search for the most appropriate inharmonicity coefficient dominates.

4 CONCLUSIONS

In this paper, an improved method for estimating pitch relevance based on an existing timbre-independent salience [6], the DLMP method, was proposed. Several modifications were envisaged to increment the original method, and their corresponding effects in the obtained salience were analysed. The modified DLMP or MDLMP consists of the DLMP method combined with all proposed modifications. The results were presented for 13 different signals picked up from the database MAPS.

From the obtained results, one can easily notice some improvements in the performance of the MDLMP method when compared to the DLMP, specially for the harmonic model. It is also possible to notice that a modification does not necessarily implies better results, but the combination of them might. Further analysis is required to define an optimum (in terms of F-measure) combination of modifications, and also the effect that different signal characteristics has on results.

REFERENCES

[1] A. Klapuri and M. Davy, Eds., *Signal Processing Methods for Music Transcription*, Springer, New York, USA, 2006.

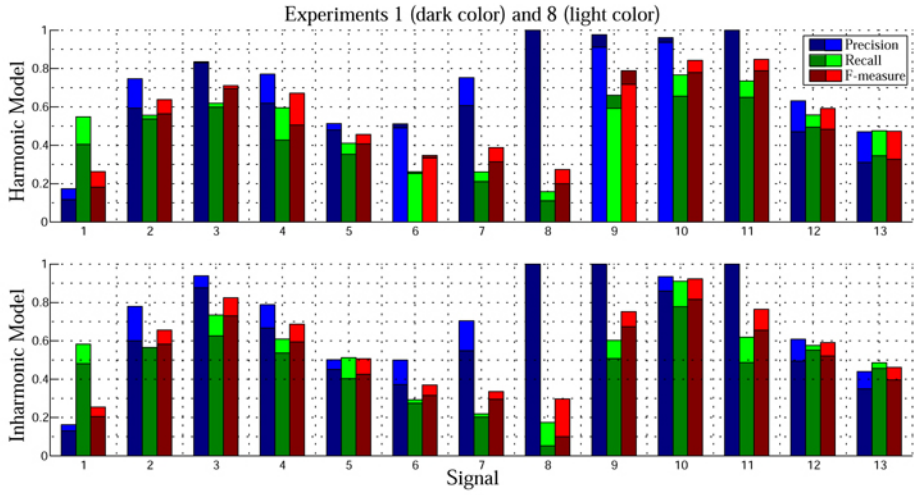


Figure 6: Results for experiments one (dark colors) and eight (light colors) for all 13 selected signals from the database MAPS. The values of precision, recall, and F-measure are shown in blue (left bin), green (middle bin), and red (right bin), respectively, for each case. The upper figure shows results for harmonic model and the lower for inharmonic model.

[2] Y. Li and D. Wang, "Separation of singing voice from music accompaniment for monaural recordings," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 4, pp. 1475–1487, May 2007.

[3] P. Cancela, E. López, and M. Rocamora, "Fan-chirp transform for music representation," in *Proceedings of the 11th International Conference on Digital Audio Effects (DAFx-10)*, Graz, Austria, September 2010, pp. 1–8.

[4] J. Salamon and E. Gómez, "Melody extraction from polyphonic music signals using pitch contour characteristics," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 6, pp. 1759–1770, August 2012.

[5] P. A. A. Esquef and L. W. P. Biscainho, "Spectral-based analysis and synthesis of audio signals," in *Advances in Audio and Speech Signal Processing: Technologies and Applications*, Hector Perez-Meana, Ed. February 2007, pp. 56–92, Hershey: IGI Global.

[6] A. Degani, R. Leonardi, P. Migliorati, and G. Peeters, "A pitch salience function derived from harmonic frequency deviations for polyphonic music analysis," in *Proceedings of the 17th Conference on Digital Audio Effects (DAFx-14)*, Erlangen, Germany, September 2014.

[7] K. Dressler and S. Streich, "Tuning frequency estimation using circular statistics," in *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, Vienna, Austria, September 2007.

[8] U. Zölzer, Ed., *Digital Audio Effects*, Wiley, Chichester, United Kingdom, 2nd edition, 2011.

[9] L. de O. Nunes, P. A. A. Esquef, and L. W. P. Biscainho, "Evaluation of threshold-based algorithms for detection of spectral peaks in audio," in *Anais do V Congresso de Engenharia de Áudio*, São Paulo, Brasil, Maio 2007, AES Brasil, pp. 66–73.

[10] Beatus Dominik Loeffler, *Instrument Timbres and Pitch Estimation in Polyphonic Music*, M.Sc. dissertation, School of Electrical and Computer Engineering, Georgia Institute of Technology, Georgia, USA, 2006.

[11] H. Fletcher, E. D. Blackham, and R. Stratton, "Quality of piano tones," *Journal of Acoustical Society of America*, vol. 34, no. 6, pp. 749–761, June 1962.

[12] V. Emiya, R. Badeau, and B. David, "Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 6, pp. 1643–1654, August 2010.

PALESTRAS / LECTURES

Sociedade de Engenharia de Áudio
Palestra em Congresso

Ministrada no 14º Congresso de Engenharia de Áudio
20ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2016, São Paulo, SP

Este artigo foi reproduzido do original final entregue pelo autor, sem edições, correções ou considerações feitas pelo comitê técnico. A AES Brasil não se responsabiliza pelo conteúdo. Outros artigos podem ser adquiridos através da Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA, www.aes.org. Informações sobre a seção Brasileira podem ser obtidas em www.aesbrasil.org. Todos os direitos são reservados. Não é permitida a reprodução total ou parcial deste artigo sem autorização expressa da AES Brasil.

Implementações de Efeitos de Áudio utilizando Arduino DUE e PedalSHIELD

Walter Antonio Gontijo, Andrew Henrique Pavei e Sidnei Noceti Filho
LINSE – Laboratório de Circuitos e Processamento de Sinais
Departamento de Engenharia Elétrica e Eletrônica, UFSC
Florianópolis, SC, 88040-900, Brasil
{walter, andrew sidnei}@linse.ufsc.br

RESUMO

Este trabalho apresenta os principais efeitos de áudio para implementações nas formas analógica e digital. Relativamente à forma analógica, são apresentados diagramas de blocos básicos, além de considerações sobre o princípio dos efeitos. Relativamente à forma digital, são apresentadas as implementações de alguns efeitos de áudio considerando o ambiente Arduino e o conjunto de hardware formado pelo Arduino DUE e PedalSHIELD. A escolha de tal conjunto se justifica pela disseminação do Arduino no desenvolvimento de diversas aplicações, bem como pelos recursos computacionais disponíveis no DUE.

0. INTRODUÇÃO

Os efeitos de áudio são amplamente usados por músicos nas suas formas analógica e digital. Na forma analógica, alguns efeitos são produzidos por circuitos analógicos contínuos e outros por circuitos analógicos amostrados [1]. Até recentemente, as implementações de efeitos de áudio na forma digital faziam uso de DSP [2]. Atualmente, devido à facilidade, tais implementações podem ser realizadas no ambiente Arduino DUE.

Em termos de grau de dificuldade de implementação, pode-se classificar da seguinte forma:

i) Circuitos analógicos contínuos são mais fáceis de implementar que circuitos analógicos amostrados.

ii) Circuitos digitais em um ambiente Arduino DUE são mais fáceis de implementar do que os circuitos que fazem uso de DSP.

A obtenção dos efeitos consiste em variar, ao longo do tempo, características como amplitude, fase e/ou frequência, conteúdo espectral ou ainda uma combinação de várias destas características. Alguns dos mais importantes tipos de efeitos são comentados a seguir.

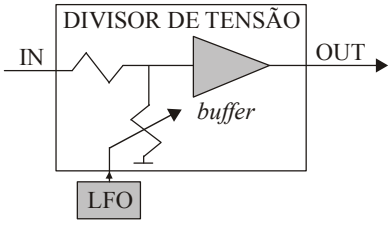
1. IMPLEMENTAÇÃO DE EFEITOS DE ÁUDIO NA FORMA ANALÓGICA

1.1 Tremolo

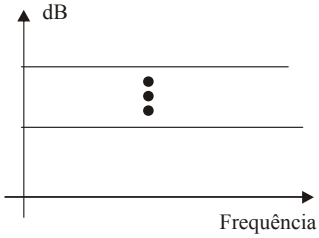
O efeito Tremolo corresponde a uma modulação da amplitude de um sinal [3]. A Figura 1(a) mostra uma forma possível de um diagrama de blocos para a obtenção desse efeito. A Figura 1(b) mostra a variação do ganho obtida com este efeito.

1.2 Vibrato

O Vibrato é um efeito que soa ao ouvido como uma modulação nas frequências do sinal. A Figura 2(a) mostra o diagrama de blocos. Dois filtros passa-tudo são suficientes para a obtenção do efeito. A Figura 2(b) mostra a variação da fase obtida com a cascata de dois filtros passa-tudo de primeira ordem.

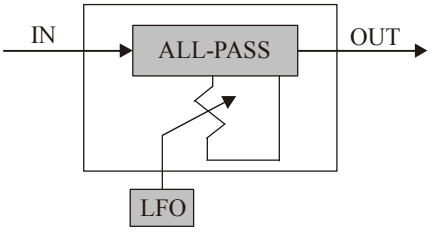


(a)

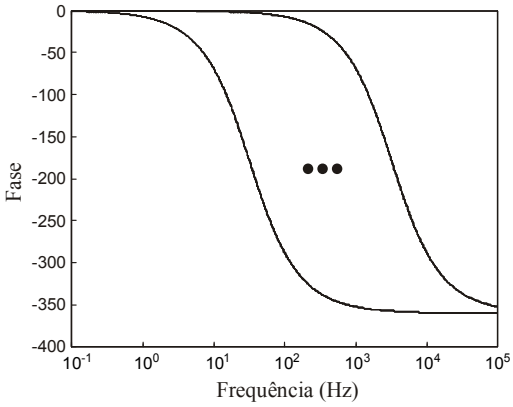


(b)

Figura 1 - Efeito Tremolo. (a) Forma possível de um diagrama de blocos. (b) Ganho das etapas.



(a)

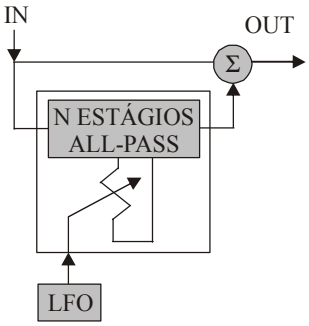


(b)

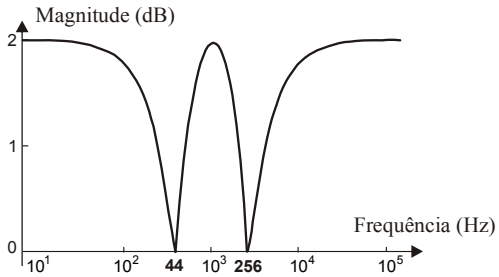
Figura 2 - Efeito Vibrato. (a) Diagrama de blocos da implementação do efeito. (b) Fase obtida ao se utilizar dois filtros passa-tudo.

1.3 *Phaser*

O efeito *Phaser* também soa ao ouvido como uma modulação nas frequências do sinal. Nesse caso, efeitos significativos são obtidos com quatro a doze estágios passa-tudo. A Figura 3(a) mostra o diagrama de blocos, que é semelhante ao diagrama de blocos do Vibrato, mas apresenta uma soma do sinal atrasado com o sinal puro. Para 4 filtros passa-tudo, são produzidos dois *notchs* (ver Figura 3(b)), que não estão igualmente espaçados em uma escala linear de frequência.



(a)

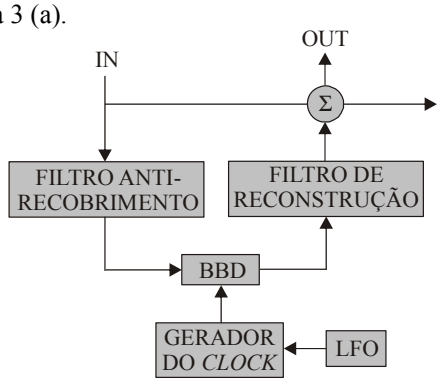


(b)

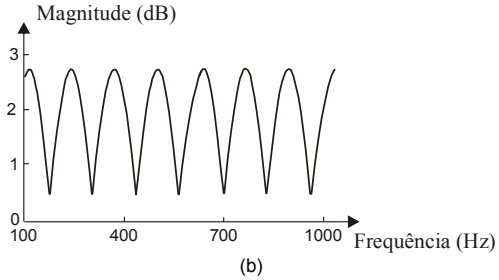
Figura 3 - Efeito Phaser. (a) Diagrama de blocos da implementação do efeito. (b) Resposta em frequência no caso de dois estágios passa-tudo.

1.4 *Chorus*

O efeito *Chorus* também soa ao ouvido como uma modulação nas frequências do sinal. É obtido com a soma de um sinal que foi semitonado com o sinal puro, conforme mostra a Figura 4. O sinal semitonado é obtido com um circuito analógico amostrado (um BBD – *bucket-brigade device*). Um sinal de *clock* com frequência variável controla os tempos de atraso entre a entrada e a saída, que é alterada com um LFO. O diagrama do *chorus* da Figura 4 é semelhante ao diagrama do *phaser* da Figura 3 (a).



(a)



(b)

Figura 4 - Efeito *Chorus*. (a) Diagrama de blocos da implementação do efeito. (b) Resposta em frequência.

1.5 *Flanger*

O efeito *Flanger* também é obtido com a soma de um sinal que foi semitonado por um BBD com o sinal puro (Figura 5), porém, esse efeito difere do efeito *chorus* porque o sinal semitonado é realimentado ao BBD. Outra diferença significativa em relação ao *chorus* é que no *flanger* os tempos de atraso entre a entrada e a saída do BBD são menores. Como no *chorus* e no *phaser*, na saída do circuito, também ocorrem interferências construtivas e destrutivas.

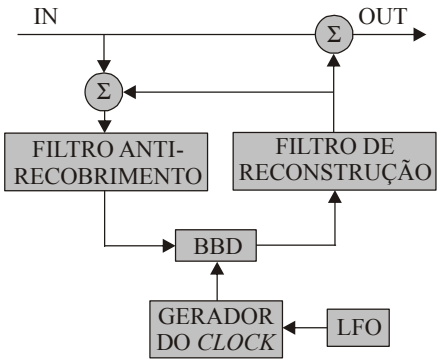


Figura 5 - Diagrama de blocos de um *Flanger*.

1.6 Eco e *Reverber*

Os efeitos Eco e *Reverber* são usados tanto para sinais de guitarra como para sinais de voz. Eles consistem em simular repetições do som. No efeito eco, ocorre quando existe um refletor a uma distância de pelo menos 8,5 m, o que corresponde a um atraso de cerca de 50 ms entre o som original e o refletido. Mas esse é aproximadamente o tempo mínimo para que o som original e o refletido soem distintamente. Efeitos mais significativos e interessantes são obtidos para tempos de atraso maiores.

O diagrama de blocos de um efeito de eco (Figura 6) é semelhante ao do *flanger*. Mas, além da diferença significativa entre os tempos de atraso envolvidos, aqui a frequência do *clock* é constante.

No efeito *Reverber*, não apenas uma, mas inúmeras reflexões ocorrem em diferentes instantes de tempo, de tal forma que não é percebido um tempo de atraso entre o som original e os vários sons refletidos. O som persiste, diminuindo de amplitude até se extinguir. Mas devido ao grande número de reflexões e inúmeros tempos de atraso que precisam ser simulados, não é prático obter o efeito *reverber* com circuitos eletrônicos analógicos. Efeitos interessantes são obtidos com processadores digitais ou com transdutores eletromecânicos e molas, por exemplo.

1.7 Wah-Wah

O efeito Wah-Wah tenta imitar o som feito pela boca quando seu nome é pronunciado. Ele é obtido quando o conteúdo espectral é variado, conforme mostrado na Figura 7, com o uso de um filtro seletor passa-faixa que apresenta um zero na origem e um zero no infinito com frequência central variando sob o comando de um pedal. Outros tipos de filtros podem ser usados tais como um passa-baixa com fator de qualidade $Q >> 0{,}707$ (para apresentar sobrepassamento em frequência) ou um equalizador *bump* que apresenta uma constante na origem e no infinito.

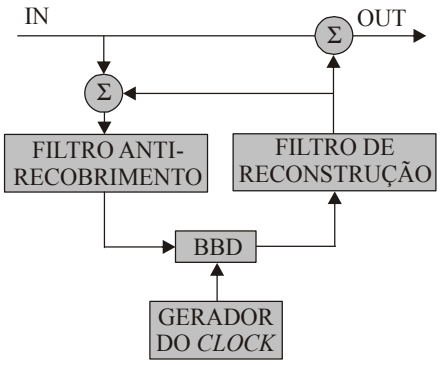


Figura 6 - Diagrama de blocos para obtenção do efeito de Eco.

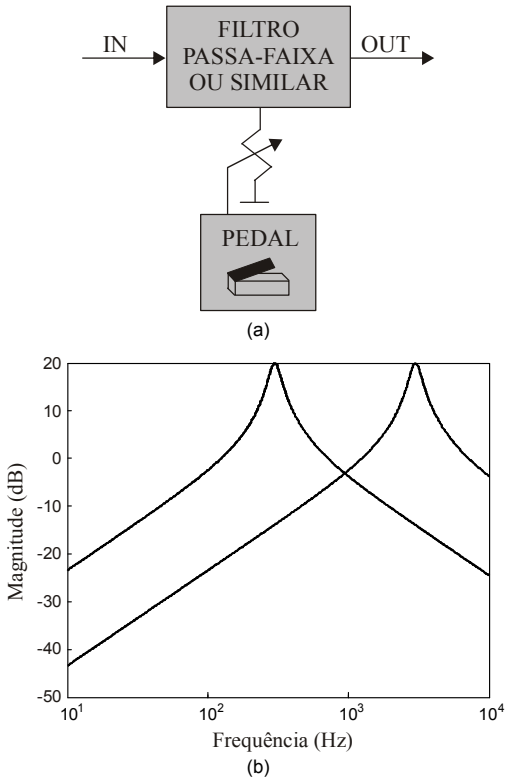


Figura 7 - (a) Diagrama de blocos do efeito Wah-Wah.(b) Curvas de resposta de um filtro passa-baixa controladas pelo acionamento do pedal.

1.8 Distorção

Os efeitos de Distorção são obtidos quando se obtém uma saturação forçada de um ou mais estágios do circuito que processa o sinal, objetivando a criação de harmônicos. Enquanto em todos os outros tipos de efeitos as saturações não são desejadas, aqui elas são propositadamente produzidas. Na literatura, esse é provavelmente o efeito que apresenta o maior número de variações em termos de circuitos. A quantidade e a proporção de harmônicos pares e ímpares dependem da topologia e, principalmente, se as saturações são obtidas com amp. op., transistores bipolares, JFET, MOSFET, diodos e válvulas.

Três dos mais conhecidos tipos de pedais que proporcionam esse efeito são o *overdrive*, o *distortion* e o *fuzz*. Com o *overdrive,* tenta-se simular as distorções que ocorrem em amplificadores valvulados saturados. Um forte componente desses sinais são harmônicos de ordem par, que soam mais suaves ao ouvido humano. No efeito *distortion*, harmônicos pares e ímpares são obtidos. Logo, em relação ao *overdrive*, tal efeito soa menos suave ao ouvido humano. No efeito *fuzz*, procura-se realçar os harmônicos ímpares, que aparecem naturalmente quando os sinais são aproximados a ondas quadradas. Por isso, dentre os três efeitos de distorção, este soa o menos suave.

1.9 Oitavador

O pedal Oitavador [4] proporciona uma sonoridade que dá a impressão ao ouvinte de dois solistas tocando simultaneamente de forma sincronizada. A Figura 8 mostra um diagrama de blocos possível para implementação de um oitavador. O primeiro bloco é um conversor que recebe o sinal da guitarra e, na saída, apresenta uma onda quadrada cuja frequência fundamental é a metade da fundamental da frequência da corda percutida. Aqui já temos um oitavador, mas o sinal de onda quadrada apresenta harmônicos de ordem ímpar, caracterizando um *octafuzz*. Assim, se é desejado minimizar a distorção, o sinal pode ser aplicado em um filtro cuja frequência de corte varia com a tensão de controle fornecida pelo bloco conversor frequência-tensão.

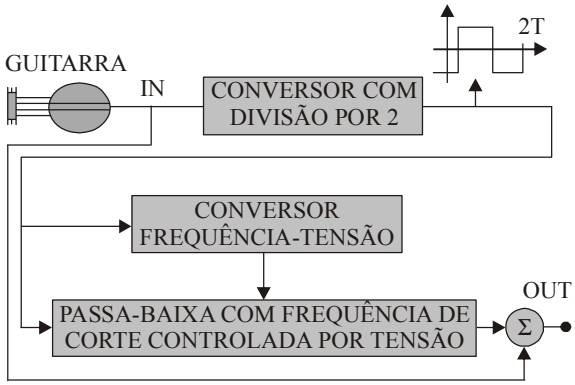


Figura 8 – Bloco conversor do sinal da guitarra em um sinal de onda quadrada.

A conversão do sinal de uma guitarra em uma onda quadrada não é uma tarefa trivial, pois dentro de um período ocorrem vários cruzamentos por zero (devido aos harmônicos) e o número de cruzamentos é dependente da corda percutida. Os circuitos que realizam essa função de conversão são relativamente complexos. O sinal filtrado e o sinal da guitarra são misturados no bloco somador.

1.10 Leslie

O Efeito Leslie [5] é obtido quando um ou mais alto-falantes são continuamente girados em um gabinete. O movimento dos alto-falantes produz um som que apresenta ao mesmo tempo uma variação aparente da frequência e variação real da amplitude dos sinais. A rotação do(s) alto-falante(s) gera dois efeitos no som percebido pelo ouvinte: alteração periódica na intensidade, devida à alteração da distância entre a fonte e o ouvinte; e alteração na frequência, devida ao efeito Doppler, pelo fato de a fonte afastar-se e aproximar-se do ouvinte.

A Figura 9(a) mostra o diagrama de blocos básico do circuito Leslie para simular o movimento giratório dos alto-falantes. A Figura 9(b) mostra uma forma possível de obter a defasagem necessária entre a variação aparente da frequência e a variação real da amplitude. A simples cascata de um vibrato e um tremolo não simula um efeito Leslie se não for respeitada a diferença de 90 graus entre os dois efeitos.

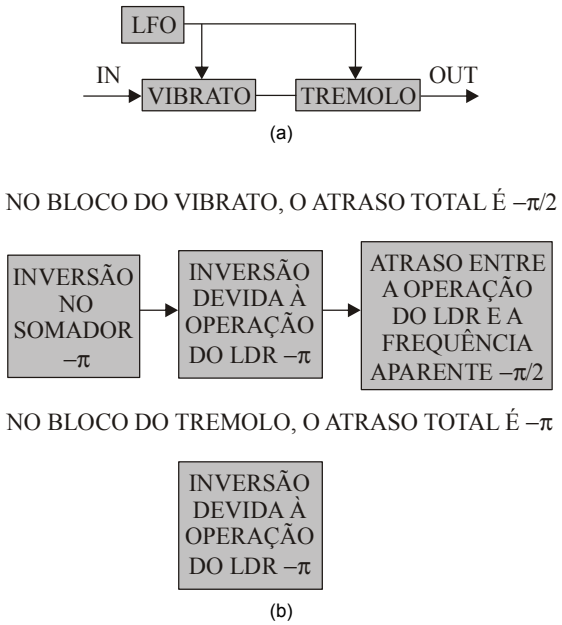


Figura 9 - (a)Diagrama de blocos. (b) Avaliação da defasagem entre variação de frequência e de amplitude.

No circuito do vibrato, o sinal sofre um atraso total de $\pi/2$ rad e no circuito do tremolo, o sinal sofre um atraso total de $-\pi$ rad [5]. Assim, a frequência instantânea estando adiantada de $\pi/2$ rad em relação à variação de amplitude cumpre o objetivo do ponto de vista das defasagens necessárias para a simulação do efeito Leslie.

1.11 Compressão (e *Sustain* para Guitarras)

O efeito de Compressão [6] é largamente usado em gravação de sinais de vários instrumentos musicais, no processamento da voz de cantores, em retransmissores de rádio, etc., com o objetivo de salientar sinais de nível baixo e diminuir o nível de sinais altos, ou seja, diminuir a faixa dinâmica. Adicionalmente, nos pedais compressores disponíveis para guitarristas, permite obter um efeito adicional de *Sustain*. A Figura 10(a) mostra um diagrama de blocos possível para a obtenção desses

efeitos. A Figura 10(b) mostra as relações de ganho possíveis de serem obtidas. Com compressão infinita é possível obter o *sustain*: depois da corda da guitarra ser percutida, e enquanto a amplitude do sinal da guitarra for maior do que o nível de *threshold*, mostrado na Figura 10(b), o sinal de saída permanece constante.

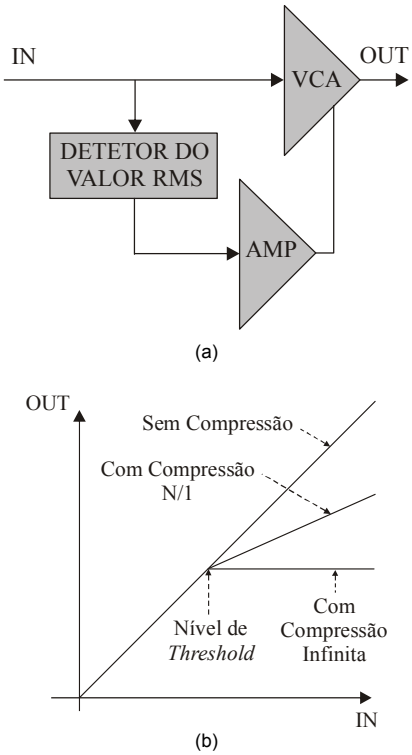


Figura 10 - (a) Diagrama de blocos possível de um compressor. (b) Relação entre a amplitude de saída e de entrada.

2 IMPLEMENTAÇÃO DE EFEITOS DE ÁUDIO NA FORMA DIGITAL

2.1 Arduino

Arduino é uma plataforma "*open-source*" de hardware e software utilizada para construir projetos eletrônicos [7]. Normalmente, o hardware utiliza microcontroladores ATMEL e são disponibilizadas diferentes placas de desenvolvimento, por exemplo, Uno, Mini, Mega, Due, etc [8]. Para o desenvolvimento de software, pode-se usar a IDE do Arduino [9] ou um compilador C, C++, por exemplo, Atmel Studio [10]. Além das placas principais, há também um extenso conjunto de placas filhas (denominadas *Shields*) que permite a utilização da plataforma Arduino em diferentes áreas de aplicação. Para a implementação dos efeitos de áudio considerados nesse artigo, são utilizados o Arduino Due [11], a *shield* PedalShield [12] e a IDE do Arduino [9].

2.2 Arduino Due

A placa Arduino Due é a primeira a utilizar um microcontrolador ARM de 32 bits e suas principais características são apresentadas na Tabela 1 [11]:

Tabela 1 - Principais Características do Arduino DUE

Microcontrolador	ATMEL AT91SAM3X8E (ARM Cortex M3);
Frequência de *clock*	84 MHz
Pinos de E/S digital	54 (12 desses podem ser configurados como saídas PWM)
Pinos de entrada analógica	12
Resolução do AD	12 bits
Frequência de amostragem	8kHZ, 16kHz, 32kHz, 44,1kHz, etc.
Pinos de entrada analógica	2
UARTs	4
Micro-usb	2
SPI	1
JTAG	1
Tensão de alimentação e E/S	3,3V (confirmar)
Memória Flash	512 KB disponíveis para o usuário
SRAM	96KB (dois bancos: 64KB e 32KB)

Devido a maior capacidade de processamento e recursos de memória do ARM Cortex M3, o Arduino DUE suporta a implementação de algoritmos de processamento digital de sinais, como filtros digitais e efeitos de áudio.

Os principais sinais usados na interligação entre o Arduino DUE e a PedalShield são mostrados na Figura 11.

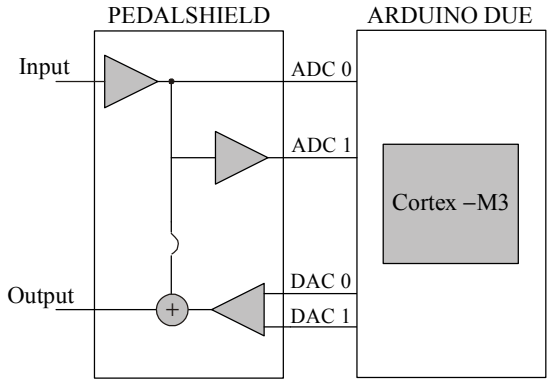


Figura 11 - Conexão entre a PedalShield e o Arduino Due [12].

Observa-se na Figura 11 que a PedalShield possui os estágios de entrada e saída para a conexão com o Arduino DUE. O estágio de entrada amplifica o sinal analógico recebido na entrada e gera saídas balanceadas que são enviadas para os conversores AD (ADC0 e ADC1) do Arduino Due. Já o estágio de saída permite que o Arduino Due externe as saídas analógicas balanceadas dos conversores DA (DAC0 e DAC1). Observa-se na Figura 11 que essas saídas podem ser somadas à entrada original. Na sequência, são apresentados os efeitos de áudio considerados nesse artigo.

3 EFEITOS DE ÁUDIO NA FORMA DIGITAL

3.1 Introdução

O *delay* (atraso) é o bloco básico usado na obtenção de diversos efeitos como, por exemplo, vibrato, *flanger*, *chorus*, *reverber*, eco, etc. O *delay* atrasa um sinal digital de um número de amostras desejadas. Dependendo de qual sinal é atrasado,

obtém-se um *Comb filter* FIR (sinal de entrada) ou *Comb filter* IIR (sinal de saída). A equação de um *Comb filter* FIR [13] é dada por:

$$y(n) = x(n) + ax(n-D) \tag{1}$$

onde $y(n)$ é a amostra de saída, $x(n)$ é a amostra de entrada, a é o fator de ponderação, D é o valor do atraso em amostras e $x(n-D)$ é a amostra de entrada atrasada de D amostras.

O diagrama de blocos do *Comb filter* FIR é mostrado na Figura 12 e sua correspondente resposta em frequência é apresentada na Figura 13.

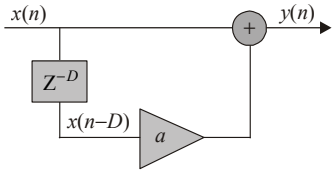


Figura 12 - Diagrama de blocos do *Comb filter* FIR.

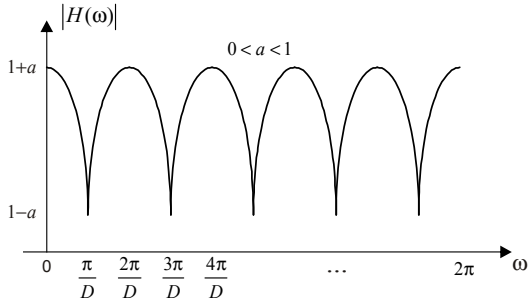


Figura 13 - Resposta em frequência do *Comb filter* FIR [13].

A equação recursiva de um *Comb filter* IIR é dada por [13]:

$$y(n) = x(n) + ay(n-D) \tag{2}$$

onde $y(n-D)$ é a amostra de saída atrasada de D amostras.

O diagrama de blocos do *Comb filter* IIR é apresentado na Figura 14 e sua correspondente resposta em frequência na Figura 15.

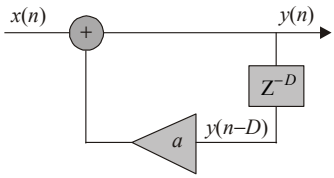


Figura 14 - Diagrama de blocos do *Comb filter* IIR.

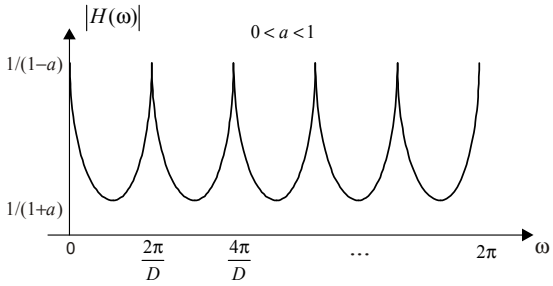


Figura 15 - Resposta em frequência do *Comb filter* IIR [13].

3.2 Vibrato

O valor escolhido para o atraso D em amostras (de forma equivalente, em segundos) altera o som percebido bem como o tipo de efeito de áudio. Por exemplo, o efeito Vibrato é implementado na forma digital variando o atraso D continuamente no tempo [13]-[15]. O diagrama de blocos que representa tal implementação é mostrado na Figura 16.

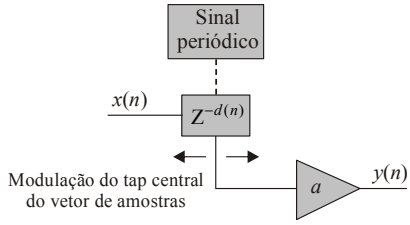


Figura 16 - Diagrama de blocos do Vibrato.

A equação correspondente do vibrato é dada por [13]:

$$y(n) = ax(n-d(n)) \tag{3}$$

onde $y(n)$ é a amostra de saída, $d(n)$ é o atraso variante no tempo, a é o fator de ponderação e $x(n-d(n))$ é a amostra de entrada atrasada de $d(n)$ amostras.

O atraso variante no tempo ($d(n)$) é obtido a partir de um oscilador de baixa frequência (LFO), que gera um sinal periódico, dado por [13]:

$$d(n) = \frac{D}{2}[1-\cos(2\pi.n.f_{cy})] \tag{4}$$

onde D é o valor máximo do atraso em amostras e f_{cy} é a frequência do LFO.

Na prática, o valor máximo utilizado para o atraso está entre 0 ms e 3 ms. Considerando, por exemplo, que a frequência de amostragem é 16kHz, obtém-se 48 amostras para o máximo atraso D. Efeitos interessantes são obtidos quando a frequência do LFO é menor do que 5Hz.

Para a implementação do algoritmo do vibrato, deve-se alocar um vetor de tamanho D para armazenar as amostras de entrada bem como calcular o atraso $d(n)$. Normalmente, existem duas opções para a obtenção desse atraso. A primeira considera uma tabela com os valores anteriormente armazenados. Na segunda opção, é utilizada uma equação recursiva para calcular o sinal periódico. Normalmente, devido à baixa frequência do LFO (menor do que 5Hz), há uma relação de compromisso entre complexidade computacional e utilização de recursos de memória, Na primeira, tem-se uma menor complexidade computacional com maior recurso de memória, enquanto na segunda opção tem-se uma maior complexidade computacional com menor recurso de memória [16].

Outra consideração relevante é sobre o valor obtido para o atraso $d(n)$, o qual pode ser fracionário. Pode-se perceber que tal situação resulta em um problema, dado que $d(n)$ corresponde a um índice do vetor de amostras. Normalmente, a solução usada em tal situação é considerar a interpolação linear entre duas amostras nos instantes de tempo $M+1$ e M, conforme apresentado por [17]:

$$y(n) = x(n-[M+1])\,frac + x(n-M)(1-frac) \quad (5)$$

onde $frac$ corresponde ao valor fracionário do atraso.

3.3 Flanger/Chorus

Os efeitos *Flanger*/*Chorus* podem ser implementados de forma equivalente ao *Comb filter* FIR, ou seja, adicionando o sinal de entrada com uma réplica atrasada no tempo, conforme diagrama de blocos apresentado na Figura 17. Equivalente ao vibrato, a réplica é obtida de um atraso variante no tempo ($d(n)$), que é controlado por um LFO. Normalmente, no *flanger*, o atraso é menor do que 10ms e a frequência do LFO é menor ou igual a 2Hz. Já para o *chorus*, o atraso deve ficar entre 10ms e 30ms e a frequência do LFO menor ou igual a 1Hz [17-18].

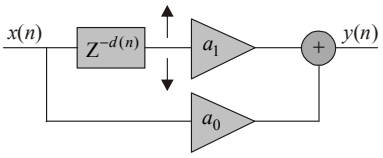


Figura 17 - Diagrama de blocos *Flanger*/*Chorus*.

A equação geral para os efeitos *flanger*/*chorus* é dada por [13]:

$$y(n) = a_0 x(n) + a_1 x(n-d(n)) \quad (6)$$

onde $y(n)$ é a amostra de saída, $d(n)$ é o atraso variante no tempo, a_0 e a_1 são os fatores de ponderação e $x(n-d(n))$ é a amostra de entrada atrasada de $d(n)$ amostras.

A correspondente resposta em frequência para $a_0 = 1$ e $0 < a_1 < 1$ é mostrada na Figura 18. Pode-se observar em tal figura que a variação de $d(n)$ causa a atenuação de diferentes componentes de frequência.

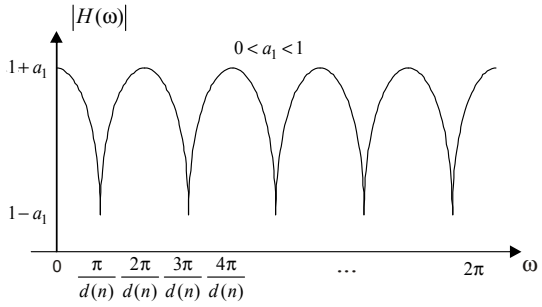


Figura 18 - Resposta em frequência do *Flanger*/*Chorus*.

Outra estratégia de implementação do *flanger*/*chorus* foi proposta por Dattorro [18]. A Figura 19 mostra o diagrama de blocos de tal estratégia.

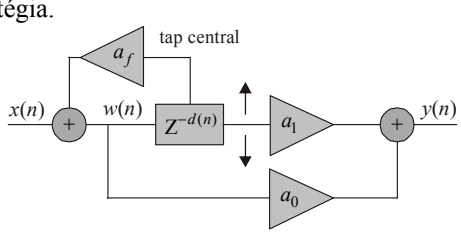


Figura 19 - Diagrama de blocos do *Flanger*/*Chorus* proposto em [18].

As equações que correspondem à implementação da Figura 19 são dadas por:

$$w(n) = x(n) - a_f w(n - \frac{D}{2}) \quad (7)$$

$$y(n) = a_0 w(n) + a_1 w(n-d(n)). \quad (8)$$

Deve-se observar que o diagrama apresentado na Figura 19 permite representar os efeitos vibrato, *flanger* e *chorus*. Para cada efeito, deve-se escolher os valores adequados para os coeficientes, tamanho do atraso, frequência e tipo do sinal usado no LFO, conforme apresentado na Tabela 2 [18].

Tabela 2 - Representação dos Efeitos Vibrato, *Flanger* e *Chorus* Usando a Estrutura Proposta por Dattorro

Efeito	Coeficientes	Atraso D (ms)	LFO (Hz)	Tipo de sinal usado no LFO
Vibrato	$a_f = 0$ $a_0 = 0$ $a_1 = 1$	0 – 3	0.1 até 5Hz	Senoidal
Flanger	$a_f = 0,7$ $a_0 = 0,7$ $a_1 = 0,7$	D < 10	<2Hz	Senoidal
Chorus	$a_f = -0,7$ $a_0 = 0,7$ $a_1 = 1$	10<D<30	< 1Hz	Senoidal Ruído filtrado Banda estreita

3.4 Tremolo

O Tremolo é implementado digitalmente multiplicando o sinal de entrada pela saída do LFO. Tal operação equivale a modular em amplitude a entrada, conforme diagrama de blocos apresentado na Figura 20.

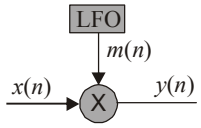


Figura 20 - Diagrama de blocos do Tremolo.

Na prática, a frequência de saída do LFO é ajustada entre 1Hz e 10 Hz. A equação para o cálculo de tal frequência é dada por [16]:

$$m(n) = \sin(2\pi.n.f_{cy}) \quad (9)$$

onde $m(n)$ é o sinal de modulação e f_{cy} é a frequência do sinal de modulação.

O sinal na saída do tremolo é dado por:

$$y(n) = m(n)x(n) \quad (10)$$

onde $x(n)$ é o sinal de entrada e $y(n)$ é o sinal de saída.

3.5 Eco

O Eco pode ser implementado digitalmente a partir do *Comb filter* FIR com a especificação do valor desejado do atraso D. Por exemplo, considerando que a frequência de amostragem seja 8kHz e que se

deseja um atraso máximo de 0,1 segundos, deve-se adotar o valor 800 para o atraso D e alocar um tamanho correspondente para o vetor de atrasos $x(n-D)$.

A equação para a implementação do eco é dada por [13]:

$$y(n) = x(n) + ax(n-D) \tag{11}$$

onde $y(n)$ é a amostra de saída, $x(n)$ é a amostra de entrada, a é o fator de ponderação, D é o valor do atraso em amostras e $x(n-D)$ é a amostra de entrada atrasada de D amostras.

O diagrama de blocos para a implementação do eco é mostrado na Figura 21.

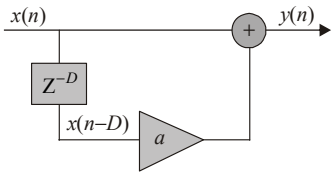


Figura 21 - Diagrama de blocos do Eco.

3.6 Reverberação

O efeito de Reverberação ou *Reverb* simula a reflexão do som em um ambiente fechado, produzindo repetições do som original. O som produzido na reverberação é classificado em três componentes: som direto, reflexões iniciais (*early*) e reflexões finais (*late*). As reflexões iniciais ocorrem de 10ms a 100ms após o som direto. Já as reflexões finais são maiores do que 100ms [13]. A Figura 22 mostra os componentes da reverberação.

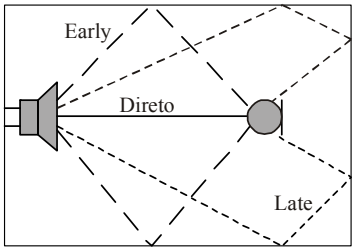


Figura 22 - Componentes da Reverberação.

A implementação digital da reverberação deve, necessariamente, considerar os três componentes citados, de forma a produzir um efeito mais próximo do "real". Os algoritmos mais utilizados para a implementação da reverberação foram propostos por Schoreder [19] e Moorer [20], consistindo em uma sequência de *Comb filter* IIR (C1 a C4) e filtros passa-tudo (A1 e A2), conforme apresentado no diagrama de blocos da Figura 23.

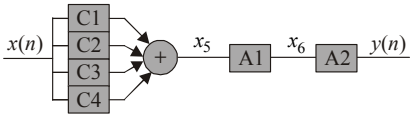


Figura 23 - Diagrama de blocos da Reverberação.

As equações dos *Comb filter* IIR são dadas por:

$$y_{C1}(n) = x(n) + a_1 y(n-D_1) \tag{12}$$

$$y_{C2}(n) = x(n) + a_2 y(n-D_2) \tag{13}$$

$$y_{C3}(n) = x(n) + a_3 y(n-D_3) \tag{14}$$

$$y_{C4}(n) = x(n) + a_4 y(n-D_4). \tag{15}$$

Já as equações correspondentes aos filtros passa-tudo são dadas por:

$$y_{A1}(n) = a_5 y(n-D_5) - a_5 x_5(n) + x_5(n-D_5) \tag{16}$$

$$y_{A2}(n) = a_6 y(n-D_6) - a_6 x_6(n) + x_6(n-D_6). \tag{17}$$

3.7 Distorção

Considerando a classificação apresentada na Seção 1.8, o efeito de Distorção considerado é do tipo *fuzz*, pois gera apenas harmônicos ímpares. Para a implementação desse efeito, é considerada a função não linear "*Symmetrical soft clipping*" dada por [17]:

$$f(x) = \begin{cases} 2x, & 0 \le |x| \le \frac{1}{3} \\ \frac{3-(2-3x)^2}{3}, & \frac{1}{3} \le |x| \le \frac{2}{3} \\ 1, & \frac{2}{3} \le |x| \le 1 \end{cases} \tag{18}$$

4 RESULTADOS EXPERIMENTAIS

Os algoritmos aqui considerados foram implementados no Matlab e no ambiente Arduino. Nos testes realizados, cada algoritmo foi submetido a arquivos de testes, gerando os correspondentes arquivos de saída. Os arquivos de testes e os algoritmos implementados encontram-se disponíveis para download no endereço apresentado em [21]. Deve-se observar que os arquivos foram gerados para uma frequência de amostragem de 16 kHz e 12 bits de codificação e que podem ser manuseados pela ferramenta de edição de áudio Ocenaudio [22].

5 CONSIDERAÇÕES FINAIS

Neste artigo, foram discutidas formas de obtenção de efeitos de áudio analógicos e digitais. A partir dos diagramas de blocos sugeridos, o projetista pode buscar soluções variadas, considerando-se a gama de circuitos propostos para implementar os diagramas apresentados. Alguns exemplos práticos foram mostrados, com o objetivo de auxiliar os *hobistas* na implementação dos circuitos desejados. Relativamente à forma digital, foram apresentadas implementações de alguns efeitos de áudio considerando o ambiente Arduino e o conjunto formado pelo Arduino DUE e pelo PedalSHIELD.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] S. Noceti Filho, A. L. Dalcastagnê, W. A. Gontijo, "Filtros e efeitos de áudio analógicos e digitais," in *Anais do 9º Congresso de Engenharia de Áudio da AES Brasil (AES Brasil 2011)*, São Paulo, SP, Maio 2011, pp. 12-37,

[2] D. Hunter, *Guitar Effects Pedals the Practical Handbook*, Backbeat Book, 2004.

[3] Gabriel C. K. da Silva, E. Costa Neto e S. Noceti Filho, "Otimização do uso de foto-acopladores artesanais para o efeito tremolo," in *Anais do 11º Congresso de Engenharia de Áudio da AES Brasil (AES Brasil 2013)*, São Paulo, SP, Maio 2013, pp. 89-93.

[4] B. G. Fernandes e S. Noceti Filho, "Implementação de um oitavador com baixa distorção," in *Anais do 13º Congresso de Engenharia de Áudio da AES Brasil (AES Brasil 2015)*, São Paulo, SP, Maio 2015, pp. 47-51.

[5] C. A. C. Wengerkievicz e S. Noceti Filho "Circuito simples para simulação do efeito Leslie" in *Anais do 11º Congresso de Engenharia de Áudio da AES Brasil (AES Brasil 2013)*, São Paulo, SP, Maio 2013, pp. 16-23.

[6] THATCorporation, "The Mathematics of Log-Based Dynamic Processors" *Application Note* 101A, pp. 01-04.

[7] https://www.arduino.cc/en/Guide/Introduction.

[8] https://www.arduino.cc/en/Main/Boards.

[9] https://www.arduino.cc/en/Main/Software.

[10] http://www.atmel.com/pt/br/tools/ATMELSTUDIO.aspx.

[11] https://www.arduino.cc/en/Main/ArduinoBoardDue.

[12] http://www.electrosmash.com/pedalshield

[13] S. J. Orfanidis, *Introduction to Signal Processing*, Prentice Hall, 2009.

[14] P. J. Bloom, "High-quality digital audio in the entertainment industry: an overview of achievements and challenges, *IEEE ASSP. Mag*, vol. 2, no. 4, pp. 2-25, Oct. 1985.

[15] A. V. Oppenheim, *Applications of Digital Signal Processing,* Prentice Hall, 1978.

[16] J. Tomarakos, D. Ledger, "Using the Low-cost, High Performance ADSP-21065L Digital Signal Processor for Digital Audio Applications," *Analog Devices Applications*, April 1998.

[17] U. Zolzer, *DAFX: Digital Audio Effects*, John Wiley & Sons, Ltd – 2002.

[18] J. Dattorro, "Effect design, part 2: delay-line modulation and chorus," *J. Audio Eng. Soc*., vol. 45, no. 10, pp. 764-788, Oct. 1997.

[19] M. R. Schroeder, "Natural sounding artificial reverberation," *J. Audio Eng. Soc.*, vol. 10; no. 3, pp. 219-223, Jul. 1962.

[20] J. A. Moorer, *About This Reverberation Business*, Centre Georges Pompidou, 1979.

[21] http://www.linse.ufsc.br/research-public.

[22] http://www.ocenaudio.com.br.

Sociedade de Engenharia de Áudio

Palestra em Congresso

Ministrada no 14º Congresso de Engenharia de Áudio
20ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2016, São Paulo, SP

Este artigo foi reproduzido do original final entregue pelo autor, sem edições, correções ou considerações feitas pelo comitê técnico. A AES Brasil não se responsabiliza pelo conteúdo. Outros artigos podem ser adquiridos através da Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA, www.aes.org. Informações sobre a seção Brasileira podem ser obtidas em www.aesbrasil.org. Todos os direitos são reservados. Não é permitida a reprodução total ou parcial deste artigo sem autorização expressa da AES Brasil.

Um relato atual da prática da bioacústica de mamíferos terrestres no Brasil

Patrícia Monticelli[1], Bruna Campos Paula[1], Flora Balieiro[1], Regis Rossi A. Faria[2]
[1]Laboratório de Etologia e Bioacústica, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo
[2]Laboratório de Acústica e Tecnologia Musical, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo
Ribeirão Preto, SP, 14040-901, Brasil
patrimonticel@gmail.com, brunacampospaula@gmail.com, flora.balieiro@gmail.com, regis@usp.br

RESUMO

Apresentamos um relato de uso de equipamentos para registro de vocalizações de animais em campo e ferramentas utilizadas para a análise destes sinais. Nosso objetivo é contextualizar a prática da pesquisa em bioacústica de mamíferos terrestres atualmente em curso no Brasil com os recursos tecnológicos disponíveis, discutindo suas virtudes e limitações, e contrapondo características existentes e recursos desejáveis para os sistemas de captura, gravação e análise sonora que ainda não atendem plenamente às necessidades da área.

0. INTRODUÇÃO

A prática da pesquisa em bioacústica fundamenta-se no estudo do comportamento dos animais por meio de suas vocalizações e relações sonoras com o meio ambiente. O acesso ao registro sonoro é viabilizado pelo uso de equipamentos eletrônicos como sistemas de captura e gravação, e além disso a análise dos sinais também requer o uso extensivo de ferramentas de processamento de áudio.

Neste relato pretendemos apresentar os diferenciais na prática da bioacústica realizada no Brasil, em suas motivações, interesses, processos adotados e animais abordados. No Laboratório de Etologia e Bioacústica (EBAC) do Departamento de Psicologia da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto (FFCLRP-USP) estudam-se os mamíferos neotropicais sob uma abordagem etológica, particularmente focalizando o comportamento de comunicação acústica de alguns ícones da fauna sul-americana, como o lobo-guará, a capivara e o quati. Além do levantamento das vocalizações usadas para a comunicação intra e interespecíficas, no EBAC analisamos a associação entre tamanhos de repertórios e o nível de sociabilidade das espécies, a influência do ruído antrópico na comunicação à longa distância (por exemplo, com o lobo-guará), a seleção sexual através de sinais acústicos, e a presença e significado de efeitos não-lineares nas vocalizações de mamíferos. Além disso, descrevemos paisagens acústicas para fins de conservação e avaliação do impacto de atividades humanas sobre a fauna, o que vem despertando um

interesse adicional em conhecer e mapear a ecologia sonora-musical destes habitats.

Em colaboração com o Laboratório de Acústica e Tecnologia Musical (LATM) do Departamento de Música da mesma faculdade, temos atuado no sentido de identificar requisitos e selecionar métodos de aquisição e processamento de sinais sonoros para obter uma melhor qualidade de captura sonora do ambiente e isolamento das interferências de múltiplas fontes, em direção a um sistema integrado para monitoramento ativo com cobertura espacial capaz, por exemplo, de localizar animais e rastrear seus movimentos em uma escala espacial maior, conforme Blumstein et al (2011) sustentam ser viável com o uso das novas tecnologias de áudio.

Considerando também a migração dos aplicativos para as plataformas móveis e a oferta de acessórios para este segmento, atualmente é praticável a captação de sons em boa resolução em dispositivos portáteis como *tablets* e *smartphones*. Os sistemas que utilizam áudio *wireless* expandem mais ainda as possibilidades, principalmente se aliados a um dispositivo de recarregamento auto-sustentável de baterias.

Neste relato buscamos identificar os materiais e tecnologias utilizados, mostrar alguns dos métodos de monitoramento que utilizamos em campo para o registro sonoro manual e alguns dos equipamentos utilizados, como o sistema "Song Meter"™ (da *Wildlife Acoustics*, *Bioacoustics Monitoring Systems*) e Arbimon™ (*Automated Remote Biodiversity Monitoring Network*) da *Sieve Analytics Inc.* (Aide et al 2013). Nosso foco em particular endereça o uso de equipamentos de gravação e monitoramento bioacústico em florestas tropicais no Sudeste brasileiro.

Os blocos funcionais e os componentes fundamentais que integram estes equipamentos incluem os microfones, o sistema de alimentação, as interfaces de áudio e sistemas de gravação. Abordamos os meios e formatos de registro sonoro mais comuns, bem como características de equipamentos e acessórios de controle que se tornam cada vez mais comuns, como *tablets* e *smartphones*.

No que concerne à análise dos sinais sonoros, identificamos algumas das ferramentas mais utilizadas e as dificuldades clássicas em se caracterizar as vocalizações de mamíferos que frequentemente estão combinadas a um fundo ruidoso.

Colocamos em pauta discussões sobre a qualidade dos registros obtidos e *reports* de uso dos sistemas em campo, destacando a falta de um sistema especializado para estudo de mamíferos terrestres, apontando limitações nos sistemas atuais e perspectivas para inovações tecnológicas desejadas para o futuro.

1. PRIMÓRDIOS DA BIOACÚSTICA NO BRASIL

Para o fundador da bioacústica no Brasil, o Francês Jacques Marie Edme Vielliard[1], a história desta ciência começa com a *Zoophonia* de Hercule Florence (Vielliard & Silva 2006), o inventor da fotografia. Foi ele o autor do relato ilustrado e detalhado da Expedição Langsdorff pela Amazônia e entorno entre os anos de 1825 e 1829. Para preservar os cantos que ouviu, Florence usou notações musicais.

A bioacústica se firmou como ciência no final da década de 1960 com o surgimento dos gravadores "portáteis" de fitas de rolo (Au 1997), ainda muito pesados, mas que serviam bem principalmente aos ornitólogos e os *bird watchers* e, logo depois, atraiu a atenção de interessados em cantos de anuros e insetos. A análise de gravações era inicialmente feita com o auxílio de sonógrafos, e depois, em programas de computador. A partir da década de 1980 os gravadores tornaram-se de fato portáteis, primeiro utilizando fitas no formato digital DAT (*Digital Audio Tape*) e hoje chegam à facilidade de se gravar e transportar uma grande quantidade de sons em algo tão pequeno e leve como um cartão de memória de estado sólido (*flash*).

É muito mais recente o interesse de pesquisadores trabalhando com mamíferos pela bioacústica. Ainda hoje se encontram trabalhos que fazem descrições qualitativas de sons de mamíferos não-humanos, vide as publicações da *Mammalian Species*, tão detalhadas na descrição da anatomia e fisiologia das espécies, mas extremamente vagas na descrição do comportamento vocal. Os estudos que oferecem descrições quantitativas, por meio de análise sonográfica, são, em sua grande maioria, de primatas, morcegos ou cetáceos. Entre as explicações para isso está a dificuldade de se localizar e observar as espécies terrestres. São animais noturnos ou muito ariscos, que evitam a presença humana e que também não vivem restritos a uma pequena área. Tudo isso torna quase impossível a gravação de um mamífero terrestre em campo por vias manuais.

A seguir apontaremos algumas das principais limitações tecnológicas dos equipamentos de gravação acústica e diretrizes para possíveis soluções, com foco no estudo de mamíferos terrestres.

[1] Vide *website* da Fonoteca Neotropical Jacques Vielliard em www.ib.unicamp.br/museu_zoologia/colecao_sonora

2. TECNOLOGIAS USADAS NA PESQUISA BIOACÚSTICA

O cenário funcional clássico de trabalho no campo envolve (a) um sistema para captura e registro, e (b) um sistema para síntese e auralização, controlados por meio de um dispositivo eletrônico/digital micro-processado com memória. O sistema de captura e registro deve dispor de um ou mais microfones e eventualmente poderá conter embutido um sub-sistema de análise acústica e semântica dos sons. O sistema de síntese e auralização, quando usado, deve dispor de alto-falantes para produzir sons no ambiente. Nos dispositivos micro-processados os dados são armazenados em inventórios digitais, que recentemente podem inclusive contar com espaços de armazenamento à distância (ex: *data clouds*) acessíveis por meio de redes de telecomunicações sem fio (*wireless*). Fora do campo, conta-se ainda com plataformas computacionais adicionais para análise e processamento do material coletado.

Em campo a operação clássica manual é ainda considerada a de maior efetividade, mas é a modalidade de operação automática que encontra o maior campo de interesse hoje e, melhorando-se a relação benefício/custo, tende a crescer. Nesta, o monitoramente é de longo termo e pode contar com vários graus de autonomia. Incluem-se nesta modalidade os sistemas conhecidos popularmente como ARS (*Automatic Recording System*) e PAM (*Passive Acoustic Monitoring*). A temporização em sistemas assim pode por exemplo realizar coletas de 1 minuto de áudio a cada 10 minutos (ex: sistema Arbimon™ com 144 1-min. gravações por dia) ou ser programável para disparo por evento acústico detectado.

Na escolha dos microfones é importante considerar o tipo de transdutor que se deseja e os materiais construtivos. Isto determinará quais serão os requisitos para alimentação da polarização (ex: de 3 a 48V tipicamente), para resposta em frequência e sensibilidade, para capacidade de embutir os dispositivos e oferecer entradas adicionais e circuitos de pré-amplificação que condicionem os sinais. Os sistemas de microfonação podem ainda considerar a modelagem do padrão de captura (característica direcional), a focalização (que poderá ter repercussão na capacidade de determinação de posição territorial de espécimes em função da resolução métrica conseguida) e a possibilidade de separação de sinais pelo arranjo espacial (ex: com uso de sistemas multi-microfonados; Blumstein et al 2011). Considere ainda que os microfones podem vir embutidos ou montáveis, usando-se conectores e cabos acessórios, caso que oferece maior flexibilidade e opções ao usuário.

Dispositivos de campo podem ter várias interfaces, portas e conexões para interligar a outros equipamentos e dispositivos. Conectores de 2 canais como miniplugs estéreo (P2) e USB são os mais comuns e de maior penetração, embora interfaces seriais (SPI) e rede IP (CAT5 ou superior) sejam possíveis de uso. Os dispositivos processadores mais populares incluem microcontroladores e microprocessadores (ex: ARM™ Cortex™ 32/64 bits), integrados DSPs, FPGA's e *codecs* dedicados, capazes de suportar vários formatos de mídia digital.

Os formatos mais comuns de registro de áudio incluem o PCM ou ADPCM (não comprimido) e os formatos de compressão perceptual (com perdas) como MP3 (MPEG 1&2 *Audio Layer* III, I & II) e MPEG 4 AAC-LC ou HE-AAC (*level* 3, SBR +PS). Formatos comprimidos sem perdas (ex: FLAC) não são ainda popularmente adotados em plataformas para bioacústica, possivelmente devido ao ainda baixo fator benefício/custo para sua adoção.

Minimamente as plataformas de monitoramento de campo devem contar com microfones, alimentação própria, um dispositivo de gravação, interfaces de controle (de usuário) e invólucros adequados à exposição ambiental. Como exemplo citamos a plataforma "Song Meter"™, basicamente um gravador de campo com recursos de microfonação e autonomia para até 300 horas.

Poucos sistemas que realizam aquisição, processamento e gerenciamento integrados estão atualmente disponíveis para o pesquisador. Entre eles citamos o Arbimon™ apresentado em Aide et al (2013). Dispositivos disponíveis em suas estações de monitoramento podem incluir painéis solares, *iPod's*™, baterias de 12V e reguladores de tensão, controladores de potência, antena (para rádio 900 MHz) e microfones de faixa plena (20-20kHz) e à prova d'água. Nas estações base de análise, encontram-se antenas de recepção, *switches* de rede e acesso à internet, computadores (servidores) para processamento e armazenamento (bases de dados ou repositórios digitais), além de programas para visualização.

Com relação à análise do material sonoro, podemos citar algumas plataformas e recursos mais comumente utilizados por equipes de pesquisadores que buscam principalmente identificar padrões de vocalização, identificar espécies para objetivos de taxonomia e estudos estatísticos sobre populações animais, e ainda identificar mensagens e funções comunicativas para compreender os modelos sociais aplicados às espécies, Este estudo a longo prazo tem impacto na geração de novo conhecimento e na definição de políticas para preservação ambiental.

São utilizados programas (*software*) em plataformas computacionais (inclusive em *grids* computacionais, como o *Biophony Grid Portal*) que perfazem a análise no domínio do tempo e da frequência. Entre eles citamos o XBAT (*Extensible*

Bioacoustic Tool), um pacote de funções para MATLAB™ para análise e anotação de sons, que opera sobre longas gravações usando técnicas de reconhecimento de padrões[2]; o Raven *Interactive Sound Analysis, software* desenvolvido no *Cornell Lab of Ornithology*, junto ao seu programa de pesquisa em bioacústica[3]; e o AviSoft SAS-Lab Pro, programa para investigação da comunicação acústica de animais da empresa *Avisoft Bioacoustics*[4] (Alemanha), com recursos para análise, edição, classificação e síntese sonora.

A título de informação sobre funções de análise, o *software* Raven, por exemplo, possui uma série de recursos como a detecção, seleção de eventos; determinação de níveis (limiares, picos e valores médios); seleção adaptativa de níveis em pontos diferentes de espectrogramas; medidores e exibidores de contornos de frequências; calibração de níveis e escalas do sinal; detectores de silêncio, indicadores de duração e posição temporal absoluta e relativa; envelopes de *fade-in* e *fade-out*; modos de audição/playback com filtros, etc.

3. LIMITAÇÕES TECNOLÓGICAS ATUAIS

Embora os equipamentos remotos automatizados sejam bons, relata-se ainda grande dificuldade com a qualidade e clareza da captura com os microfones e técnicas de captura existentes nestes sistemas. Os sons frequentemente são prejudicados por ruídos difíceis de se isolar em comparação ao que se consegue com sistemas de captura/operação manuais, em que temos mais ajustes a serem feitos presencialmente.

A título de ilustração de um tipo de vocalização de mamífero terrestre brasileiro e suas características, a figura 1 mostra dois registros bem caracterizados de uma vocalização de longa distância do lobo-guará obtidos de animais de cativeiro, manualmente por Flora Balieiro (à esquerda), e através de um sistema de gravação automático (à direita, com muito ruído ao fundo). O aulido da espécie é uma nota que soa grave (maior energia nas frequências inferiores) e ruidosa em uma ampla faixa de frequências. A figura mostra a forma de onda temporal do som acima, e seu espectrograma abaixo.

As vocalizações emitidas por mamíferos vão desde ultrassons (e.g. morcegos) até infrassons (e.g. elefantes) e para cada faixa de frequência de interesse é necessário que o pesquisador use um equipamento específico (Fletcher 2004, McQuay 2008). Microfones, como instrumentos transdutores que convertem energia mecânica (ondas de pressão sonora) em energia elétrica, funcionam também como "filtros" registrando determinadas faixas de frequência e intensidade, e descartando ou atenuando outras (Burbidge et al. 2001).

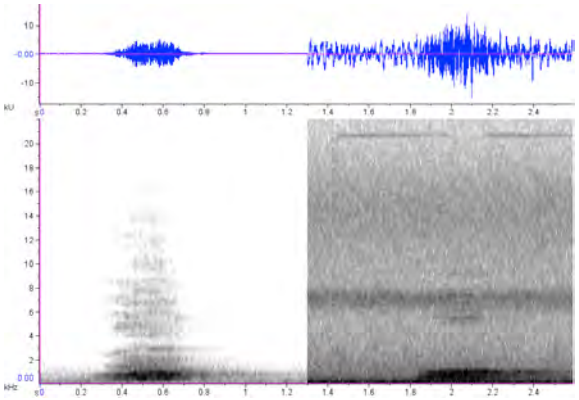


Figura 1 – Duas notas de aulido de lobos-guará em cativeiro, obtidos manualmente com gravador Marantz PMD660 acoplado a Sennheiser ME67 (esq.) e por gravador automático "Song Meter"™ (dir.). Repare na diferença no nível de ruído entre as duas gravações, tomadas em situações similares.

O desenvolvimento da maioria dos microfones portáteis disponíveis no mercado foram impulsionados pela indústria da música e, portanto, criados para captar sons audíveis para humanos (i.e. 20-20kHz) e para operar em ambientes controlados (i.e. longe de condições adversas comuns durante trabalhos em campo com mamíferos terrestres, tais como alta temperatura e umidade). Embora a disponibilidade de microfones capazes de captar a diversidade de vocalizações de mamíferos esteja crescendo, ela ainda é restrita em certa medida.

Os gravadores portáteis digitais foram um importante avanço tecnológico para o estabelecimento da bioacústica como ciência (Au 1997). No entanto, a duração da carga de suas baterias ainda é um fator limitante para trabalhos de campo, pois gravações contínuas sem fonte de alimentação externa normalmente duram poucas horas. Tipicamente encontramos células alcalinas, de NiMH ou Li-íon com duração de até 300 horas. A autonomia entretanto depende de diversos fatores, inclusive da possibilidade de alimentação externa (como no uso de painéis solares) e da boa regulação de tensão e gerenciamento de potência.

Outro parâmetro importante a ser considerado por pesquisadores é a escolha do formato de gravação dos arquivos de áudio e a capacidade de armazenamento do gravador. Gravações em formatos compactados como (ex: MPEG-1/2 *Audio Layer* III, I ou II) ocupam menos espaço, o que aumenta a quantidade de dados que podem ser armazenados. Por outro lado, a baixa fidelidade deste tipo de arquivo de áudio impede que alguns tipos de análises

[2] https://code.google.com/archive/p/xbat-devel/

[3] www.birds.cornell.edu/brp/raven/RavenVersions.html

[4] www.avisoft.com

acústicas sejam realizadas (Fischer et al. 2013). Gravações fidedignas à vocalização original são gravadas em formatos de arquivo não compactados (ex: WAV/AIFF PCM – *Pulse Code Modulation*) e requerem mais espaço de armazenamento, o que pode aumentar muito o custo de certos tipos de pesquisa.

Dentre os principais problemas enfrentados com gravações de mamíferos terrestres em campo destacamos:

(1) microfones com resposta de frequência não plana dentro da faixa de frequências da vocalização de interesse;

(2) baixa relação sinal/ruído das gravações, especialmente quando lidamos com microfones onidirecionais em ambientes ruidosos;

(3) sensibilidade do microfone que pode ser permanentemente afetada quando exposto a temperaturas acima da faixa de tolerância de seus parâmetros operacionais;

(4) baixa resiliência em condições de alta umidade (especialmente quando lidamos com microfones condensadores);

(5) alto consumo de energia mesmo nos gravadores digitais, o que restringe o tempo de gravação em campo;

(6) capacidade restrita de armazenamento de dados acústicos para monitoramento de longo prazo por alguns gravadores digitais para determinados tipos de pesquisa;

Com relação aos métodos de análise de sinais e detecção de vocalizações, os problemas mais comuns incluem a ocorrência e as dificuldades de tratamento de casos de falsos positivos e falsos negativos. Exemplos de causas de falsos positivos incluem ruídos meteorológicos (ex: vento), mecânicos e vocalizações de espécies parecidas. Como exemplo de causas para falsos negativos citamos a microfonação deficiente (em que vocalizações distantes, não sendo marcadas, não serão achadas) e parametrização excessivamente restritiva para evitar falsos positivos (que acaba por elevar o número de falsos negativos, como relata Aide et al 2013).

4. PERSPECTIVAS PARA INOVAÇÕES TECNOLÓGICAS

Nesta sessão desejamos sugerir algumas diretrizes a serem observadas no desenvolvimento de soluções para as limitações atuais. Soluções tecnológicas focadas em gravações acústicas de mamíferos terrestres em ambientes naturais devem levar em conta aspectos diretamente relacionados às espécies de interesse e ao ambiente em que vivem. Elencamos abaixo alguns aspectos fundamentais:

(1) conhecer aspectos do repertório vocal das espécies de interesse (ex: qual é a faixa de frequências que se quer amostrar; qual é a amplitude média das vocalizações alvo);

(2) caracterização da biofonia do ambiente (i.e. leque de sons da paisagem acústica que tem origem biológica). Esta etapa servirá, por exemplo, para verificar previamente se existem outras vocalizações que podem se sobrepor às de interesse e que poderão ser eliminadas pelo equipamento;

(3) caracterização da geofonia (i.e. sons provenientes de fenômenos naturais como chuva e vento; Krause 1993) e antropofonia (i.e. sons oriundos de atividades humanas; Krause 1993) do ambiente, que podem interferir no registro de sons de interesse;

(4) características climáticas, fitofisionômicas e de relevo do ambiente que podem influenciar na propagação do som e durabilidade do equipamento de gravação;

(5) tipos de pesquisa que se beneficiarão com a tecnologia a ser empregada. Por exemplo, pesquisas com levantamento de fauna podem se beneficiar de gravadores autônomos (i.e. que fazem gravações programadas sem a necessidade da presença do pesquisador em campo) que tenham microfones capazes de amostrar amplas faixas de frequências, enquanto pesquisas que buscam gravar uma única espécie ou vocalização podem se beneficiar de microfones desenhados especialmente para este fim.

5. CONSIDERAÇÕES FINAIS

O avanço da tecnologia eletrônica e o aperfeiçoamento de equipamentos de gravação e análise acústica tem ampliado as fronteiras de pesquisas na bioacústica. Todavia existem ainda barreiras tecnológicas que limitam a transposição no uso destes para o estudo de mamíferos terrestres. Não se trata de melhorar o sucesso no registro de sons por eles produzidos somente ajustando a qualidade dos instrumentos de gravação. É necessário se investir em automação no processo. Assim, existe hoje uma demanda crescente de equipamentos pensados para registro de vocalizações em campo na ausência do pesquisador, e com uma cobertura maior de alcance.

Equipamentos para monitoramento de campo na forma de plataformas autônomas parece ser uma tendência, em que a autonomia, além da questão da durabilidade de baterias ou do tempo máximo de operação sem intervenção humana, acrescenta funções de decisão e análise, de auto-gestão, ou de maior controlabilidade à distância. Não obstante à aparente robotização dos dispositivos que isto possa sugerir, caber notar que em casos onde a intervenção do aparelho no meio ambiente for maior, também será maior a necessidade de disfarçá-lo.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Faria, R. R. A.; Gasco, A. D. C.; Monticelli, P. F. *Entre notas e alaridos: perspectivas na bioacústica de mamíferos terrestres*. In: 15th Brazilian Symposium on Computer Music – SBCM 2015, Campinas, 2015. Disponível em: http://compmus.ime.usp.br/sbcm2015.

[2] Blumstein, D. T. et al. *Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus*. Journal of Applied Ecology, v.48, p.758-767, 2011.

[3] Au, W.W.L. *Some hot topics in animal bioacoustics*. Journal of Acoustical Society of American, 1997.

[4] Burbidge, A; Iannantuoni, M. & Riddett, P. *An introduction to wildlife sound recording*. 1ª revision, Wildlife Sound Recording Society, 2001.

[5] Fischer, J., Noser, R., & Hammerschmidt, K. *Bioacoustic field research: a primer to acoustic analysis and playback experiments with primates*. American Journal of Primatology, 75(7), 643-663, 2013.

[6] Fletcher, N.H. *A simple frequency-scaling rule for animal communication*. Journal Acoustical Society of America, 115: 2334-2338, 2004

[7] Krause, B.L. *The Niche Hypothesis: A Virtual Symphony of Animal Sounds, the Origins of Musical Expression and the Health of Habitats*. The Soundscape Newsletter, 6, 6-10, 1993.

[8] McQuay, B. *Audio equipment*. Macaulay Library, Cornell Lab of Ornithology. Ithaca, NY, 2008.

[9] Vielliard, J. M. E. & Silva, M. L. A. *Bioacústica como ferramenta de pesquisa em Comportamento animal*. In: Grauben Assis, Regina Brito e William Lee Martin. Estudos do Comportamento II. Belém: Editora da UFPA, 2006.

[10] Aide et al. *Real-time bioacoustics monitoring and automated species identification*. PeerJ 1:e103; 2013. DOI 10.7717/peerj.103.

Sociedade de Engenharia de Áudio
Artigo de Congresso

Apresentado no 14º Congresso de Engenharia de Áudio
20ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2016, São Paulo, SP

Este artigo foi reproduzido do original final entregue pelo autor, sem edições, correções ou considerações feitas pelo comitê técnico. A AES Brasil não se responsabiliza pelo conteúdo. Outros artigos podem ser adquiridos através da Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA, www.aes.org. Informações sobre a seção Brasileira podem ser obtidas em www.aesbrasil.org. Todos os direitos são reservados. Não é permitida a reprodução total ou parcial deste artigo sem autorização expressa da AES Brasil.

Qual o futuro do MP3? Áudio espacial e codificação orientada a objetos

Bruno Masiero[1]

[1] Universidade Estadual de Campinas (Unicamp),
Faculdade de Engenharia Elétrica e Computação, Departamento de Comunicações
Campinas, SP, 13083-852, Brasil

masiero@unicamp.br

RESUMO

Os avanços tecnológicos das últimas décadas resultaram em uma mudança significativa na nossa relação com a mídia. A popularização da Internet permitiu que uma enorme quantidade de documentos de áudio estejam disponíveis a qualquer hora e de praticamente qualquer lugar, uma condição que só foi atingida graças a técnicas de codificação que permitem a compactação de arquivos de áudio, sendo a codificação MP3 a porta estandarte deste processo. Por outro lado, a crescente popularização de sistemas surround para uso doméstico e do uso de fones-de-ouvido associados a telefones portáteis com grande capacidade de processamento permite a integração da informação espacial como um elemento integrante do processo e do produto artístico. Isto significa que se faz necessário técnicas de codificação pensadas para arquivos de áudio espacial. Neste documento iremos apresentar as principais técnicas de captura e de reprodução de áudio espacial, que são essenciais para se entender os requisitos de um formato de codificação de áudio espacial e ao fim iremos discutir uma proposta para atender a estes requisitos, a codificação baseada em objetos.

0 INTRODUÇÃO

A música está intimamente ligada com a evolução humana, sendo, em muitas culturas, um elemento fundamental em cerimônias, festividades e rituais religiosos. Apesar de cada cultura ter sua própria interpretação do que seria música, um fato comum ao evento musical é que a apreciação da música era um evento *ao vivo*, ou seja, para apreciar a música o ouvinte precisava necessariamente estar presente no local e instante que a música esta sendo executada.

Mas isto começou a mudar no fim do século XIX com novas invenções como o *Théâtrophone*, que permitia que performances de óperas ou concertos orquestrais pudessem ser escutados pelo telefone, dispositivo que

havia sido patenteado em 1876 por Alexander Graham Bell. Isto era realizado colocando-se vários telefones em uma sala ou teatro que eram conectados as telefones (operados por moedinhas) geralmente disponíveis em hotéis e bares. Percebe-se que a ideia de vender música a distância já nascia aí, apesar de que ainda limitada a execuções ao vivo.

Mas tudo isto mudou com o *Fonógrafo*, inventado em 1877 por Thomas Edison. Esta invenção deu início a uma verdadeira revolução na forma como a música poderia ser "consumida", já que o fonógrafo permitia que a música pudesse ser registrada (gravada) para posterior reprodução, possivelmente em outro local. Ou seja, esta invenção decretou o fim da exclusividade da música ao vivo e permitiu a criação de uma nova indústria, a indústria fonográfica [1].

As primeiras gravações apresentavam uma qualidade sonora muito baixa, tanto que, após sua invenção, Thomas Edison abandou o projeto do fonógrafo por dez até que uma nova tecnologia foi desenvolvida que permitisse uma gravação de melhor qualidade [1]. Ainda assim, nas primeiras décadas da indústria fonográfica a qualidade da reprodução era tão baixa e, por outro lado, a possibilidade de escutar música em casa sem a presença de músicos tão incrível que neste momento não importava o fato de o sistema ter um único alto-falante e, portanto, não reproduzir as características espaciais do som que poderiam ser apreciadas durante uma performance ao vivo.

Conforme as técnicas de gravação e transmissão[1] foram avançando e o custo dos transdutores caiu, a evolução lógica foi incluir informação espacial ao conteúdo musical. Note que o Théâtrophone já permitia o uso de duas linhas telefônicas (ao dobro do preço), para serem escutadas uma com cada ouvido, dando assim uma certa sensação de espacialidade.

Em de 1931 Alan Blumlein patenteou um sistema que é considerado a gênese do que hoje conhecemos como estéreo [2]. Neste documento Blumlein sugere, entre muitas outras coisas, a gravação com dois microfones e a posterior reprodução com dois alto-falantes. Muitos experimentos e tentativas foram feitas por cientistas, engenheiros e audiófilos para desenvolver um sistema de reprodução de alta-fidelidade com dois canais. Mas foi só por volta da década de 1960 que sistemas estereofônicos[2] se tornaram economicamente viáveis e que a gravação em estéreo se tornou o padrão *de facto* da indústria fonográfica, o que permanece sendo até os dias de hoje.

Após a introdução do cinema falado, a indústria cinematográfica sempre se manteve na vanguarda da reprodução sonora em alta-fidelidade e logo assimilou o uso da estereofonia. Inicialmente, as gravações eram feitas em dois ou mais canais para facilitar a associação da imagem com o som, mas logo cinemas passaram a usar a reprodução multi-canal, muitas vezes colocando o diálogo em um canal ligado a um alto-falante no centro da tela e a trilha sonora ligada a dois alto-falantes nos cantos da tela. Com o passar do tempo foram adicionados mais canais para incrementar a sensação de espacialidade até a indústria cinematográfica chegar ao seu padrão *de facto*, o som *surround* ou 5.1, que depois ganhou as casas com o advento dos *home theatres*.

Mas a busca pela sensação de imersão no espaço sonoro, ou seja, pelo realismo, não parou por aí. Novos sistemas de reprodução sonora continuaram a ser desenvolvidos, como por exemplo o Ambisonics [3] e o *wave field synthesis* (WFS) [4]. Estes sistemas de reprodução espacial requerem um número elevado de canais de transmissão, sendo, portanto, sistemas cada vez mais complexos e que requerem técnicas de gravação também cada vez mais complexas.

A indústria fonográfica sofreu uma nova guinada com a era digital e com o advento do compartilhamento de arquivos de áudio pela internet, que resultou em um novo modelo de consumo de música, que não é mais comprada em uma mídia física como o vinil ou o CD, mas sim é baixada diretamente da rede de computadores. Como esta nova forma de consumo possui uma restrição intrínseca na taxa de transmissão de dados da rede, se tornou necessário a redução do tamanho dos arquivos de áudio. Por esta razão técnicas de compressão de áudio com perdas, como a famosa MP3, se popularizaram para permitir o envio de arquivos de áudio por canais com taxas de transmissão reduzidas.

Pelo que foi apresentado, nota-se um claro desencontro entre técnicas de reprodução espacial mais modernas e os novos hábitos de consumo de conteúdo musical, já que para transmitir o elevado número de canais de áudio requeridos pelas novas técnicas se faz necessária uma taxa de transmissão muito alta, o que ainda continua sendo uma situação restritiva, ainda mais com o aumento do uso da internet por redes de telefonia celular.

Ou seja, se faz necessário uma forma de compactação eficiente dos arquivos de áudio espacial. Por outro lado, como acima mencionado, o uso cada vez mais frequente de telefones celulares para acesso à internet também resulta num incremento no número de usuários que escutam música por fones de ouvido.

Um contraponto ao aumento desenfreado no número de canais é o sistema biauricular [5], que assume que são necessário apenas dois canais de gravação e reprodução (cada canal reproduzindo diretamente em um ouvido através do uso de fones-de-ouvido), uma vez que nós possuímos apenas dois ouvidos. Este sistema já era estudo, por exemplo, por Blumlein na década de 1930 e continuou como objeto de curiosidade de muitos estudioso, tendo tido algumas tentativas frustradas de

[1] Estações comerciais de rádio começam a transmitir conteúdo musical e de notícias no início da década de 1920, após o fim da Primeira Guerra Mundial.

[2] Do grego, *stereo* significa "sólido". Ou seja, na concepção do nome, o sistema estéreo não é limitado a dois canais, mas se tratava, sim, de um sistema de reprodução sonora espacial. Ao menos esta era a impressão quando saímos da monofonia!

comercialização. Atualmente, contudo, o sistema biauricular vive um período de renascimento.

Com tudo que foi exposto, percebe-se que há uma nova demanda para a codificação de áudio que deve conseguir codificar gravações com conteúdo espacial de forma a ser transmitida em canais com baixas taxas de transmissão e cujo mesmo arquivo possa ser reproduzido através de diferentes técnicas de reprodução. Tendo em vista estes requisitos o *Moving Picture Expert Group* (MPEG) propôs um novo paradigma de codificação, uma codificação baseada em objetos sonoros ao invés de canais de áudio com o intuito de atender à demanda de reduzir a taxa de informação sendo transmitida e ao mesmo tempo permitir que o arquivo de áudio possa ser reproduzido da melhor maneira possível em sistemas de reprodução de diferentes complexidades, ou seja, que o sistema de áudio sejam escaláveis.

Este artigo pretende dar uma visão geral sobre as principais técnicas de gravação e reprodução sonora para nortear uma discussão sobre os requisitos da codificação de áudio espacial orientada a objetos, ou, em inglês, *Spatial Audio Object Coding* (SAOC).

1 CAPTAÇÃO ESPACIAL DO SOM

A captação e o armazenamento de um evento sonoro só se faz possível com o uso de um transdutor que converta a onda sonora em uma onda em algum outro meio. O primeiro dispositivo a realizar a transdução de uma onda sonora em uma onda elétrica foi desenvolvido na década de 1870 por David Hughes [6] e proporcionou uma melhoria tão considerável na transdução do som que só a partir de sua invenção que a telefonia, a radiodifusão e a indústria fonográfica puderam ganhar corpo.

Durante a primeira metade do século XX as gravações eram feitas usando-se, predominantemente, apenas um microfone. Realizar a gravação com um único microfone é equivalente a dizer que a onda sonora de interesse foi amostrada em um único ponto. Se quisermos melhorar a informação espacial uma maneira seria amostrar o espaço em mais pontos, ou seja, usar mais microfones [7]. Por exemplo, em sua patente de 1931, Blumlein já previa o uso de dois microfones para a gravação estereofônica [2].

Com a popularização do estéreo e posterior desenvolvimento de outras técnicas de reprodução espacial (ver seção 2) diversos outros tipos de arranjos[3] de microfones foram desenvolvidos para atender a diferentes propósitos. Ainda assim, podemos classificar estes arranjos de microfones para captação espacial entre dois paradigmas principais: arranjos compactos ou arranjos distribuídos. Dentro do primeiro grupo destacamos ainda um tipo de arranjo especial para captação biauricular, que usa um manequim (com cabeça e torço) provido de dois microfones onde seriam as entradas do canais auditivos.

[3] Um arranjo de microfones é um agrupamento ordenado de dois ou mais microfones.

1.1 Captação biauricular

O nosso sistema auditivo se faz valer do fato de que possuímos duas orelhas (posicionadas em lados opostos da cabeça) para determinar a direção de chegada de uma frente de onda sonora, o que é conhecido por audição biauricular [8]. Uma frente de onda ao atingir nossa cabeça irá gerar um evento sonoro diferente em cada ouvido, dependente do ângulo de chegada desta onda.

Blauert define a tecnologia biauricular da seguinte maneira [8]: "A tecnologia biauricular são métodos que envolvem o sinal acústico captados pelos dois ouvidos de um ouvinte para algum objetivo prático, como, por exemplo, gravação, análise, síntese, processamento, avaliação e reprodução destes sinais."

A ideia principal é que todos os eventos auditivos percebidas por nós são extraídos dos sinais pelos nossos dois ouvidos. Se formos capazes de gravar os sinais que adentram nosso canal auditivo, os sinais biauriculares, e posteriormente reproduzir exatamente os mesmos sinas na mesma posição (com o uso, por exemplo, de fones-de-ouvido), o evento auditivo gerado por esta reprodução deve ser o mesmo que aquele causado durante a gravação.

É importante ressaltar que a percepção espacial está intimamente ligada às características anatômicas do ouvinte, representadas pela função de transferência anatômica (*head-related transfer function*, HRTF) [9, 10]. Portanto, a gravação biauricular deveria, idealmente, ser feita com a própria cabeça do ouvinte, o que, venhamos e convenhamos, não é sempre praticável.

Gravações biauriculares acabam sendo, em geral, feitas com uma cabeça artificial, ou manequim [11]. No entanto, apesar de algumas experimentações, como por exemplo o álbum *Street Hassle* de Lou Reed, lançado em 1978 como o primeiro disco produzido comercialmente com tecnologia biauricular, esta tecnologia não logrou sucesso comercial e ficou relegada a cientistas e entusiastas. Apenas recentemente, com a popularização do uso de fones-de-ouvido por causa da miniaturização dos sistemas de áudio portátil é que a técnica biauricular tem vivido uma renascença.

1.2 Arranjos compactos de dois canais

A audição biauricular usa dois parâmetros para estimar a direção de chegada da onda: a diferença de amplitude entre os ouvidos (*interaural level difference*, ILD) ou a diferença do tempo tempo de chegada[4] da onda aos ouvidos (*interaural time difference*, ITD). Como veremos na seção 2.3.1, a reprodução estereofônica busca reproduzir estes parâmetros de forma a posicionar fontes virtuais entre os alto-falantes.

Os arranjos para gravação estereofônicas buscam justamente capturar uma diferença de fase ou amplitude entre os dois canais de gravação. Note que este tipo de arranjo apresenta os microfones agrupados de forma a

[4] Uma diferença no tempo de chegada pode ser interpretado como uma diferença na fase do espectro deste sinal.

captar o som em uma pequena região do espaço, por isso o nome *compacto*.

Listamos abaixo os três tipos de arranjos para gravação estereofônica mais usuais.

Diferença de intensidade Também conhecido como arranjo X-Y ou par de Blumlein, este arranjo usa dois microfones direcionais (com diretividade do tipo cardioide ou figura-de-oito) posicionados perpendicularmente entre si e com suas membranas justapostas. Esta configuração dá ganhos diferentes para o sinal de acordo com sua direção de chegada por causa de sua diretividade não ser omnidirecional. Isto significa que os dois canais resultantes irão apresentar diferença de amplitude, mas não de fase (por estarem justapostos) [12].

Diferença de tempo de chegada Também conhecido como arranjo A-B, este arranjo usa dois microfones em paralelo e distantes (20 a 50 cm) um do outro. Esta configuração irá gerar dois sinais com diferença de nível sonoro e principalmente de tempo de chega. Ou seja, os dois canais resultantes irão apresentar diferença de fase em seu espectro [12].

Diferença diretividade Também conhecido como arranjo M-S, de *mid/side*, este arranjo usa um microfone omnidirecional e outro com diretividade figura-de-oito justapostos. Diferentemente das configuração anteriores, neste caso os canais de reprodução não são os mesmos da gravação, mas sim a combinação linear (some e subtração) dos dois sinais gravados. Este tipo de gravação permite uma maior flexibilidade para o pós-processamento do sinal, além de apresentar compatibilidade direta com a reprodução monofônica [12].

1.3 Arranjos compactos multicanais

A técnica de reprodução *surround*, ou 5.1, se popularizou nas últimas décadas, se tornando o padrão *de facto* da indústria cinematográfica. Em geral as faixa de áudio e efeitos sonoros de um filme são mixadas em pós-produção, onde as fontes sonoras são distribuídas no espaço usando a técnica de panorama (seção 2.3.1). Mas para situações onde se deseja gravar uma cena sonora para reprodução em 5.1 sem edição ou pós-processamento usa-se um arranjo com cinco microfones tipo cardioides, cada um apontando para cada uma das posições padronizadas para os alto-falantes no arranjo 5.1 (seção 2.3.1) [13]. Note que se utilizarmos apenas os sinais dos microfones apontando para -45° e 45° temos uma configuração X-Y e, portanto, compatibilidade com a reprodução em estéreo.

Analisando os arranjos compactos de dois canais da seção anterior verifica-se que os arranjos X-Y e A-B buscam reproduzir características de como o sistema auditivo humano extrai informações espaciais (ITD e ILD) enquanto o arranjo M-S busca descrever a variação local do campo acústico, uma vez que a diretividade figura-de-oito pode ser interpretada como a primeira derivada de um microfone omnidirecional em uma dada direção.

Podemos imaginar que queremos refinar a descrição da variação espacial do campo sonoro. Para isto, seria necessário posicionar três microfones figura-de-oito ortogonalmente entre si (frente-trás, direita-esquerda, cima-baixo), além do microfone omnidirecional. Mas esta é uma construção difícil de ser feita.

Michael Gerzon, um matemático que trabalhava no campo da física quântica, propôs uma maneira de solucionar este problema usando uma ferramenta matemática do seu campo de trabalho, a decomposição em harmônicas esféricas (DHE). Esta ferramenta permitia transformar uma gravação feita com um arranjo de microfones em forma de um tetraedro em uma gravação que teria sido feita com um microfone do tipo M-S expandido discutido acima.

O primeiro arranjo tetraédrico de microfones, batizado de *Soundfield microphone* começou a ser comercializado em 1978. Os sinais gerados pelo arranjo são conhecidos por formato-A enquanto que os sinais obtidos após a transformação pela DHE são conhecidos por formato-B. O formato-B é então pós-processado e eventualmente distribuído para alto-falantes. Gerzon também propôs uma maneira de realizar este pós-processamento baseada na DHE, que deu o nome de Ambisonics (seção 2.3.1).

Apesar de o soundfield ser construído com apenas quatro microfones, a teoria do DHE pode ser aplicada para arranjos com um maior número de microfones, permitindo extrair-se um maior número de harmônicas e com isso realizar uma melhor descrição da variação espacial do campo acústico. Esta técnica de gravação vem ganhando cada vez mais prestígio e novos microfones vêm sendo desenvolvidos para este fim. Um desdes arranjos compactos está disponível comercialmente, o *Eigenmike*, com 32 microfones distribuídos em uma esfera.

1.4 Arranjos distribuídos

Vimos que arranjos compactos são projetados para representar localmente algum parâmetro do campo sonoro. Já arranjos distribuídos, como o próprio nome diz, estão distribuídos pelo espaço e, portanto, permitem amostrar uma maior região do campo sonoro. Este tipo de arranjo é bastante usado para imageamento acústico [14] e imageamento sísmico [15], mas praticamente não é usado pela indústria fonográfica.

Para gravações de grandes grupos tocando ao vivo, como orquestras sinfônica, técnicos de som geralmente lançam mão de uma espécie de arranjo distribuído, posicionando diversos microfones espalhados pelo orquestra. Já quando a gravação é feita em estúdio, é comum o instrumentos serem gravados individualmente ou em pequenos grupos (close miking). Neste segundo caso as trilhas gravadas são mixadas posteriormente (pós-produção) e a maneira como são mixadas levam em conta o tipo de sistema que deverá ser usado para reprodução.

Interessante é notar que esta técnica de captação individual é a que mais se aproxima do conceito de objetos sonoros explorado pela codificação SAOC, já que cada instrumento ou naipe é considerado um objeto sonoro.

2 REPRODUÇÃO ESPACIAL DO SOM

Passamos agora da captação para o outro extremo da cadeia fonográfica, a reprodução, que por muitas décadas era feita usando-se uma única fonte acústica. Inicialmente estas fontes eram feitas usando-se uma corneta metálica ligada a uma membrana que era excitada por um objeto pontiagudo. Alguns fonógrafos também usavam um sistema de ar comprimido para amplificar o som produzido. Mas foi só após o desenvolvimento dos alto-falantes de bobina móvel que se tornou possível a reprodução sonora de alta fidelidade.

O próximo passo na busca da melhoria destes sistemas foi introduzir *espacialidade* ao som reproduzido. Diversos tentativas foram feitas para desenvolver um sistema de reprodução de alta-fidelidade com dois (ou mais) canais. Mas foi só por volta da década de 1960 que sistemas estereofônicos (com dois canais) se tornaram economicamente viáveis.

Invariavelmente, as técnicas de reprodução continuaram sua evolução através do aumento do número de canais, o que faz necessário o uso de arranjos com múltiplos alto-falantes. Assim como fizemos com os arranjos de microfones, podemos classificar os arranjos de alto-falantes em arranjos compactos ou distribuídos, sendo a segunda opção a mais comumente utilizado.

Os sinais reproduzidos pelos arranjos distribuídos podem ser classificados sob três paradigmas distintos: biauricular, panorama ou síntese de campo.

Como já apresentado na seção 1.1, a reprodução biauricular apresenta sons diferentes para cada ouvido do ouvinte, garantindo assim a reprodução dos parâmetros interaurais. Este tipo de reprodução é geralmente feito através de fones-de-ouvido, que podem ser interpretados como um arranjo compacto individual (de dois alto-falantes). Também é possível a reprodução de sinais biauriculares através de dois ou mais alto-falantes, mas para este fim é necessário o uso de um banco de filtros para cancelamento de diafonia (*crosstalk cancellation*, CTC) [16].

Os sistemas tipo panorama buscam recriar em uma área do espaço (o *sweet spot*) diferenças interaurais (ILD e ITD) de forma que o ouvinte ali presente tenha uma impressão de espacialidade. Por outro lado, sistemas tipo síntese de campo buscam efetivamente recriar dentro do ambiente de reprodução o campo sonoro existente no local de gravação.

2.1 Reprodução individual

A reprodução de gravações biauriculares pode ser feita de duas maneiras: com fones-de-ouvido ou com alto-falantes.

Fones-de-ouvido A reprodução através de fones de ouvido é, em primeira instância, mais simples, uma vez que a separação de canal obtida desta maneira é quase ideal. No entanto, cada fone-de-ouvido possui uma resposta em frequência distinta, que, ainda por cima, varia de acordo com a geometria do ouvido externo do ouvinte e também com o posicionamento do dispositivo [17]. Desta maneira, é desejável que a equalização do fone-de-ouvido seja individualizada [18].

Um aspecto crítico para a qualidade da reprodução biauricular é que estamos, mesmo que inconscientemente, constantemente nos movimentando. No entanto, a disposição da cena sonora apresentada continua fixa em relação à posição do fone-de-ouvido. Isto significa que ao nos movimentarmos as fontes sonoras irão se mover conosco ao invés de ficarem fixas no espaço, podendo resultar no desaparecimento da ilusão acústica criada pela reprodução biauricular.

Para solucionar este problema se faz necessário o uso de um sistema dinâmico, que rastreie a posição da cabeça do ouvinte e compense seus movimentos. Isto é praticável com sinais biauricular que tenham sido auralizados [19], de forma a permitir que a posição das fontes seja atualizada [20]. Mais uma vez motivando um sistema de armazenamento de áudio orientado a objeto que realize a auralização da cena sonora no fim da cadeia de reprodução.

CTC Se os sinais biauriculares fossem reproduzidos através de dois alto-falantes teríamos o efeito de diafonia, ou seja, o sinal esquerdo, que deveria ser ouvido apenas pelo ouvido esquerdo, ao ser reproduzido por um alto-falante à esquerda do ouvinte atingirá primeiro o ouvido esquerdo, mas depois também o direito e vice-versa para o sinal do ouvido direito. Esta mistura dos sinais resulta na distorção dos parâmetros interaurais contidos no sinal e no consequente desaparecimento da sensação de espacialidade contida no sinal biauricular.

Para anular este efeito de diafonia são necessários filtros CTC. Estes filtros misturam os dois sinais biauriculares gerando o que chamamos de sinais transaurais. Os sinais transaurais são então reproduzidos por alto-falantes e ao chegar aos ouvidos do espectador eles interagem de forma que o sinal resultante é o próprio sinal biauricular [21, 22, 23].

Estes filtros, no entanto, dependem da posição dos alto-falantes em relação ao ouvinte e da própria HRTF (que é individual). Se filtros não individualizados forem usados a qualidade da localização obtida com o sistema fica comprometida [24, 25].

Como mencionado na seção anterior estamos constantemente nos movendo. Isto implica que a direção dos alto-falantes em relação à cabeça do ouvinte também está em constante variação e por isso os filtros CTC precisam ser constantemente atualizados levando em conta a posição do ouvinte em relação ao arranjo de alto-falantes [20, 26].

2.2 Arranjos compactos

Chamamos de arranjos compactos de alto-falantes dispositivos com um grande número de transdutores posicionados muito próximos entre si. Geralmente estes transdutores são fixados sobre uma esfera rígida [27, 28] ou sobre sólidos platônicos (como o dodecaedro ou o icosaedro) [29] e permitem direcionar o feixe sonoro, criando fontes móveis e super-diretivas.

Estes dispositivos são muito usados para caracterização acústica de salas, como o cálculo do tempo de reverberação ou o coeficiente de transmissão de paredes [30]. No entanto, eles não costumam ser usados para reprodução musical e por esta razão não serão aqui discutidos em detalhes.

2.3 Arranjos distribuídos

Os sistemas de reprodução espacial mais frequentemente utilizados são arranjos distribuídos de dois ou mais alto-falantes. O termo "distribuídos" é usado porque nesta configuração os alto-falantes estão distribuídos ao redor dos ouvinte. Este tipo de sistema geralmente possui um *sweet spot*, um ponto (geralmente no centro do arranjo) que garante a melhor qualidade de reprodução espacial.

Os sistemas de reprodução com arranjos distribuídos se baseiam em dois paradigmas: panorama e síntese de campo, que serão discutidos a seguir.

2.3.1 Panorama

Como já discutido na seção 1.2, nós localizamos sons no plano transversal através dos parâmetros interaurais ILD e ITD. A dependência destes dois parâmetros foi denominada de "teoria duplex" por Lord Rayleigh, primeiro cientista a descrever este efeito. Posteriormente determinou-se que a ITD é utilizada para localização de baixas frequências (até ≈1500 Hz) e a ILD para a localização de altas frequências (a partir de ≈700 Hz).

Algumas das técnicas de captação discutidas na seção 1.2 buscam reproduzir esta diferença de amplitude e fase entre os dois canais gravados, que serão posteriormente reproduzidos por dois alto-falantes. Mas quando é usada a técnica de *close miking*, como é possível "posicionar" esta fonte sonora entre os alto-falantes?

A técnica mais utilizada para este fim é o panorama, que envia o mesmo sinal com ganhos distintos para cada falante. Quando são usados dois alto-falantes esta técnica ganha o nome de panorama estéreo, ou simplesmente estéreo.

Estéreo Vamos assumir que tenhamos dois alto-falantes posicionados em duas pontas de um triângulo equilátero e o ouvinte se encontra na outra ponta. A técnica de panorama consiste em enviar o mesmo sinal com ganhos diferentes para cada alto-falante, de forma a posicionar o som entre aos falantes.

Mas por que percebemos um evento sonoro entre os falantes e não dois eventos sonoros distintos?

Por causa da geometria da cabeça do ouvinte, o som gerado pela fonte à esquerda do ouvinte irá chegar primeiro ao seu ouvido esquerdo e depois ao seu ouvido direito. A cabeça do ouvinte também irá cria uma sombra acústica tal que o som chegará ao ouvido esquerdo com maior intensidade que no ouvido direito. O inverso é válido para a fonte à direita do ouvinte.

Suponhamos que o sinal enviado para à fonte da esquerda tem um ganho de 3 dB em relação à fonte da direita. Para altas frequências a sombra gerada pela cabeça no ouvido oposto (maior do que 10 dB para frequências acima de 2 kHz) faz com que o som direto predomine. Neste caso, o som no ouvido esquerdo será 3 dB maior que o do ouvido direito, gerando portanto uma ILD que será interpretada como um som vindo de uma direção mais à esquerda.

Mas e para as baixas frequências onda a sombra acústica não é tão pronunciada? Neste caso o sinal gerado por cada fonte irá se combinar em cada um dos ouvidos de forma vetorial, resultando em um sinal de mesma amplitude mas com diferença de fase [31]. Ou seja, a diferença de amplitude resulta, para baixas frequências, em uma ITD, que é justamente o parâmetro relevante para o nosso sistema auditivo nesta faixa de frequência.

Mas como definir qual a relação de ganho necessária para posicionar a fonte em uma dada direção? Testes subjetivos [32] mostraram que a curva que melhor estima a direção ϕ percebida para uma dada combinação de ganhos g_R e g_L é

$$\frac{\tan\phi}{\tan\phi_0} = \frac{g_R - g_L}{g_R + g_L}. \tag{1}$$

Aqui assume-se que o ouvinte esteja centrado em relação aos falantes e que estes estão posicionados em um ângulo de ϕ_0 em relação à direção de visada do ouvinte.

Esta técnica possui a vantagem de ser de muito fácil implementação, no entanto, as fontes virtuais geradas por panorama são percebidas como largas e apresentam bastante coloração espectral [33]. Felizmente, esta coloração só é percebida em ambientes anecoicos. Quando ouvido dentro de uma sala a reverberação acaba mascarando este efeito, resultando em um sistema simples e com boa localização, provavelmente a razão de seu grande sucesso.

Surround Este é o termo utilizado para as expansões do panorama de amplitude para um formato envolvente e comercialmente viável, mais frequentemente utilizado pela industria cinematográfica. Como não houve uma padronização neste sentido, muitos sistemas diferentes foram desenvolvidos e comercializados, diferindo principalmente em como a faixa sonora era codificada e gravada/sincronizada nos rolos de filme. No que tange o arranjo, ouve uma convergência pelo formato conhecido por 5.1, com uma fonte no centro (0°), duas fontes em ±45° e duas fontes em ±110°.

Esta geometria do arranjo acabou sendo escolhida porque pares de falantes na lateral do ouvinte (por exem-

plo em 45° e 135°) não permitem realizar a tradução de ILD em ITD, ou seja, não é possível posicionar fontes laterais. O surround compensa este fato colocando os falantes posteriores mais abertos, ou seja, mais próximos ao eixo dos ouvidos, gerando assim uma fonte estável nas laterais. Além disto, um canal central é usado para garantir que diálogos seja ouvidos sempre vindo da tela, mesmo que o ouvinte não esteja mais no sweet spot.

VBAP Pulkki estendeu o conceito de panorama para arranjos tridimensionais. Ele sugeriu que o trio de alto-falantes mais próximos da direção da fonte virtual deveriam ser selecionados, deixando o restante dos falantes do arranjo inativos. O ganho destas três fontes era então calculado como os ganhos que permitiam descrever a direção da fonte virtual como combinação linear da direção das três fontes ativas, daí o nome *vector-base amplitude panning* (VBAP).

Note que a argumentação de que pares de fontes próximos do eixo dos ouvidos não permitem gerar fontes virtuais entre eles também é válida para pares na vertical. Pulkki já havia notado isto, mas ele argumenta que, apesar de a percepção de elevação variar de ouvinte para ouvinte, ela ainda assim fica sempre limitada dentro da área delimitada pelas três fontes ativas [34].

Ambisonics Enquanto VBAP propõem usar três alto-falantes por direção, a técnica Ambisonics propõem o uso de todos os alto-falantes de um arranjo distribuído para sintetizar uma fonte virtual.

Conforme discutido na seção 1.3, microfones compactos permitem descrever o campo sonoro ao seu redor em função de harmônicas esféricas, que são funções próprias (base ortonormal) da onda acústica quando representada em coordenadas esféricas [35].

Para determinar o peso de cada alto-falante do arranjo Gerzon assumiu que estes se comportavam como uma fonte de onda plana, aproveitando o fato de que a onda plana possui uma representação analítica em harmônicas esféricas, dependente apenas da posição do falante no arranjo. Bastava então ponderar os alto-falante tal que, para cada harmônica, a soma das influências de todos alto-falantes fosse igual ao formato-B, ou seja, ao valor da harmônica esférica gravada com o arranjo compacto de microfones [36].

Como as harmônicas esféricas são um conjunto de bases ortogonais, eu posso truncar meu resultado na ordem necessária, definida pelo menor número de microfones ou alto-falantes [37, 38]. Gerzon sugeriu que quando truncado em primeira ordem[5], esta técnica funcionaria baseada em princípios psicoacústicos, ou seja, seria capaz de gerar parâmetros interaurais de forma similar à técnica de panorama [3].

[5] A ordem 0 possui uma harmônica enquanto a ordem 1 possui três harmônicas distintas, resultando em um total de quatro harmônicas para um sistema de ordem 1.

No entanto, conforme se aumenta a quantidade de harmônicas usadas para descrever o campo sonoro, este sistema promove uma transição suave do paradigma de panorama para o paradigma de síntese de campo sonoro [38].

2.3.2 Síntese de campo

Enquanto técnicas baseadas em panorama tentar recriar parâmetros interaurais, as técnicas de síntese de campo buscam justamente recriar um campo sonoro com a maior fidelidade possível. Note que até agora todas estas técnicas são capazes apenas de variar a direção de chegada percebida para a fonte virtual, mas não a sua distância do ouvinte, já que todas elas assumem que as fontes se encontram em campo distante e atuam como geradores de ondas planas.

Já as técnicas de síntese de campo, como são capazes de controlar o campo sonoro e consequentemente a curvatura da frente de onda, permitem sintetizar (em parte do interior do arranjo) fontes pontuais a distâncias menores que a dos próprios alto-falantes [39].

No entanto, estas técnicas se baseiam na suposição de haver um infinito número de fontes no arranjo. No momento que voltamos à realidade e construímos tais sistemas com um número limitado de alto-falantes, problemas mundanos começam a aparecer, como uma área limitada onde o campo sonoro é fiel ao planejado (sweet spot) e o surgimento de artefatos na reprodução em altas frequências (dependente da distância entre as fontes) [40].

Higher Order Ambisonics Com já foi dito, Ambisonics permite uma cadeia completa de gravação, transmissão e reprodução *escalável* de som espacial, tal que conforme aumentamos a ordem máxima das harmônicas esféricas utilizadas realizamos uma transição entre panorama e síntese.

Isto significa dizer que o Ambisonics de alta ordem (*Higher Order Ambisonics*, HOA) nada mais é do que a técnica Ambisonics descrita na seção anterior, mas truncada em uma ordem de harmônicos esféricos mais elevada. E assim como sua versão de baixa ordem ela permite tanto a mixagem de sons (geração de fontes virtuais) quanto a gravação de uma performance com um arranjo concêntrico de microfones.

O sweet spot relacionado a esta técnica se torna maior conforme aumentamos a ordem máxima do sistema. É interessante notar que o sweet spot de sistemas Ambisonics se localizam sempre no centro do arranjo [41].

Wave Field Synthesis Vimos que o HOA é baseado em sinais captados com arranjos concêntricos, que captam como o som chega a um dado ponto e cuja reprodução busca fornecer sinais para os alto-falantes que, quando combinados, reproduzam no centro do arranjo o mesmo campo sonoro que foi captado pelo arranjo de microfones.

Já a técnica de síntese de campo sonoro (*wave field syntesis*, WFS) é baseada no princípio de Huygens, que postula que uma frente de onda pode ser decomposta na soma de infinitas fontes pontuais [4]. Ou seja, se eu conheço o campo sonoro que incide sobre o lado externo de um volume, eu posso reproduzir este mesmo campo na parte interna deste volume distribuindo infinitas fontes pontuais pela superfície limítrofe deste volume e fazendo que cada uma delas emita um sinal equivalente ao sinal gerado pelo campo sonoro externo naquele ponto infinitesimal.

Ou seja, diferentemente do HOA, que amostra o campo sonoro em uma região compacta do espaço, o WFS assume que o campo sonoro será amostrado em todo contorno do meu espaço de reprodução. Esta suposição não é viável na praticamente, pois precisaríamos de um arranjo distribuído de grandes proporções. Por esta razão, o WFS é usado somente com sinais auralizados, ou seja, para campos sonoros virtuais. Isto o torna um sistema praticamente "nativo" para reprodução de arquivos com compactação orientada a objetos.

Infelizmente, apesar de teoricamente o WFS permitir uma reprodução muito fiel do campo acústico simulado, isto só acontece de fato se o arranjo utilizado for suficientemente denso, o que o torna este sistema caro e ainda não muito viável para uso cotidiano.

3 CODIFICAÇÃO ORIENTADA A OBJETOS

Como vimos nas seções anteriores, existe uma vasta gama de sistemas de gravação e reprodução de áudio espacial e a maioria destes sistemas só são compatíveis com seus sistemas duais, ou seja, uma gravação com um arranjo compacto do tipo X-Y deve ser reproduzida com um arranjo distribuído tipo estéreo. Para poder ser reproduzido com outro tipo de sistema, como VBAP ou Ambisonics, seria necessário um processamento conhecido como "up-mixing", que busca extrair as informações espaciais do arquivo original e sintetizar um novo arquivo compatível com um outro sistema de reprodução.

Como a tendência da reprodução de áudio doméstica é a oferta de soluções proprietárias baseada em um ou vários dos paradigmas discutidos neste artigo, é importante criar um tipo de arquivo que possa ser reproduzido em qualquer tipo de sistema. Isto significa que os arquivos devem fornecer todas as informações necessárias para que a cena acústica desejada possa ser sintetizada pelo próprio cliente, que deve fornecer informação sobre as características de seu sistema de reprodução.

E é justamente nesta direção que caminha o mais novo formato de compressão de áudio proposto pela MPEG, o formato SAOC, que propõe uma codificação orientada a objetos sonoros.

Este novo paradigma de codificação é completamente diferente de todos outros formatos de codificação já propostos pela MPEG, como o *MPEG Surround*, que realiza um "down-mix" de um sinal 5.1 para um sinal estereofônico, que é então comprimido para transmissão, além de gerar uma sequência de metadados espaciais com baixa taxa de informação (que é posteriormente usada para tentar reconstituir o arquivo multi-canal). O fato é que o MPEG Surround foi desenvolvido sob o paradigma de múltiplos canais, e não de objetos sonoros.

ISO/IEC 23003-2:2010 Esta norma padroniza as interfaces do SAOC, ditando os rumos que o desenvolvimento deve seguir e garantindo compatibilidade entre diferentes soluções.

O SAOC utiliza o conceito de *objetos sonoros*, isto é, sinais monofônicos não-correlacionados, que são então compactados conjunta ou separadamente. Além disto, uma sequência de metadados espaciais com baixa taxa de informação também é adicionada ao fluxo de dados. Estas informações são então utilizadas pelo receptor para gerar os sinais para cada canal de saída de acordo com o sistema de reprodução disponível informado pelo usuário.

Mas e se não tenho as gravações individuais de cada objeto da cena? É aqui que entram novos paradigmas de processamento de sinais, que não foram padronizados e ainda são objetos de pesquisa. Estes novos algoritmos serão baseados em modelos da audição espacial que possuirão como entrada gravações feitas com arranjos de microfones, sejam eles densos ou esparsos. Um exemplo nesta direção é o algoritmo DirAc, que extrai estes parâmetros de gravações feitas com microfones do tipo soundfield ou eigenmic [42, 43].

No lado de entrada do codec será necessário desenvolver algoritmos que permitam a separação cega de fontes e extração de informações espaciais da cena sonora. Do lado de saída serão necessário desenvolver sintetizadores robustos capazes de fornecer uma reprodução de qualidade para qualquer tipo de geometria de fontes.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Randall Stross, "The Incredible Talking Machine," *TIME*, 2010.

[2] Alan Dower Blumlein, "Improvements in and relating to sound-transmission, sound-recording and sound-reproducing systems," 1931.

[3] Michael Gerzon, "Surround-sound psychoacoustics: Criteria for the design of matrix and discrete surround-sound systems," *Wireless World*, vol. 80, no. December, pp. 483–486, 1974.

[4] A. J. Berkhout, "A Holographic Approach to Acoustic Control," *Journal of the Audio Engineering Society*, vol. 36, no. 12, pp. 977–995, 1988.

[5] Henrik Møller, "Fundamentals of binaural technology," *Applied Acoustics*, vol. 36, no. 3-4, pp. 171–218, 1992.

[6] Anton Huurdeman, *The Worldwide History of Telecommunications*, John Wiley & Sons, 2003.

[7] Philip A. Nelson and Stephen J Elliott, *Active Control of Sound*, Academic Press, San Diego, CA, 3rd edition, 1992.

[8] Jens Blauert, *Spatial hearing: the psychophysics of human sound localization*, MIT Press, 1997.

[9] Henrik Møller, Michael Friis Sørensen, Dorte Hammershøi, and Clemen Boje Jensen, "Head-Related Transfer Functions of Human Subjects," *J. Audio Eng. Soc.*, vol. 43, no. 5, pp. 300–310, 1995.

[10] William M. Hartmann, "How We Localize Sound," *Physics Today*, vol. 52, no. 11, pp. 24, nov 1999.

[11] Stephan Paul, "Binaural Recording Technology: A Historical Review and Possible Future Developments," *Acta Acustica united with Acustica*, vol. 95, no. 5, pp. 767–788, sep 2009.

[12] Ville Pulkki and Matti Karjalainen, *Communication Acoustics*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2015.

[13] Bill Whiston, "Microphone systems used for Surround Sound pickup - And their use at Wimbledon tennis and the Proms," *EBU Technical Review*, , no. Q1, pp. 1–9, 2008.

[14] Vítor Heloiz Nascimento, Bruno Sanches Masiero, and Flávio P. Ribeiro, "Acoustic Imaging Using the Kronecker Array Transform," in *Signals and Images: Advances and Results in Speech, Estimation, Compression, Recognition, Filtering, and Processing*, Rosangela Fernandes Coelho, Vitor Heloiz Nascimento, Ricardo Lopes de Queiroz, Joao Marcos Travassos Romano, and Charles Casimiro Cavalcante, Eds., pp. 153–178. CRC Press, 2015.

[15] Lawrence C. Wood and Sven Treitel, "Seismic Signal Processing," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 649–661, 1975.

[16] Jerald L Bauck and Duane H Cooper, "Generalized Transaural Stereo," in *93rd AES Convention*, San Francisco, USA, 1992.

[17] Bruno Sanches Masiero and Janina Fels, "Perceptually Robust Headphone Equalization for Binaural Reproduction," in *130th AES Convention*, London, England, 2011, pp. 1–7.

[18] Josefa Oberem, Bruno Sanches Masiero, and Janina Fels, "Authenticity and naturalness of binaural reproduction via headphones regarding different equalization methods," *Proceedings of AIA-DAGA 2013 Conference on Acoustics*, 2013.

[19] Michael Vorländer, *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality (RWTHedition)*, Springer, 2007.

[20] Tobias Lentz, "Dynamic crosstalk cancellation for binaural synthesis in virtual reality environments," *J. Audio Eng. Soc.*, vol. 54, no. 4, pp. 283–294, 2006.

[21] B. B. Bauer, "Stereophonic Earphones and Binaural Loudspeakers," *J. Audio Eng. Soc.*, vol. 9, no. 2, pp. 148–151, 1961.

[22] Jerald L Bauck and Duane H Cooper, "On Transaural Stereo for Auralization," in *95th AES Convention*, New York, 1993, vol. 3728.

[23] Henrik Møller, "Reproduction of artificial-head recordings through loudspeakers," *J. Audio Eng. Soc.*, vol. 37, no. 1, pp. 2–5, 1989.

[24] Michael A Akeroyd, John Chambers, David Bullock, Alan R Palmer, A Quentin Summerfield, Philip A. Nelson, and Stuart Gatehouse, "The binaural performance of a cross-talk cancellation system with matched or mismatched setup and playback acoustics," *J. Acoust. Soc. Am.*, vol. 121, no. 2, pp. 1056—-1069, 2007.

[25] Piotr Majdak, Bruno Sanches Masiero, and Janina Fels, "Sound localization in individualized and non-individualized crosstalk cancellation systems.," *Journal of the Acoustical Society of America*, vol. 133, no. 4, pp. 2055–2068, apr 2013.

[26] Bruno Sanches Masiero and Michael Vorländer, "A Framework for the Calculation of Dynamic Crosstalk Cancellation Filters," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 9, pp. 1345–1354, sep 2014.

[27] Martin Pollow, Gottfried K Behler, and Bruno Sanches Masiero, "Measuring directivities of natural sound sources with a spherical microphone array," in *Proceedings of Ambisonics Symposium*, Graz, AT, jun 2009, IEM Graz.

[28] Alexander Mattioli Pasqual, *Sound Directivity Control in a 3-D Space by a Compact Spherical Loudspeaker Array*, Phd, Universidade Estadual de Campinas, 2010.

[29] Franz Zotter, *Analysis and Synthesis of Sound-Radiation with Spherical Arrays*, Ph.d., University of Music and Performing Arts, Graz, Austria, jan 2009.

[30] Internacional Standard, "ISO 3382:1997 - Acoustics - Measurement of the reverberation time of rooms with reference to other acoustical parameters," 1997.

[31] Klaus Wendt, *Das Richtungshören bei der Überlagerung zweier Schallfelder bei Intensitäts- und Laufzeitstereophonie*, Phd, RWTH Aachen, 1963.

[32] JC Bennett, K Barker, and FO Edeko, "A new approach to the assessment of stereophonic sound system performance," *J. Audio Eng. Soc*, vol. 33, no. 5, pp. 314–321, 1985.

[33] Ville Pulkki, *SPATIAL SOUND GENERATION AND PERCEPTION BY AMPLITUDE PANNING TECHNIQUES*, Ph.D. thesis, Helsinki University of Technology, 2001.

[34] Ville Pulkki, "Localization of amplitude-panned virtual sources II: Two-and three-dimensional panning," *J. Audio Eng. Soc.*, vol. 49, no. 9, pp. 753–767, 2001.

[35] Angelo Farina, Ralph Glasgal, Enrico Armelloni, and Anders Torger, "Ambiophonic Principles for the Recording and Reproduction of Surround Sound for Music," in *19th AES Conference on Surround Sound, Techniques, Technology and Perception*, Schloss Elmau, Germany, 2001.

[36] Bruno Sanches Masiero and Michael Vorländer, "SPATIAL AUDIO REPRODUCTION METHODS FOR VIRTUAL REALITY," in *Anais del 42o Congreso Español de Acústica*, Cáceres, Spain, 2011.

[37] Michael A. Gerzon, "Ambisonics in multichannel broadcasting and video," *J. Audio Eng. Soc.*, vol. 33, no. 11, pp. 859–871, 1985.

[38] Jérôme Daniel, Rozenn Nicol, and Sébastien Moreau, "Further investigations of high order ambisonics and wavefield synthesis for holophonic sound imaging," in *114th AES Conference*, Amsterdam, Netherlands, 2003, pp. 2764–2778.

[39] Filippo Maria Fazi, *Sound Field Reproduction*, Ph.D. thesis, Uuiversity of Southampton, 2010.

[40] M Kolundzija, C Faller, and Martin Vetterli, "Sound Field Reconstruction: An Improved Approach For Wave Field Synthesis," in *Preprint 126th Conv. Aud. Eng. Soc*, Munich, Germany, 2009.

[41] Sébastien Moreau, Jérôme Daniel, and Stéphanie Bertet, "3D sound field recording with higher order ambisonics-objective measurements and validation of spherical microphone," in *120th AES Convention*, 2006, vol. 5, pp. 1–24.

[42] Juha Vilkamo, Tapio Lokki, and Ville Pulkki, "Directional audio coding: Virtual microphone-based synthesis and subjective evaluation," *AES: Journal of the Audio Engineering Society*, vol. 57, no. 9, pp. 709–724, 2009.

[43] Jürgen Herre, Cornelia Falch, Dirk Mahane, Giovanni Del Galdo, Markus Kallinger, and Oliver Thiergart, "Interactive teleconferencing combining spatial audio object coding and DirAC technology," *Journal of the Audio Engineering Society*, vol. 59, no. 12, pp. 924–935, 2012.

Sociedade de Engenharia de Áudio

Palestra em Congresso

Ministrada no 14º Congresso de Engenharia de Áudio
20ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2016, São Paulo, SP

Este artigo foi reproduzido do original final entregue pelo autor, sem edições, correções ou considerações feitas pelo comitê técnico. A AES Brasil não se responsabiliza pelo conteúdo. Outros artigos podem ser adquiridos através da Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA, www.aes.org. Informações sobre a seção Brasileira podem ser obtidas em www.aesbrasil.org. Todos os direitos são reservados. Não é permitida a reprodução total ou parcial deste artigo sem autorização expressa da AES Brasil.

Análise modal e representação musical aplicadas ao design de estúdios de produção fonográfica

José Augusto Mannis[1 2], Tiago F. Mattos[3]

[1]Universidade Estadual de Campinas – UNICAMP
Instituto de Artes – Depto. de Música
Laboratório de Acústica e Artes Sonoras – LASom
Campinas, SP, 13083-854, Brasil

[2]Universidade Federal Fluminense – UFF
Programa Nacional de Pós-Doutorado PNPD CAPES
Instituto de Arte e Comunicação Social
Laboratório de Produção Musical – GATP0001
Niterói, RJ, 24210-510, Brasil

[3]Estúdio Codimuc
Taubaté, SP, 12000-000, Brasil

jamannis@unicamp.br tiagomattos@codimuc.com.br

RESUMO

Este trabalho relata e desenvolve uma análise acústica realizada em 2015, apresentando a fundamentação teórica e a metodologia adotadas, discutindo resultados obtidos e apontando para novas perspectivas de pesquisa. Como métodos inovadores destacam-se a análise do comportamento modal em acústica de salas pelo *critério de Bonello* após resultados de simulação por elementos finitos; a inclusão de representação musical para dar mais visibilidade a imagens mentais próprias à escuta humana. A inclusão de medições realizadas nas duas salas simuladas, depois de construídas e devidamente instaladas de mobiliário e equipamentos, permitiu estender a análise a outros aspectos de interesse como a disposição dos monitores de áudio.

0. INTRODUÇÃO

Os comprimentos de onda das baixas frequências que caracterizam as salas são comparáveis às suas próprias dimensões. Isso significa que temos muito mais frentes de onda que se deslocam, do que raios acústicos. Segundo Gerges, a teoria de geometria ou raio acústico é inadequada para estudar o comportamento acústico de salas. A mais adequada é a teoria da análise modal, na qual a sala pode ser tratada como um ressonador complexo comportando vários modos acústicos, cada um com uma frequência característica de ressonância e amortecimento [1]. Os modos de vibração em acústica de salas estão especificamente vinculados às ondas estacionárias que se estabelecem em seu volume interior única e exclusivamente por propagação aérea. Existem três tipos de modos característicos de uma sala retangular: (1) *Modos Axiais*, referentes a ondas

estacionarias resultantes da propagação na direção de uma única linha, ou melhor, reta, portanto entre superfícies paralelas. Se a sala for perfeitamente retangular e a origem estiver num de seus vértices, os modos axiais se alinharão aos eixos (*x, y* e *z*); (2) *Modos Tangenciais*, referentes a ondas estacionárias que se estabelecem pela propagação sonora em duas direções, seguindo não mais uma linha, mas agora um plano definido por duas retas, como por exemplo os planos *xy*, *xz* e *yz*; (3) *Modos Oblíquos* nos quais as ondas estacionárias resultam de propagações nas três direções ao mesmo tempo.

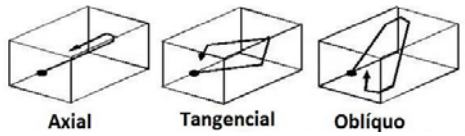


Figura 1 - Modos de vibração de uma sala produtos do percurso de propagação de ondas estacionárias em relação à geometria da sala: (1) Axial – em uma única direção, portanto em uma reta (eixo da sala); (2) Tangencial – em um plano: propagação paralela à superfícies de seu contorno; (3) Oblíquo – propagação em todas as direções [2].

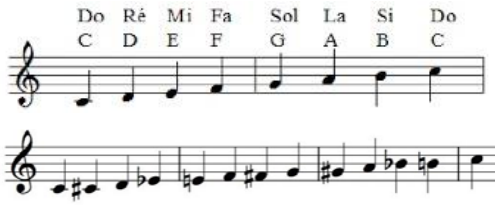


Figura 2 - Alturas musicais (C, D, E, F, G A, B – do, ré, mi, fá, sol, lá, si) *com as respectivas alterações: sustenido* (♯) *um semitom acima; bemol* (♭) *um semitom abaixo; bequadro* (♮) *cancela cada um dos anteriores.*

As alturas musicais inscritas em pentagrama nas representações empregadas neste trabalho, não se referem a nenhum sinal gerado por fonte sonora, mas apenas a potências de ressonâncias que somente ocorrerão mediante a incidência de ondas sonoras com frequências correspondentes.

Cabe aqui ainda explicitar a distinção a ser feita nas áreas de conhecimento de acústica, psicologia e afins entre os conceitos de *frequência*, grandeza atribuída a uma característica da onda sonora enquanto fenômeno físico, e *altura*, sensação produto de um fenômeno da percepção. Cada um desses conceitos remete, portanto, a uma instância distinta. O primeiro refere-se a um fenômeno físico externo que se pretende observar, enquanto que o segundo, a um fenômeno subjetivo, na instância do observador, como uma sensação que se apresenta à sua consciência. Portanto atenção para que a associação de alturas aos modos de vibração de uma sala não induza ninguém a ter a equivocada ideia de que uma sala possa gerar sons de alguma maneira. Salas por si não geram sons, elas simplesmente soam e ressoam. Os modos de vibração de uma sala farão com que determinadas ondas sonoras permaneçam ressoando por mais tempo do que outras, simplesmente por que sua conformação geométrica favorece em alguns casos encontros de máximos e mínimos de pressão que se somam, e em outros cancelamentos entre máximos e mínimos que tendem a ocorrerem no mesmo lugar. Por essa simples razão é que o principal elemento de design acústico de uma sala é antes de tudo a cuidadosa definição de suas proporções, ou seja, seu volume em um contorno geral delineado, harmonizando altura, largura e profundidade.

0.1 Nomenclatura dos modos

Os modos obtidos na simulação de cada sala estão ordenados de *M001* a *M200.* Os modos axiais entre teto e piso levam a letra *H* acompanhada de um número *n* referente à sua ordem: *H1* (primeiro Modo Axial entre teto e piso).

1. METODOLOGIA

A análise modal deste estudo não foi realizada a partir de predição dos modos inferida sobre uma superfície regular com faces paralelas (SRP), um paralelepípedo, com dimensões totais médias de largura (L_x), profundidade (L_y) e altura (L_z), mas através de um aplicativo de elementos finitos simulando o comportamento dos modos de vibração de um volume de ar contido no contorno de uma superfície irregular (SIR) com admitância nula, ou seja, considerando todas as superfícies como totalmente reflexivas. O método a partir das dimensões medidas consiste por exemplo em obter um contorno equivalente como na figura a seguir, compensando as irregularidades da superfície de contorno equilibrando os pontos além e aquém da linha média, mantendo a área e o volume constantes.

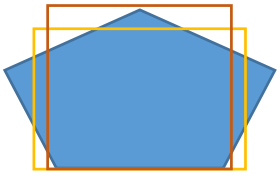


Figura 3 – Exemplo de duas tentativas de soluções de contornos médios, cada um reduzido a dois pares de linhas paralelas, buscando equilibrar as áreas que excedem e as que faltam em relação à figura inicial, de maneira a que se chegue a uma área equivalente.

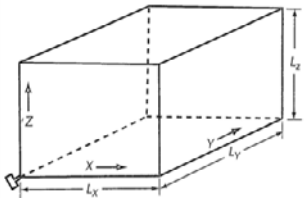


Figura 4 – Modelo de sala retangular (paralelepípedo) de referência para compreensão dos Modos Normais de Vibração de uma sala. Davis (2006, p. 180) [3]

Uma vez obtidas as três dimensões médias da sala L_x, L_y e L_z, os modos de vibração no volume de ar no interior do contorno médio é dado, como proposto por Davis [3], por

$$f_N = \frac{c}{2}\sqrt{\left(\frac{N_x}{L_x}\right)^2 + \left(\frac{N_y}{L_y}\right)^2 + \left(\frac{N_z}{L_z}\right)^2}$$

Equação 1

Sendo,

c, a velocidade de propagação do som

N_x, N_y e N_z são números inteiros a partir de 0, 1, 2, 3...

O diferencial significativo do presente método, baseado em simulação em aplicativo de elementos finitos, está em lidar diretamente com modos efetivos que se verificam na sala, e não com modos de um modelo aproximado como o que acabou de ser exposto logo acima.

2. PROJETOS ANALISADOS E SIMULAÇÃO

Realizamos a análise modal de dois projetos de salas de controle de áudio multicanal denominados *Sala de controle de áudio A* e *Sala de controle de áudio B*, com as seguintes características:

Tabela 1 – Coordenadas (em metros) dos vértices localizados na estrutura de alvenaria de cada sala (piso, paredes e laje) da Sala de controle de áudio A

	x	y	z
P01	0.986	0	0
P02	5.709	0	0
P03	6.696	2.413	0
P04	4.860	6.700	0
P05	1.836	6.700	0
P06	0	2.413	0
P07	0.986	0	3.560
P08	5.709	0	3.560
P09	6.696	2.413	3.560
P10	4.860	6.700	3.560
P11	1.836	6.700	3.560
P12	0	2.413	3.560

Área: 34,6 m^2
Perímetro: 22,3 m
Área total: 148,6 m^2
Volume: 123,2 m^3

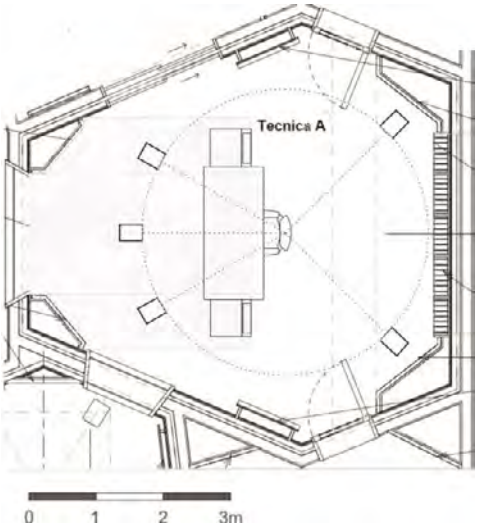


Figura 5 - Planta baixa da Sala A.

Tabela 2 - Coordenadas (em metros) dos vértices localizados na estrutura de alvenaria de cada sala (piso, paredes e laje) da Sala de controle de áudio B

	x	y	z
P01	0	0	0
P02	6.540	0	0
P03	6.540	1.632	0
P04	6.041	4.625	0
P05	0.500	4.625	0
P06	0	1.632	0
P07	0	0	2.950
P08	6.540	0	2.950
P09	6.540	1.632	2.950
P10	6.041	4.625	2.950
P11	0.500	4.625	2.950
P12	0	1.632	2.950

Área: 28,7 m^2
Perímetro: 21,4 m
Área total: 120,5 m^2
Volume: 84,7 m^3

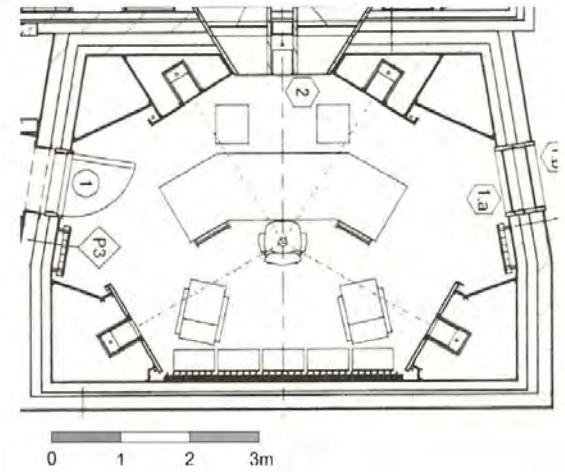


Figura 6 – Planta baixa da Sala B.

Tabela 3 - Parâmetros aplicados para a análise modal - simulação em aplicativo de elementos finitos

Pressão de referência	2.10^{-5} Pa
Densidade do ar	1,2 kg/m^3
Temperatura do ar	21ºC
Velocidade de propagação do som no ar	343,78 m/s
Admitância da superfície de contorno	0 (zero)
Elemento da malha de elementos finitos	22 cm
Frequência máxima a ser considerada ($\lambda/6$ = 22cm)	262 Hz
Quantidade de modos extraídos	300
Frequência Inicial	10 Hz
Frequência Final	3000 Hz

3. RESULTADOS, ANÁLISE E DISCUSSÃO

Os dados resultantes foram (1) uma listagem com os Modos ordenados de 1 a 300 e suas respectivas frequências; (2) plotagem em perspectiva compreendo três dimensões da conformação da variação de pressão no volume de cada sala para cada um dos seus modos. Foram desconsiderados os modos com frequência acima da *frequência máxima*, sendo esta delimitada pelo elemento empregado para compor a malha estrutural (22 cm) que não deve exceder 1/6 do comprimento de onda da onda estacionária simulada. Finalmente foram considerados os primeiros 270 modos da *Sala A* e 196 modos da *Sala B*.

Tabela 4 - Frequências dos 200 primeiros Modos de Vibração da Sala A obtidas na simulação

SET	FREQ (Hz)		FREQ (Hz)		FREQ (Hz)
1	27,71	68	153,72	135	199,85
2	31,27	69	153,90	136	200,09
3	48,35	70	155,04	137	200,30
4	49,40	71	155,07	138	200,45
5	50,74	72	155,11	139	200,71
6	55,77	73	155,47	140	202,54
7	57,63	74	155,52	141	203,22
8	60,81	75	156,74	142	203,51
9	64,34	76	157,49	143	203,63
10	69,23	77	157,94	144	203,71
11	70,20	78	159,36	145	204,12
12	76,12	79	160,72	146	204,16
13	77,84	80	161,47	147	204,96
14	79,44	81	163,14	148	205,54
15	80,62	82	164,42	149	207,33
16	81,82	83	164,76	150	207,79
17	90,36	84	165,10	151	208,17
18	92,86	85	165,82	152	208,18
19	92,97	86	166,04	153	208,28
20	93,19	87	166,28	154	209,32
21	95,22	88	167,71	155	209,49
22	97,06	89	168,81	156	209,62
23	101,00	90	169,95	157	209,67
24	101,04	91	170,45	158	212,27
25	102,09	92	170,97	159	212,54
26	104,29	93	171,31	160	213,28
27	104,91	94	173,11	161	214,00
28	105,02	95	173,25	162	214,10
29	109,16	96	174,59	163	214,79
30	109,81	97	174,83	164	215,12
31	110,17	98	174,96	165	215,20
32	112,22	99	175,82	166	215,21
33	112,35	100	176,56	167	215,52
34	114,94	101	176,70	168	216,47
35	115,20	102	177,00	169	216,52
36	116,85	103	178,05	170	217,18
37	119,53	104	179,43	171	217,68
38	119,59	105	180,25	172	217,75
39	120,58	106	181,37	173	218,16
40	122,57	107	182,16	174	218,56
41	123,23	108	182,73	175	219,25
42	123,89	109	183,19	176	219,84
43	126,01	110	183,53	177	220,00
44	127,53	111	184,71	178	220,95
45	129,26	112	185,12	179	221,24
46	129,28	113	185,91	180	221,34
47	131,62	114	186,40	181	222,21
48	132,66	115	186,51	182	222,88
49	133,56	116	186,70	183	223,60
50	135,06	117	187,85	184	223,72
51	135,18	118	188,84	185	224,69
52	138,56	119	191,16	186	225,24
53	139,55	120	191,29	187	226,93
54	140,54	121	192,95	188	227,54
55	140,90	122	193,25	189	227,56
56	142,34	123	193,46	190	227,66
57	143,32	124	193,75	191	227,89
58	143,52	125	193,80	192	228,22
59	145,75	126	194,32	193	228,47
60	146,57	127	194,42	194	228,73
61	147,07	128	197,18	195	229,43
62	147,76	129	197,27	196	230,31
63	148,01	130	198,38	197	230,94
64	149,30	131	199,22	198	231,64
65	149,42	132	199,30	199	232,52
66	150,05	133	199,34	200	232,74
67	151,80	134	199,38		

Tabela 5 - Frequências dos 200 primeiros Modos de Vibração da Sala B obtidas na simulação.

SET	FREQ (Hz)		FREQ (Hz)		FREQ (Hz)
1	27,55	68	172,92	135	226,40
2	37,47	69	173,03	136	226,61
3	47,31	70	173,09	137	227,28
4	54,85	71	173,56	138	227,57
5	58,38	72	177,29	139	228,77
6	64,60	73	177,83	140	229,00
7	68,31	74	180,14	141	229,79
8	69,46	75	180,70	142	230,84
9	74,86	76	181,80	143	231,16
10	75,27	77	182,01	144	232,82
11	79,52	78	182,30	145	232,97
12	80,27	79	182,66	146	233,10
13	81,81	80	184,55	147	236,01
14	90,07	81	184,75	148	236,61
15	93,37	82	186,70	149	236,91
16	94,41	83	187,28	150	237,01
17	95,20	84	187,78	151	238,08
18	98,92	85	189,35	152	238,53
19	100,77	86	190,01	153	238,71
20	108,37	87	190,36	154	238,96
21	110,43	88	191,39	155	239,00
22	111,33	89	191,80	156	240,02
23	111,81	90	191,92	157	240,25
24	112,22	91	192,07	158	240,60
25	116,76	92	192,95	159	240,73
26	117,43	93	194,16	160	240,83
27	120,73	94	195,19	161	241,79
28	121,64	95	196,02	162	242,18
29	123,45	96	197,00	163	243,42
30	123,46	97	197,22	164	244,33
31	125,09	98	198,12	165	244,60
32	126,50	99	198,71	166	244,63
33	126,92	100	198,77	167	244,73
34	126,92	101	199,41	168	245,27
35	130,00	102	199,70	169	245,69
36	130,95	103	201,05	170	245,73
37	135,17	104	202,37	171	245,82
38	135,35	105	202,40	172	246,57
39	136,41	106	202,51	173	247,16
40	136,62	107	202,93	174	247,71
41	138,46	108	203,03	175	250,29
42	139,97	109	204,82	176	250,87
43	140,27	110	205,91	177	250,99
44	142,58	111	206,92	178	252,49
45	143,90	112	207,07	179	253,13
46	147,70	113	210,20	180	253,28
47	149,01	114	210,85	181	253,36
48	149,53	115	211,06	182	254,05
49	150,42	116	212,21	183	254,53
50	150,97	117	212,72	184	254,95
51	151,70	118	213,63	185	255,12
52	152,39	119	214,59	186	255,27
53	153,29	120	215,03	187	255,47
54	157,81	121	215,23	188	255,89
55	160,92	122	215,55	189	259,47
56	161,02	123	217,85	190	260,02
57	161,93	124	218,00	191	260,04
58	162,26	125	218,31	192	260,45
59	162,32	126	219,28	193	260,55
60	162,90	127	219,33	194	261,39
61	163,42	128	220,23	195	261,69
62	163,82	129	220,79	196	261,78
63	164,54	130	224,07	197	262,29
64	167,01	131	225,40	198	262,31
65	168,79	132	225,86	199	262,73
66	170,51	133	226,21	200	263,77
67	172,49	134	226,26		

Everest [4] (p. 58-59) [5] (p. 343-344) cita os critérios desenvolvidos por Oscar Bonello [6] [7] [8], *Bonello Criteria*, para avaliação da qualidade acústica de uma sala no tocante à seus modos de vibração. Bonello prescreve a observação e análise da configuração e do comportamento dos modos de vibração de uma sala. O *Critério 1* verifica como os modos estão distribuídos, mais particularmente quanto à densidade modal em função da frequência. Para verificar a progressão da acumulação de modos, monta-se um diagrama contendo o número de frequências modais por banda de frequência de um terço de oitava e observa-se como o gráfico evolui. Se for progressivamente crescente, de forma regular, as frequências modais estarão com uma distribuição aceitável. Entretanto, se a linha for irregular, poderá haver problemas de coloração na resposta acústica da sala.

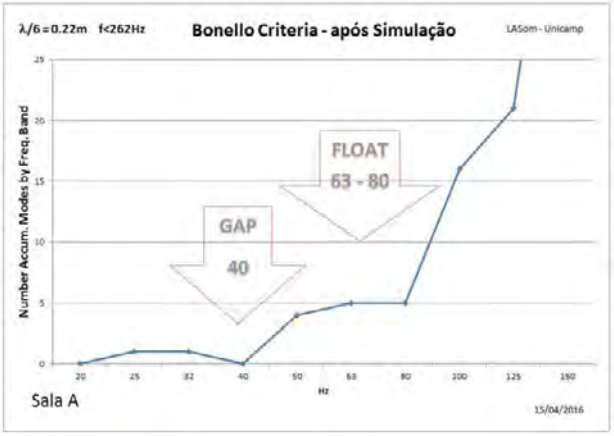


Figura 7 – Análise pelo critério de Bonello das frequências dos Modos de Vibração da Sala A obtidas na simulação.

Na análise através do *Critério 1* de Bonello constata-se que a *Sala A* apresenta uma lacuna na densidade modal na banda de 40 Hz e uma forte contenção de progressão de densidade modal nas bandas de 63 e 80 Hz.

Na análise da *Sala B* pelo critério de Bonello *(Figura 8)* temos uma progressão com uma falsa lacuna [FAUX GAP] em 32 Hz, portanto uma progressão contínua, havendo uma leve contenção na progressão da densidade modal na banda 80 Hz após a qual, de 90 a 180 Hz, ocorre uma retomada na progressão. A aparente ausência de modo na banda de 32 Hz corresponde a uma falsa lacuna, pois os modos iniciais estão em constante e regular progressão, tendo ocorrido de ter dois dos modos posicionados, um muito próximo, mas antecedendo o início da banda de 32 Hz e o seguinte, sucedendo imediatamente o final da mesma banda, conforme se observa na *Figura 10.*

Nesta análise de comportamento da progressão modal, empregaremos, dentre outros procedimentos, um método apresentado por Renato Cipriano na AES de 2010, infelizmente em palestra sem publicação, mas cujo processo consiste em distribuir os modos em um diagrama logarítmico de maneira a ter uma impressão mais próxima da sensação auditiva quanto à distribuição dos modos em função das alturas percebidas pelo ouvido humano, como apresentado em *Figura 9* e *Figura 10*.

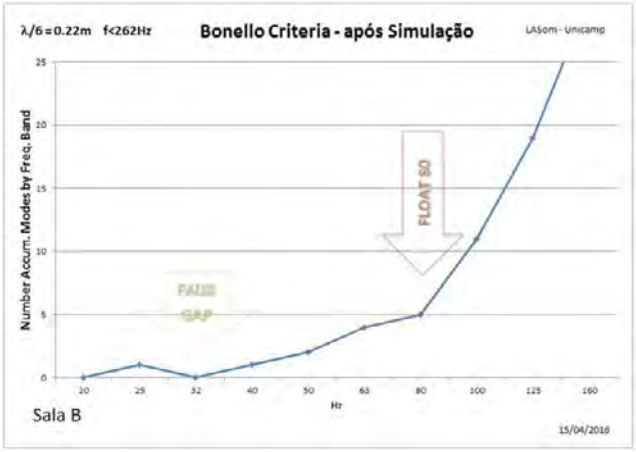


Figura 8 - Análise pelo critério de Bonello das frequências dos Modos de Vibração da Sala B obtidas na simulação.

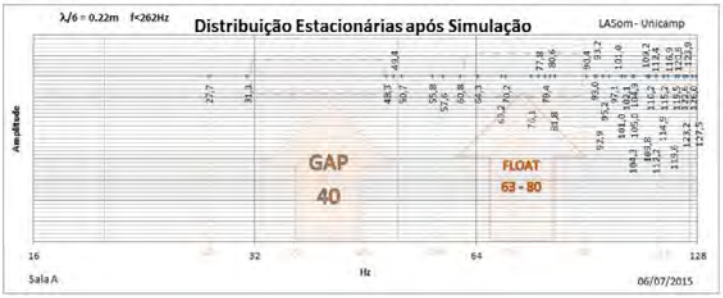


Figura 9 - Distribuição das frequências dos Modos de Vibração da Sala A (16 – 128 Hz) obtidas na simulação, plotadas em eixo logarítmico.

A *Figura 9* explicita o vazio que se forma na região de 40 Hz [GAP 40] da *Figura 7* bem como a progressão fortemente contida da acumulação de modos nas bandas de 63 e 80 Hz [FLOAT 63-80], aglutinando vários modos muito próximos um do outro em pequenos grupos, cercados por segmentos vazios, proporcionalmente iguais ou maiores que o próprio agrupamento.

Observando na *Figura 10* como o ritmo dos modos representados pelos pontos que se acumulam progride regularmente, pode-se concluir que a lacuna em 32 Hz no gráfico da *Figura 8* é unicamente produto de um efeito de fase entre o período das bandas e do passo de progressão modal, não reportando a uma irregularidade de progressão. Por efeitos como esse é sempre recomendável checar as possibilidades de predição de qualidade com diversos métodos. Quando todos apontam para a mesma tendência é uma confirmação. Quando há contradições, aprofunda-se a verificação.

Discrepâncias como essa em relação à análise pelo critério de Bonello ou ainda efeitos contrários ao expresso no gráfico de progressão de densidade modal encontrados na posição do *sweetspot* podem ser

evidenciados somente após o exame das plotagens de pressão para cada modo resultantes da análise modal com simulação. Evidências sobre essas discrepâncias encontram-se no *item 4,* mais à frente.

Figura 10 - Distribuição das frequências dos Modos de Vibração da Sala B (16 – 128 Hz) obtidas na simulação.

A *Sala B,* portanto, começa a progressão modal de maneira regular e suave, contrastando neste aspecto com o GAP 40 Hz, o FLOAT 63-80 e a retomada da progressão nas bandas de 90 a 180 Hz da *Sala A,* plotada na *Figura 11*, apresentando truncamentos irregulares em pequenos segmentos. Os agrupamentos inconstantes tendem a reforçar a intensidade nas frequências dos modos reunidos em detrimento de uma menor intensidade nas bandas de frequências vizinhas, proporcionalmente descobertas, sem a presença de modos. Razão pela qual a isso denominamos *flutuação*.

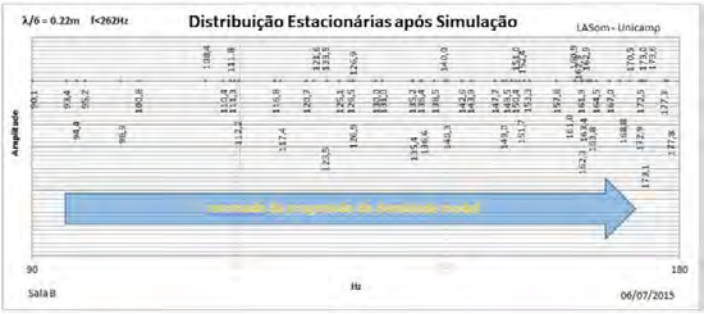


Figura 11 - Distribuição das frequências dos Modos de Vibração da Sala B (90 – 180 Hz) obtidas na simulação.

Buscando aproximar ainda mais a apresentação dos dados de maneira semelhante à sensação auditiva, as frequências dos modos foram convertidas em alturas musicais com precisão de 1/8 de tom[1], através da comparação da frequência de cada modo com uma tabela de frequências com precisão de 1/16 de tom, sendo cada altura musical de referência centrada e compreendida num âmbito com intervalo de ±1/16 de tom.

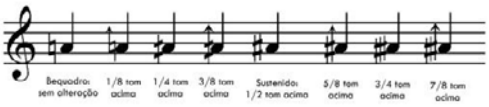


Figura 12 – Legenda da notação de alturas musicais em intervalos de 1/8 de tom. A nota de referência no pentagrama é um la 3 (A4) 440 Hz.

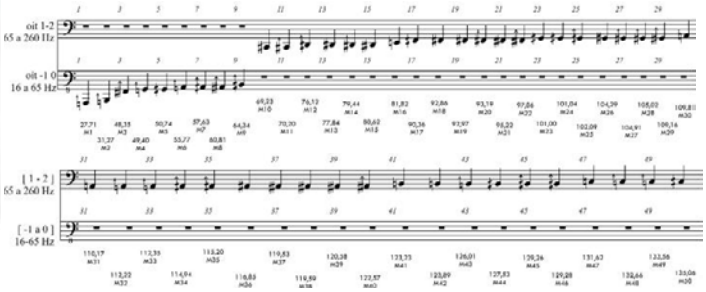


Figura 13 – Representação em notação musical dos 50 primeiros modos da Sala A. Abaixo de cada nota musical, a indicação da ordem do modo e sua respectiva frequência.

Contudo, apesar de nesta última representação ter havido um sensível e significativo ganho de precisão em alturas, a percepção das irregularidades de distribuição das frequências ficou menos clara do que estava na plotagem sobre eixo de frequência logarítmico (*Figura 9, Figura 10, Figura 11*). Com o intuito de torna-la novamente perceptível, atribuímos à posição das notas musicais sequenciadas no pentagrama não mais a ordem do modo *(1, 2, 3...)* cuja frequência lhe corresponde, mas uma progressão constante de alturas em intervalos de 1/8 de tom. Cada posição à direita de cada nota, está reservada para uma nota 1/8 de tom acima, de maneira que percorrendo horizontalmente o pentagrama temos as posições para todas as alturas ordenadas em sequência de 1/8 de tom. Porém, só estarão inscritas as notas cuja altura corresponder à frequência de um determinado modo encontrado na sala, caso contrário o espaço permanecerá vazio. Assim, nas posições em que houver uma nota inscrita, terá como significado que há um modo com a frequência correspondendo à altura anotada. Quando não houver nenhum modo a ser associado, o espaço permanece vazio, ou seja, com a representação musical de pausa (𝄽). Os espaços vazios no pentagrama corresponderão à progressão de frequências cujas ondas sonoras não serão influenciadas por ondas estacionárias. Além disso, acrescentamos à cada nota, sempre que possível, a imagem da plotagem de pressão correspondente a seu modo de vibração, conforme obtido na simulação.

É importante salientar que todas as zonas de máximos e mínimos de pressão mostradas nas plotagens produto da simulação são potências de respostas da sala, podendo ser mais ou menos evidenciadas em função do

[1] O intervalo de 1/8 de tom é dado por uma razão entre frequências vizinhas, tal que $\frac{f_2}{f_1} = \left(1 + \sqrt[48]{2}\right)$, sendo $f_2 > f_1$

posicionamento dos monitores de áudio, como mostrado no *item 4*. Ou seja, todos os modos da sala aqui mapeados, podem ser excitados colocando fontes sonoras em várias posições, tendo uma resposta sonora em frequência diferente para cada posição. Porém é importante salientar que todas essas respostas serão derivadas da mesma estrutura modal da sala, representada pelo produto do mapeamento, cujo processo está aqui apresentado e proposto como metodologia de análise.

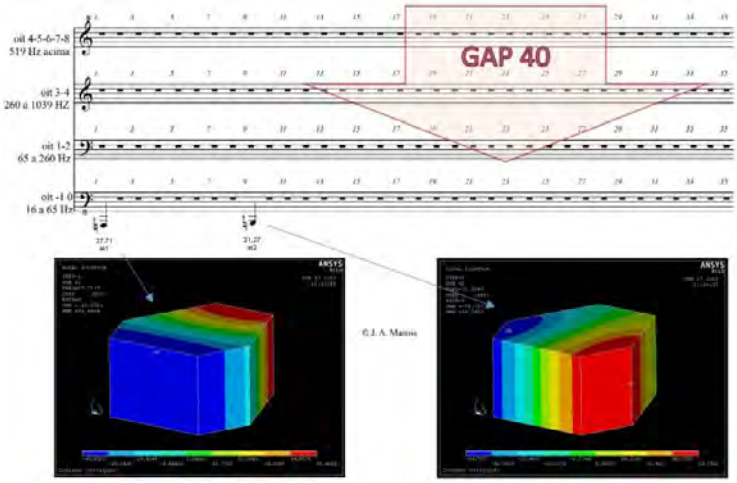


Figura 14 – Representação da progressão modal da Sala A em alturas musicais em 1/8 de tom (01 a 35 oitavos de tom) – Modos M01 e M02.

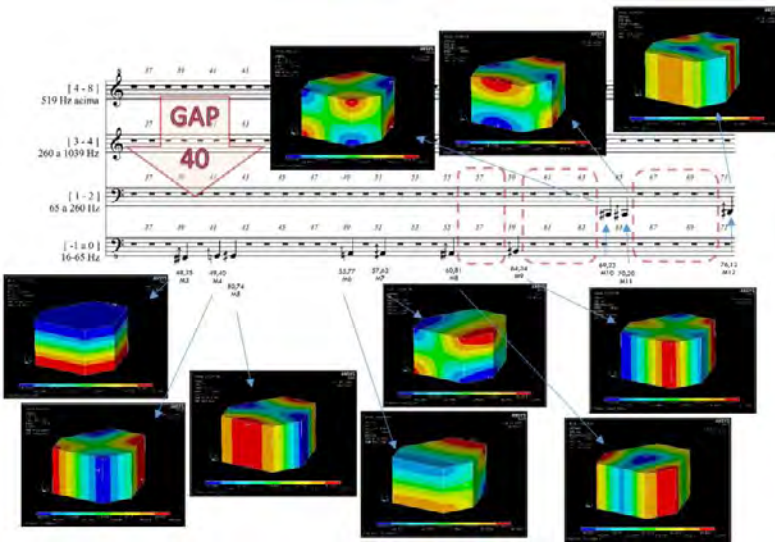


Figura 15 - Representação da progressão modal da Sala A em alturas musicais em 1/8 de tom (36 a 71 oitavos de tom) – Modos M03 a M12.

Julgamos relevante priorizar a análise e o tratamento acústico em relação ao *sweetspot* da sala de controle de áudio, pois esta é a posição de escuta mais relevante nesse tipo de sala. Após identificarmos todos os modos interferindo com máximos e mínimos de pressão na zona do *sweetspot,* os mesmos foram reunidos e distribuídos em representação de alturas musicais em sequência de passos de 1/8 de tom, evidenciando assim as incidências e os vazios de modos, numa representação de sensação linear para a percepção humana e bastante clara à interpretação e análise. Essa representação implementa um novo recurso permitindo antever a variação de resposta acústica da sala para cada nota (altura) musical e, com isso, tornar possível um controle estreito e preciso por parte da equipe de produção, integrando o conteúdo musical às técnicas de produção fonográfica. Esse esquema também representa um mapeamento da resposta acústica da sala de controle em função das alturas musicais. Se a tonalidade principal da música gravada coincidir com frequência de um modo, alguns efeitos, como por exemplo o de finalização em cadência podem ser potencializados. Se uma eventual influência desse tipo proporcionar uma impressão de maior plenitude e amplidão, isso estará falseando a mixagem, pois é um efeito acontecendo na sala e não na mixagem. Da *Sala A* separamos na *Figura 18* quatro modos que nos pareceram particularmente críticos: *M08*, *M14*, *M22* e *M60*. Todos atuam fortemente no *sweetspot* e na altura da cabeça humana, em posturas em pé ou sentado.

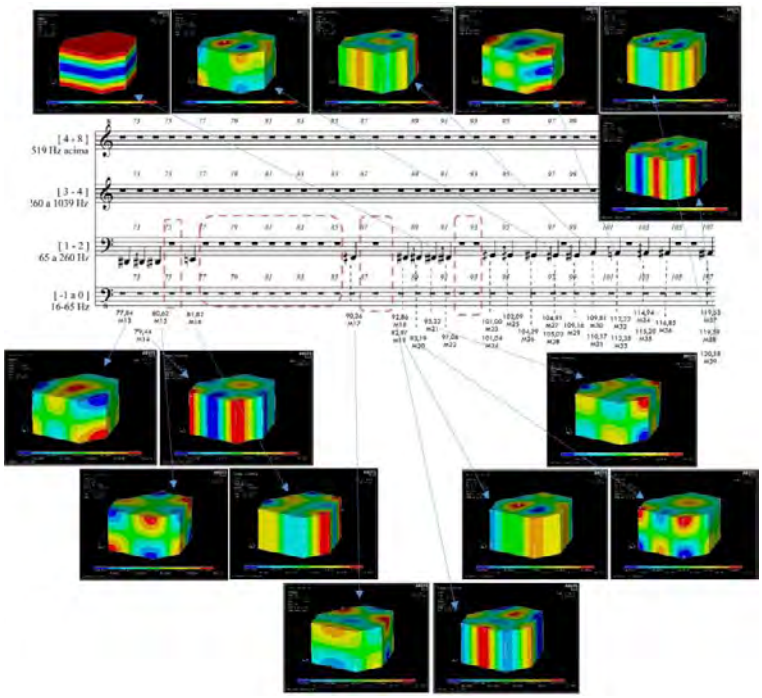


Figura 16 - Representação da progressão modal da Sala A em alturas musicais em 1/8 de tom (72 a 107 oitavos de tom) – Modos M13 a M39.

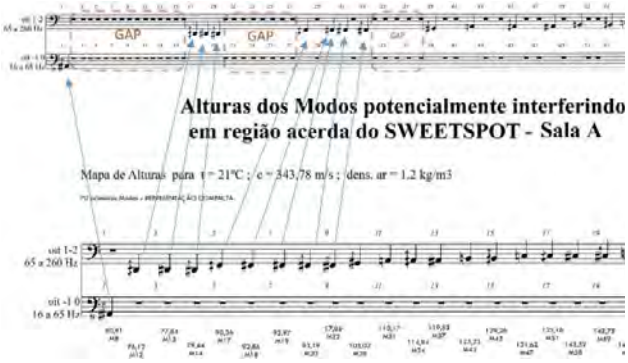
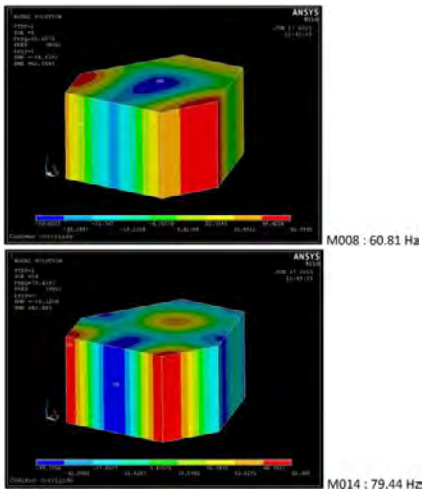


Figura 17 – Distribuição em passos de 1/8 de tom das alturas dos modos com impacto sobre o sweetspot da Sala A.

Na concepção de contorno de sala em forma de pentágono, como é o caso da *Sala A*, os modos axiais laterais sempre tenderão a atuar na direção do *sweetspot* como *M02*, *M08* e *M37*, neste caso reforçando frequências de modos associadas a alturas musicais *la_{-1} 1/8 de tom acima, si_{-1} 1/8 de tom acima,* alturas entre *sib_0 ¼ de tom abaixo e do_1 1/8 de tom abaixo, e entre la_1 e si_1 (Figura 13)*, além de envolver percursos lineares de propagação das estacionárias na composição de modos tangenciais próximos como *M12 (mi_1 ¼ de tom abaixo)*, *M13 (mib_1)*, *M18* e *M19 ($fá\#_1$) (Figura 14, Figura 15, Figura 16).* Essa característica de resposta é inerente à conformação do contorno e só pode ser alterada, mudando sua forma ou implementando *bass trap* com interferência em ¼ de comprimento de onda na direção lateral mais ampla. O dimensionamento do *bass trap,* compreendendo sua efetiva eficiência, pode ser verificado em simulação em aplicativo de elementos finitos. Um outro ponto que chama a atenção na *Sala A* é a grande concentração de modos acerca de si_0 e *sib_0 ¼ de tom abaixo e si_1* e *sib_1 ¼ de tom abaixo,* pelo fato de ter, na raiz dos modos, três frequências fundamentais próximas decorrentes das proporções entre as dimensões da sala: Profundidade *M1 (la_{-1} 1/8 de tom acima)*, Largura *M2 (si_{-1} 1/8 de tom acima)* e altura *M3 (sol_0 ¼ de tom abaixo ou $fa\#_0$ ¾ de tom acima)*.



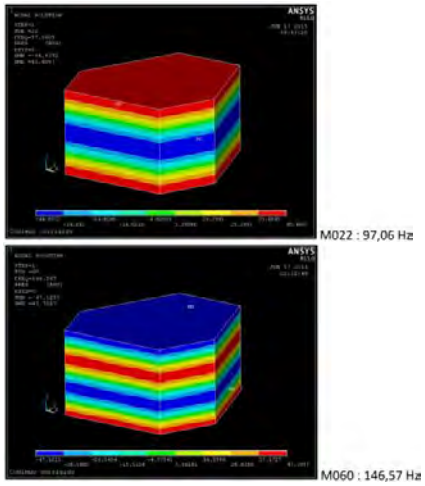


Figura 18 – Sala A - Impacto no sweetspot: (1) Modo de Vibração Axial L1: primeiro modo lateral. M008 60,81 Hz; (2) Modo de Vibração Tangencial: M014 79,44 Hz; (3) Modo de Vibração Axial H2: M022 97,06 Hz; (4) Modo de Vibração Axial H3: M060 146,57 Hz.

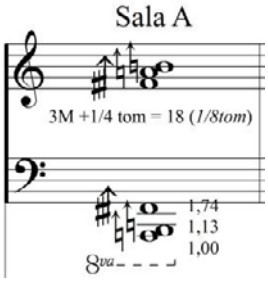


Figura 19 – Representação musical das Alturas correspondendo às frequências dos primeiros modos axiais da Sala A. Pentagramas inferior: (1) altura real e (2) razão proporcional entre as frequências (intervalos musicais) sendo a mesma para as dimensões dos comprimentos de onda entre x, y e z. Pentagrama superior: (3) representação das alturas na configuração mais compacta possível. (4) intervalo entre a nota mais aguda e a mais grave desse agrupamento, em unidades de 1/8 de tom.

Observa-se na parte superior da *Figura 19* três alturas musicais concentradas em apenas um intervalo de terça maior mais ¼ de tom, equivalente a 18 oitavos de tom, ou seja, 1/3 da oitava. Já a *Sala B* apresenta uma relação entre dimensões mais espaçada, tendo na representação compacta um intervalo de 21 oitavos de tom entre a nota mais aguda e a mais grave.

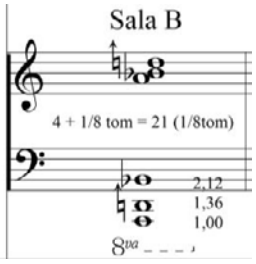


Figura 20 – Representação musical das Alturas correspondendo às frequências dos primeiros modos axiais da Sala B.

Se houver correlação para a qualidade da resposta modal entre a configuração de alturas sonoras e as

proporções das dimensões dos comprimentos de onda dos três primeiros modos axiais poderíamos investigar, mas em outro trabalho, a hipótese de que a constituição desse acorde compacto, quanto mais ela for ampla, mais equilibradamente estariam distribuídos os modos, com menos coincidências de alturas sobrepostas. Comparando a progressão modal de Salas Retangulares (SRT), dada pela *Equação 1*, com outras salas com superfícies irregulares (SIR) como as *Salas A e B* aqui estudadas, com pé direito *H* constante, observa-se que os modos axiais apresentam frequências mais elevadas do que as esperadas para as frequências múltiplas inteiras da frequência de seu modo axial *H1*, como se a altura da sala `diminuísse` virtualmente em relação à altura (pé direito) real. No caso da *Sala A* o primeiro Modo Axial *H1* corresponde ao modo *M003* 48,35 Hz cujo comprimento de onda a 21º C é 7,11 m praticamente duas vezes o pé direito da sala, 3,56 m, com um desvio de 0,01 m.

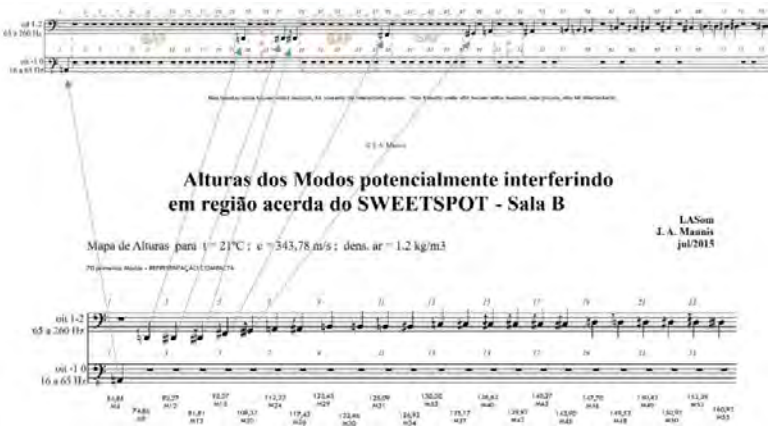
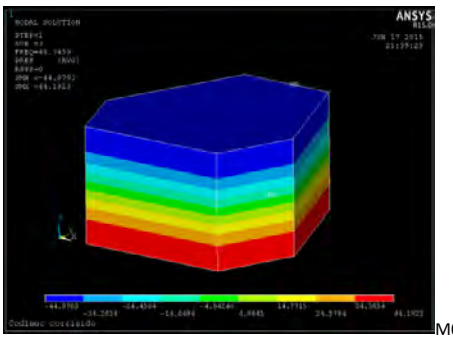


Figura 21 - Distribuição em passos de 1/8 de tom das alturas dos modos com impacto sobre o sweetspot da Sala B.



M003: 48,35Hz

Figura 22 - Sala A: Modo de Vibração Axial H1: M003 48,35 Hz.

O segundo modo axial vertical *H2* é o *M022* 97,06 Hz (*Figura 18 - 3*) cujo comprimento de onda a 21º C é 3,54 m. Seu *H1* virtual seria de 48,53 Hz correspondendo a um comprimento de onda de 7,08 m. 97,06 Hz para o modo axial *H2* é ligeiramente superior à frequência esperada para o segundo modo axial *H2* 96,70 Hz (2 x *H1*). O terceiro modo axial vertical *H3* 146,57 Hz (*Figura 18 - 4*), correspondendo ao *M060*, está também acima do *H3* esperado 3 x *H1* = 145,05 Hz, assim como *H4* 197,18 Hz *(M128)* (*Figura 24*) está acima do esperado 4 x *H1* = 193,40 Hz. Esses desvios encontram-se em progressão linear, conforme apresentado a seguir (*Figura 23*). Observa-se que o modo axial *H4* (*Figura 24*), correspondendo ao modo *M128* 197,18 Hz apresenta na conformação de seus máximos e mínimos de pressão a influência de modos tangenciais, pois as camadas em si estão delineadas em alturas estáveis e as concentrações de máximos (vermelhos) se dão em áreas definidas em um único plano horizontal, intercaladas de camadas estáveis de mínimos (azuis). Essa interferência de modos tangenciais pode estar vinculada à resposta modal causada pela geometria irregular de contorno das paredes laterais, neste caso mais complexa do que numa sala retangular, desestabilizando progressivamente os modos axiais. Isso poderia ser verificado em posteriores trabalhos e seria uma influência bastante positiva, pois os modos axiais são os mais intensos e potencialmente problemáticos podendo induzir ao desequilíbrio na resposta em frequência.

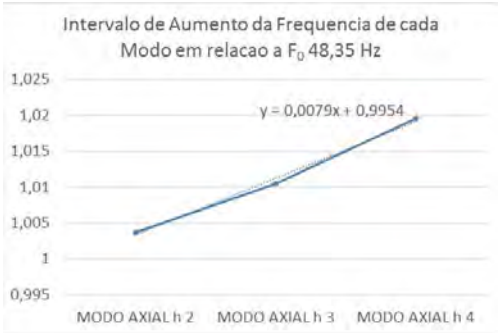
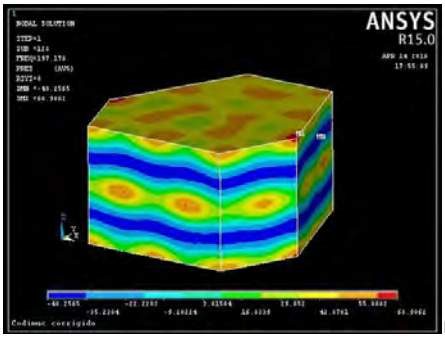


Figura 23 – Intervalo crescente de aumento da frequência dos Modos Axiais Verticais (H2, H3, H4) em relação à frequência do Modo Axial H1 48,35 Hz, modo M003 da Sala A.



M128:197,18Hz

Figura 24 - Sala A: Modo de Vibração Axial H4: com influência de modo tangencial. M128 197,18 Hz.

Medições realizadas ulteriormente, depois dos estúdios concluídos, apontam para discrepâncias em relação à análise do critério de Bonello, no tocante à progressão da densidade modal, mesmo para as frequências produto da simulação. Alguns máximos de pressão localizados próximo ao *sweetspot* podem sugerir tendências contrárias às do critério de Bonello, como ocorre nas medições que resultaram nas Curvas

de Respostas Operacionais (CRO), abaixo ilustradas, das *Sala A* e *Sala B*.

4. MEDIÇÕES APÓS CONCUSÃO DAS OBRAS E INSTALAÇÃO DOS ESTÚDIOS

Depois de concluídas as obras e instalações de ambos os estúdios, procedeu-se à medição dos mesmos, mas visando a aplicação prática prevista para atividades de produção fonográfica, ou seja, a resposta em frequência medida no *sweetspot*.

Observando os resultados das medições do Nível de Pressão Sonora em função da frequência em bandas de 1/3 de oitava apresentadas a seguir para cada uma das salas, pergunta-se como uma mesma sala pode dar respostas distintas, mesmo algumas opostas em relação a outras, ou ainda salientar bandas de frequência que estavam em desvantagem no diagrama de Bonello? Esta questão será abordada ao final deste item, após a exposição dos resultados.

Figura 25 - CRO da Sala A: Nível de Pressão Sonora em função da Frequência, medido no sweetspot. Monitores em 5.1: Ls e Rs em 120º e 135º, com e sem interferência acústica do monitor de vídeo.

Observa-se que as curvas de respostas operacionais extrapolam, em alguns pontos, os limites de tolerância. Conforme a EBU Tech 3276 [9] obter todos os pontos dentro da margem de tolerância é bastante complexo. É comum o uso de equalizadores para pequenas correções em salas de controle. Na mesma figura são apresentadas as linhas que correspondem aos limites de tolerância indicados na referida recomendação.

Os resultados das CRO da *Sala A* para os monitores de áudio frontais esquerdo e direito podem ser vistos no gráfico da *Figura 26*. Os mesmos foram significativamente semelhantes acima de 160 Hz, devido à geometria simétrica da sala. A curva da resposta do monitor esquerdo apresenta uma queda de 3 dB em relação à tolerância em 125 Hz, uma banda na qual temos *M034* 114,94 Hz, *M037* 119,53 Hz, *M042* 123,23 Hz, *M045* 129,26 Hz e *M047*, 131,62 Hz (*Figura 27, Figura 28*), todos atuando ao mesmo tempo no *sweetspot* ou bem perto deste. Para o monitor direito a resposta foi um pouco mais plana nessa região, porém houve uma queda na CRO abaixo de 63 Hz apesar do *TR* mais elevado nessa mesma banda, coincidindo com *M008* 60,81 Hz (si ¼ de tom abaixo) atuando no *sweetspot*. Nas frequências acima de 160 Hz o resultado está dentro da faixa recomendada.

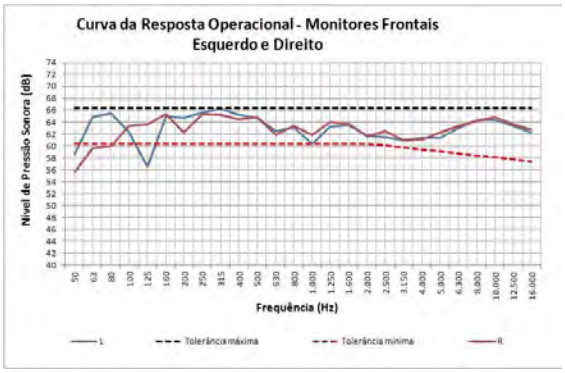


Figura 26 – CRO da Sala A: Nível de Pressão Sonora em função da Frequência, medido no sweetspot para monitores na posição L e R.

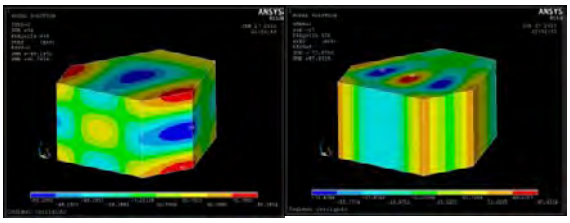


Figura 27 – Sala A : (esq) M034 : 114,94 Hz (dir) M037 : 119,53 Hz

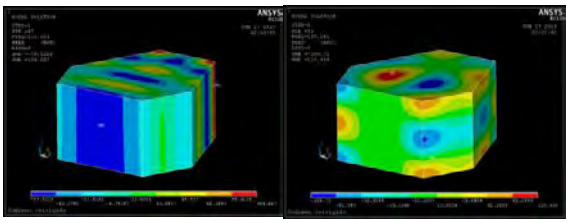


Figura 28 – Sala A: (esq) M047 : 131,62 Hz; (dir) M051 : 135,18 Hz.

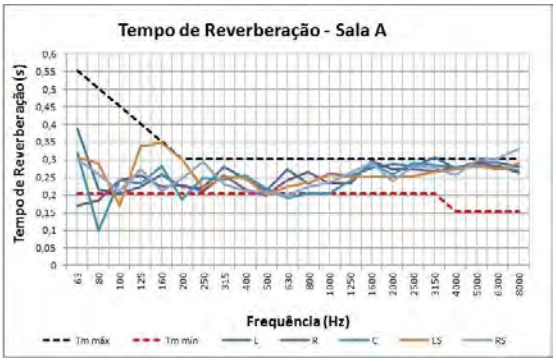


Figura 29 – Tempo de reverberação medido na Sala A em função da frequência em 1/3 de oitava.

Observa-se que o Tempo de Reverberação (TR) está dentro ou muito próximo da faixa de tolerância recomendada, com exceção da frequência de 80 Hz na curva do monitor de áudio central, posição na qual incidem nessa mesma banda de frequência os modos

M012 76,12 Hz, *M013* 77,83 Hz, *M014* 79,44Hz (*Figura 30*). O monitor de áudio esquerdo traseiro *(LS)* apresentou uma elevação no *TR*, em relação às demais curvas, em 160 Hz.

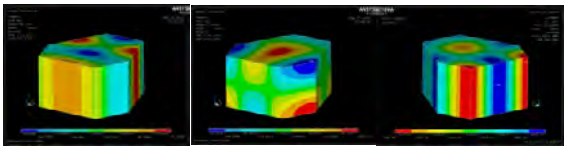


Figura 30 – Sala A: (esq) M012 76,12 Hz, (centr) M013 77,83 Hz, (dir) M014 79,44Hz

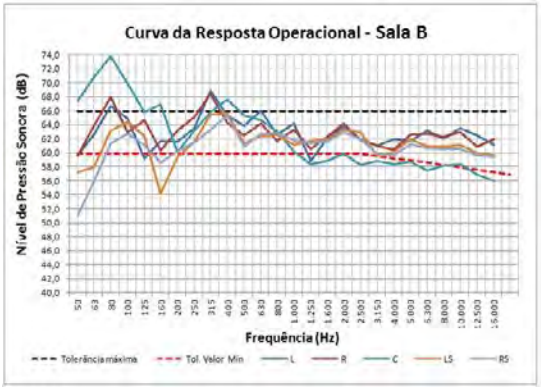
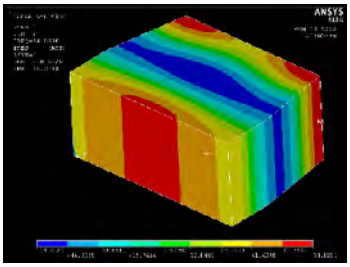


Figura 31 - CRO da Sala B: Nível de Pressão Sonora em função da Frequência, medido no sweetspot.

As CRO da *Sala B* e as curvas de tolerância recomendadas pela EBU Tech. 3276 [9] são apresentadas na *Figura 31*. Observa-se um valor elevado do Nível de Pressão Sonora em 80 Hz na CRO para o monitor central, apesar do Diagrama de Bonello apontar para uma flutuação de contenção nessa banda específica [FLOAT 80] (*Figura 8, Figura 10*). Efetivamente nessa posição e nessa banda de frequência temos a incidência do segundo modo axial de profundidade *P2*, *M09* 74,86 Hz *(si_1 1/8 de tom acima)* e do terceiro modo axial lateral *L3*, *M13* 81,81 Hz *($mi_1$1/8 de tom abaixo) (Figura 32)*. O monitor de vídeo altera significativamente o resultado da CRO, uma vez que parte do som direto não chega ao ouvinte. Nesse estúdio, o monitor de vídeo é de 32 polegadas (~80 cm) e exerce grande impacto no resultado final da CRO para o monitor central ($91 < \lambda/4_{\ 80\ Hz} < 121$cm) podendo atuar para reflexões alimentando os lóbulos de máximos e mínimos de *M13*.



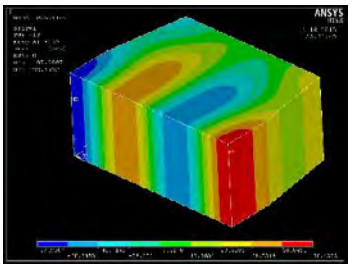


Figura 32 – Sala B: (1) M09 74,86 Hz; (2) M13 81,81 Hz.

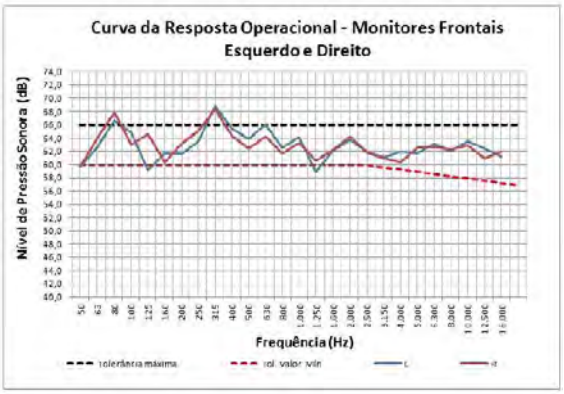


Figura 33 - CRO da Sala B: Nível de Pressão Sonora em função da Frequência, medido no sweetspot para monitores na posição L e R.

Na CRO da *Sala B* referente aos monitores frontais, esquerdo e direito, apresentada na *Figura 33*, observam-se alguns valores de Nível de Pressão Sonora fora da faixa de tolerância, porém na maior parte do âmbito de frequência audível o resultado está dentro da faixa recomendada. Observa-se pouca variação acima de 400 Hz. Os estúdios de ambas as salas analisadas neste trabalho possuem um equalizador destinado a pequenos ajustes de frequência. Contudo tomamos a precaução de desliga-los durante as medições.

Figura 34 - TR medido na Sala B em função da frequência em 1/3 de oitava.

A *Sala B* possui bastante revestimento reflexivo em madeira na parte frontal, o que justifica o tempo de reverberação bem próximo do limite superior, e em algumas frequências ultrapassando a faixa de tolerância indicada pela EBU Tech 3276 [9]. A região dos graves, onde se concentram as ondas estacionárias, apresentou resultado dentro do tempo de reverberação

recomendado, com exceção do monitor direito em 63 Hz, próxima ao modo axial lateral *L2*, *M04* 54,85 Hz *(la_0) (Figura 35)* cujo lóbulo central incide nessa exata região. Nessa posição, o resultado obtido ultrapassou em 0,15 s a tolerância máxima do *TR*.

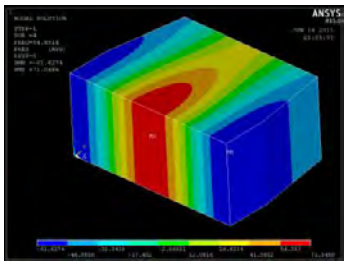


Figura 35 – Sala B: M04 54,85 Hz

Diversas interferências modais podem estar associadas a respostas irregulares ou inesperadas como as que destacamos neste item, pois quando uma fonte sonora se posiciona em uma área na qual há potencialmente um máximo ou mínimo de pressão, o respectivo modo associado a essa pressão se instala e a estacionária se estabelece. Quando houver mais de um modo com máximos ou mínimos na mesma região, há que se considerar as interferências de fase entre os diferentes modos, o que pode inclusive, conforme o caso, causar cancelamentos. Entenda-se assim como um pequeno deslocamento de uma fonte sonora pode causar uma mudança, até surpreendente, na resposta em frequência do sinal sonoro reproduzido pelos monitores de escuta. Isso explica como uma mesma sala pode estimular diferentes respostas para variados posicionamentos das fontes sonoras, neste caso, os monitores de áudio.

5. CONCLUSÕES

Apresentamos, discutimos e comentamos métodos aplicados para análise modal de dois projetos de estúdios de produção fonográfica, implementando novos recursos como a inclusão de notação musical para aperfeiçoar a representação de maneira a que se aproxime o máximo possível de imagens mentais da percepção sonora humana. As análises expostas foram discutidas, por um lado nos levando a algumas conclusões e por outro suscitando suposições e formulações de novas hipóteses a serem verificadas em trabalhos ulteriores. Destacam-se dentre essas conclusões e desdobramentos: (1) para a predição da resposta modal de uma sala, além da tradicional redução a uma sala retangular média *(Figura 3, Figura 4)*, pode-se proceder a uma simulação por elementos finitos *(item 2. PROJETOS ANALISADOS E SIMULAÇÃO)* centrando a atenção da análise nos modos axiais que se revelarem em função da conformação do contorno; (2) contornos irregulares podem influenciar a evolução dos modos axiais a partir de uma certa ordem, interferindo em sua progressão *(Figura 24)*; (3) as posições dos monitores de áudio influenciam a resposta em frequência do sistema formado por eles mesmos integrados aos modos de vibração da sala *(Figura 26, Figura 29, Figura 31, Figura 33, Figura 34)*. Para cada posição de monitores haverá uma resposta correspondente, porém todas remetendo à estrutura do modelo modal geral da sala; (4) Como elemento de design, além das proporções entre as dimensões da redução a uma sala retangular média, pode-se proceder ao exame das relações de proporção entre as frequências dos primeiros modos axiais que se revelarem na simulação *(Figura 19, Figura 20)*; (5) Poderá ser feito um estudo ulterior sobre a hipótese de que determinadas relações intervalares de frequências estejam correlacionadas a constituição de determinados acordes, orientando a uma definição de proporções com resultados de distribuição modal mais equilibrada e homogênea.

6. PÚBLICO ALVO

Pesquisadores e estudantes de acústica de salas aplicada a estúdios de produção fonográfica; Arquitetos; Pesquisadores nas áreas de Acústica musical, Sonologia e afins; Produtores fonográficos.

7. NECESSIDADES DE INFRAESTRUTURA

Datashow para projeção da apresentação.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] GERGES, Samir *Ruído: fundamentos e controle.* 2. ed. Florianópolis: NR Editora, 2000. 676 p.

[2] EVEREST, F. Alton. Small budget recording studio. 2. ed. New York: TAB Books, MacGraw-Hill, 1988. 295 p.

[3] DAVIS, Don; PATRONIS Jr., E. *Sound system engineering*. 3. ed. New York: Focal Press ; Elsevier, 2006. 489 p

[4] EVEREST, F. Alton. Acoustic of small rooms. In: *Handbook for sound engineers: the new audio cyclopedia.* Glen Ballou, editor. 2. ed. Indiana (EUA): Sams / MacMillan, 1987. 1506 p.

[5] EVEREST, F. Alton. *The master handbook of acoustics*. 4. ed. New York: McGraw-Hill, 2001. 615 p.

[6] BONELLO, Oscar J. Acoustical evaluation and control of normal room modes. J. Acoust. Soc. Am., suppl.1, v. 66, n. 2, 1979. In: *Acoustical Society of America*, 98th meeting, 1979b, Salt Lake City, Utah, EUA.

[7] BONELLO, Oscar. A new computer aided method for the complete acoustical design of broadcasting and recording studios, In: *IEEE international conference on acoustics, Speech and signal processing*, 1979. Anais… v. 4, April, p. 326-329, 1979a.

[8] BONELLO, Oscar. A new criterion for the distribution of normal room modes. *Journal of the AES*. v. 29, n. 9, p. 597-606, September, 1981.

[9] EUROPEAN BROADCASTING UNION. EBU Tech.3276: *Listening conditions for the assessment of sound programme material*: *Monophonic and two–channel stereophonic*. Geneva, 1998 2. Ed.

LISTA DE AUTORES / *AUTHOR INDEX*

APOLINÁRIO, I. F. **72**

BISCAINHO, LUIZ W.P **72**
BRANDÃO, E. **33**

DA SILVA, M.J. **56**
DASSIE-LEITE, A. P. **56**
CORDIOLI, A. J. **33**

FARIA, R. R. A. **90**
FORNARI, JOSÉ **39**

GONÇALVES CIENIO, JR. **14**
GONTIJO, W. A. **81**

HOMEM, M. **14**

MASIERO,B **96**
MANNIS, J. A. **106**
MATTOS, T. **106**
MEDEIROS, E. B **23**
MENDES, R. **64**
MENESES, E.A. **39**
MERLO, C. A. **23**
MONTICELLI, P. **90**
MURTA, BERNARDO H. **33**

NOCETI FILHO, S. **81**

PASQUAL, A. M. **23**
PAULA, B. **90**
PAVEI, A. H. **81**
PIRES, L. **46**

ROQUE, T. R. **64**
ROSA, M. **56**

SETTE SILVA, H. **33**
SODRÉ, B. **56**

VIEIRA, M. N. **46**

YEHIA, H. C. **46**

AUDIO ENGINEERING SOCIETY

BRAZIL SECTION

Conceito gráfico,
direção de arte e editoração:

totum
agenciatotum.com.br

Imagens e vetores:

freepik