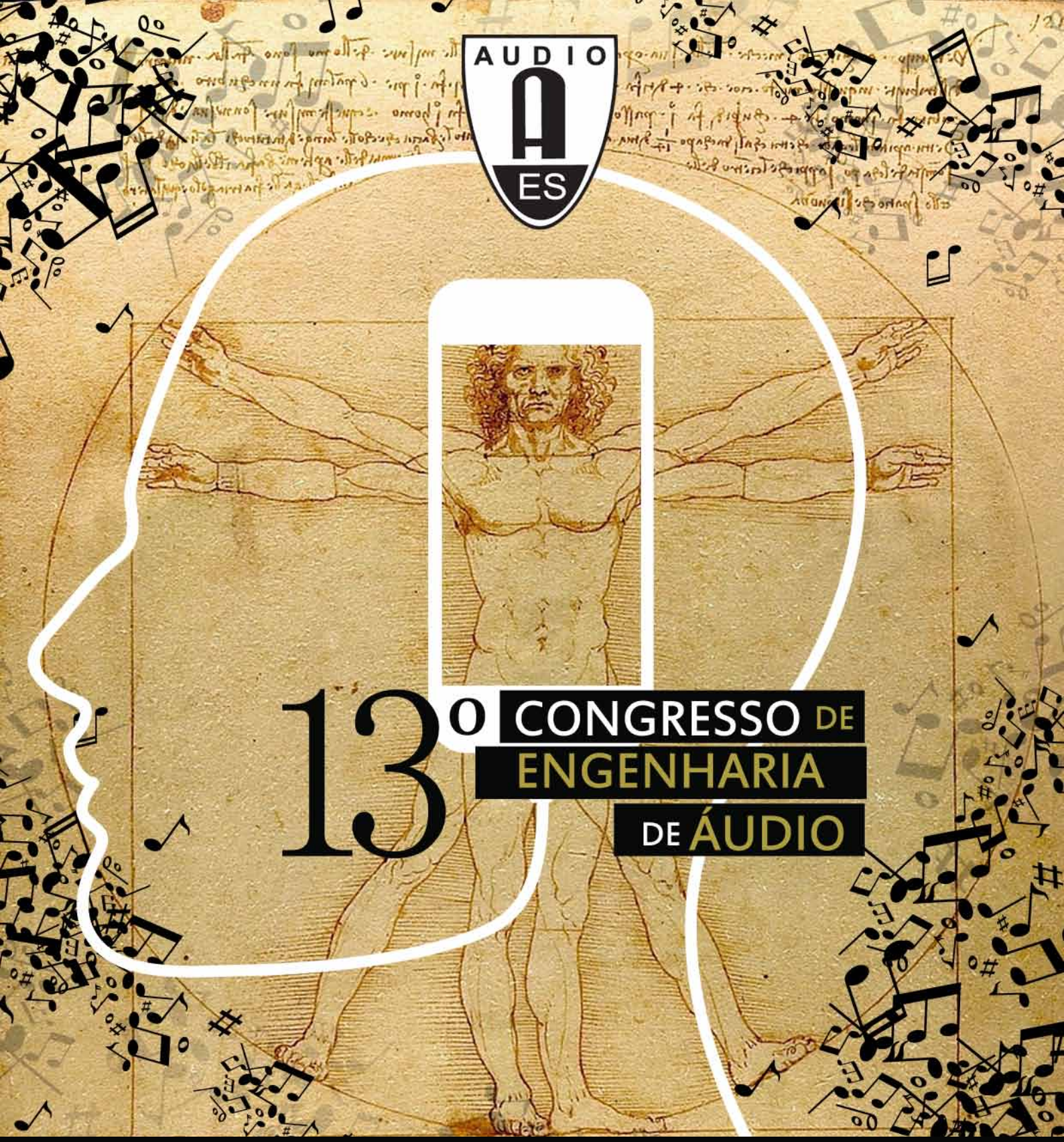


Anais do 13º Congresso de Engenharia de Áudio da AES Brasil
19ª Convenção Nacional da AES Brasil

Proceedings of the 13th AES Brazil Conference
26-28 Maio/May, 2015

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

AUDIO ENGINEERING SOCIETY

BRAZIL SECTION

Anais do 13º Congresso de Engenharia de Áudio da AES Brasil

Proceedings of the 13th AES Brazil Conference

26-28 Maio/*May*, 2015

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

Copyright ©2015

AUDIO ENGINEERING SOCIETY
BRAZIL SECTION

C76a Congresso de Engenharia de Áudio, 13., São Paulo, 2015.
Anais do 13º Congresso de Engenharia de Áudio AES Brasil e
19ª Convenção Nacional AES Brasil / editado por Paulo R. A. Marins
Rio de Janeiro: Sociedade de Engenharia de Áudio, 2015.

Disponível em: www.aesbrasil.org/congressos

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, 19., São Paulo, 2015. 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 Rosa

totum
agenciatotum.com.br

PATROCÍNIO

APOIO

REALIZAÇÃO

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)
Sidnei Noceti Filho (LINSE-EEL-CTC-UFSC)
Regis Rossi A. Faria (FFCLRP-USP)

Coordenadores de Artigos ***/ Papers Chairs***
Luiz Wagner Biscainho (UFRJ)
Diego Barreto Haddad (CEFET/RJ)
Hugo Tremonte Carvalho (UFRJ)

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

Coordenador de infra-estrutura ***/ Infrastructure Chair***
Júlio C. Lucchi (USJT-IMT-FEI-ITA)

Coordenador de Comunicação ***/ Communication Chair***
Márcio Brandão (UnB)

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

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

Alexandre Leizor Szczupak	- Universidade Federal do Rio de Janeiro - UFRJ
Amaro Lima	- Centro Federal de Educação Tecnológica do Rio de Janeiro - CEFET-RJ
André Dalcastagnê	- Instituto Federal de Santa Catarina - IF-SC
Andreas Ellmauthaler	- Halliburton
Carlos Araújo	- Universidade Federal da Paraíba - UFPB
Carlos Ynoguti	- Instituto Nacional de Telecomunicações - Inatel
Carlos F. T. Soares	- Universidade Federal do Rio de Janeiro - UFRJ
Diego Haddad	- Centro Federal de Educação Tecnológica do Rio de Janeiro - CEFET-RJ
Fabio Figueiredo	- Universidade de São Paulo - USP
Fernando Iazzetta	- Universidade de São Paulo - USP
Fernando Pacheco	- Instituto Federal de Santa Catarina - IF-SC
Flávio Ávila	- Universidade do Estado do Rio de Janeiro - UERJ
Hugo Carvalho	- Universidade Federal do Rio de Janeiro - UFRJ
Isabela Apolinário	- Universidade Federal do Rio de Janeiro - UFRJ
Ivandro Sanches	- Centro Universitário da FEI
José Gomes	- Universidade Federal do Rio de Janeiro - UFRJ
José A. Mannis	- Universidade Estadual de Campinas - UNICAMP
Leonardo Nunes	- Microsoft
Lisandro Lovisolo	- Universidade do Estado do Rio de Janeiro - UERJ
Lucas Maia	- Universidade Federal do Rio de Janeiro - UFRJ
Luiz Biscainho	- Universidade Federal do Rio de Janeiro - UFRJ
Marcelo Johann	- Universidade Federal do Rio Grande do Sul - UFRGS
Marcelo Pimenta	- Universidade Federal do Rio Grande do Sul - UFRGS
Marcelo Queiroz	- Universidade de São Paulo - USP
Marcio Brandao	- Universidade de Brasília - UnB
Marcus Bittencourt	- Universidade Estadual de Maringá - UEM
Maurício Costa	- Universidade Federal do Rio de Janeiro - COPPE / UFRJ
Maurício Loureiro	- Universidade Federal de Minas Gerais - UFMG
Michael Vorlander	- Aachen University - RWTH
Michel Tcheou	- Universidade do Estado do Rio de Janeiro - UERJ
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 Santos	- Universidade Estadual de Campinas - UNICAMP
Rodolfo Caesar	- Universidade Federal do Rio de Janeiro - UFRJ
Thiago Prego	- Centro Federal de Educação Tecnológica do Rio de Janeiro - CEFET-RJ
Tadeu Ferreira	- Universidade Federal Fluminense - UFF
Victor Lazzarini	- National University of Ireland - Irlanda
Wallace Martins	- Universidade Federal do Rio de Janeiro - UFRJ
Walter Gontijo	- Universidade Federal de Santa Catarina - UFSC

EDITORIAL

O Congresso de Engenharia de Áudio da AES Brasil chega à sua 13a edição. Ao longo de mais de 10 anos de existência, acima de 200 artigos científicos inéditos foram publicados, o que consolida o congresso como fórum de discussões científicas acerca das pesquisas na área de áudio no Brasil.

Nesta 13a edição estão sendo publicados 12 trabalhos inéditos, sendo 5 artigos completos (full papers), 4 artigos curtos (short papers), e 3 artigos convidados (invited papers) de palestrantes. Os temas desses trabalhos estão relacionadas aos seguintes tópicos: psicoacústica e percepção auditiva, processamento analógico e digital de sinais de áudio, sistemas eletroacústicos, computação musical, acústica de ambientes e salas, tecnologias de microfone e captação sonora, áudio espacial e multicanal, e tecnologia musical. Esta diversidade de temas abordados reforça a capacidade deste fórum produzir uma ampla discussão no que tange à pesquisa científica na área de áudio.

É importante destacar também as 3 palestras apresentadas no Congresso: "Passado, Presente e Futuro do Codificadores Perceptuais de Áudio" por Karlheinz Brandenburg do Fraunhofer Institute - Alemanha ; "Leis Básicas da Psicofísica e Psicoacústica" por Stephan Paul da Universidade Federal de Santa Catarina – Brasil, e "High Resolution Sound Field Analysis-Synthesis for 3d Audio Aplications" por Olivier Warusfel do Institute de Recherche et Coordination Acoustique/Musique – França. Tais palestras refletem novamente a relevância científica do evento, visto que os referidos palestrantes – além de serem pesquisadores reconhecidos na área – são oriundos de instituições renomadas internacionalmente.

Destaca-se também a oferta de 2 mini-cursos: "Arranjo de Microfones Utilizado Arduíno", ministrado por Thiago de Moura Prego do CEFET-RJ, e "Desenvolvimento de Aplicativos para Dispositivos Móveis na Plataforma Android", ministrado por Francisco Henrique de Freitas, também do CEFET-RJ. Tais cursos – além de abordarem temáticas atuais e pertinentes - podem servir de importante fonte de construção de conhecimento para os participantes do evento.

Um outro ponto a se ressaltar é a concomitância da realização do 13o Congresso de Engenharia de Áudio da AES Brasil com a 19a Convenção Nacional da AES Brasil. O fato de os dois eventos acontecerem no mesmo local e simultaneamente possibilita aos participantes não só estar presentes em discussões científicas, como também assistir demonstrações de fabricantes, ou mesmo manusear equipamentos de áudio.

Por fim, gostaríamos de destacar novamente a importância do evento lembrando que o número de participantes vem aumentando a cada edição, o que demostra um crescente interesse por parte de pesquisadores e da comunidade de áudio em geral no que concerne aos temas abordados no Congresso de Engenharia de Áudio da AES Brasil.

Paulo R. A. Marins
Coordenador Editorial

PROGRAMAÇÃO
Terça-feira / Tuesday, Maio / May 26, 2015

10:00-11:20am
Sala 1
Room 1

Sessão de Artigos 1 / *Paper Session 1*
ARTIGOS CURTOS / *SHORT PAPERS*
Chair: Hugo Carvalho

SENSITIVITY ANALYSIS OF AN ACOUSTIC SENSOR LOCALIZATION TECHNIQUE
Diego Haddad - CEFET-RJ, Brasil
Luiz Biscainho, Wallace Martins, Markus Lima - Universidade Federal do Rio de Janeiro

IMPLEMENTAÇÃO DE UM OITAVADOR COM BAIXA DISTORÇÃO
Bruno Griep Fernandes - Universidade Federal de Santa Catarina, UFSC, Brasil
Sidnei Noceti Filho - Universidade Federal de Santa Catarina - UFSC - LINSE/EEL/CTC, Brasil

ANÁLISE E REPROJETO DE MESA DE ALTA QUALIDADE PARA GRAVAÇÃO ACÚSTICA
Marcelo Johann, Andrei Yefinczuk, Rodolfo Pirotti - UFRGS, Brasil
Marcio Chiaramonte - Metalúrgica Meber, Brasil

ESTUDO DO DESENVOLVIMENTO DE UM SISTEMA COMPUTACIONAL PARA TAQUIGRAFIA MUSICAL
Antonio Penteado - Universidade Estadual de Campinas, UNICAMP, Brasil
Jose Fornari - Universidade Estadual de Campinas,UNICAMP, Brasil

1:30-3:00pm
Auditório D
Lecture Room D

Palestra / *Lecture*

LEIS BÁSICAS DA PSICOFÍSICA E DA PSICOACÚSTICA
Stephan Paul - Universidade Federal de Santa Santa Catarina- UFSC, Brasil

3:30-5:00pm
Auditório A
Lecture Room A

Palestra / *Lecture*

HIGH-RESOLUTION SOUND FIELD ANALYSIS-SYNTHESIS FOR 3D AUDIO APPLICATIONS
INTERCÂMBIO DE PESQUISA FRANÇA E BRASIL VIA FAPESP
Olivier Warusfel - IRCAM/CNRS/Universidade Paris VI - França

PROGRAMAÇÃO
Quarta-feira / Wednesday, Maio / May 27, 2015

9:00-11:30pm
Sala 1
Room 1

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

FAN CHIRP TRANSFORM WITH NONLINEAR TIME WARPING
Isabela Apolinário - Universidade Federal do Rio de Janeiro, UFRJ ,Brasil
Luiz Biscainho - Universidade Federal do Rio de Janeiro, UFRJ , Brasil
Martín Rocamora, Pablo Cancela - Universidad de la República, Uruguay

TEMPOS DE REVERBERAÇÃO EM SALAS COMUNS DE RESIDÊNCIAS TÍPICAS NO SUL DO BRASIL: RESULTADOS DE MEDIÇÕES IN-SITU
Michael Klein - UUniversidade Federal de Santa Maria - UFSM, Brasil
Stephan Paul - Universidade Federal de Santa Catarina- UFSC, Brasil

PROJETO E CONSTRUÇÃO DE UMA INTERFACE GESTUAL PARA IMPROVISAÇÃO LIVRE
Eduardo Meneses - Unicamp - Universidade Estadual de Campinas, Brasil
Jose Fornari - Unicamp - Universidade Estadual de Campinas, Brasil
Marcelo Wanderley - McGill University, Canada

AQUISIÇÃO SIMULTÂNEA MULTICANAL PARA APLICAÇÕES EM ARRANJOS DE MICROFONES
Marcelo Costa, Ivandro Sanches
Departamento de Engenharia Elétrica do Centro Universitário da FEI, Brasil

ANÁLISE DA FATORAÇÃO DE MATRIZES NÃO-NEGATIVAS PARA A SEPARAÇÃO DE FONTES ACÚSTICAS VIA DISTÂNCIA EUCLIDIANA QUADRÁTICA E DIVERGÊNCIA DE KULLBACK-LEIBLER
Marcelo Santos, Flávia Freitas - Pontifícia Universidade Católica de Minas Gerais, PUC-MG, Brasil
Zelia Peixoto - PPontifícia Universidade Católica de Minas Gerais, PUC-MG, Brasil

01:30- 03:00pm
Auditório C
Lecture Room C

Palestra / Lecture

O COMPOSITOR E O SOM
Rodrigo Cicchelli Velloso - Universidade Federal do Rio de Janeiro, UFRJ , Brasil

03:30-5:00pm
Auditório A
Lecture Room A

Palestra / Lecture

PASSADO, PRESENTE E FUTURO DOS CODIFICADORES PERCEPTUAIS DE ÁUDIO
Karlheinz Brandenburg - IDMT Fraunhofer - Alemanha

PROGRAMAÇÃO
Quinta - feira/ Thursday, Maio / May 28, 2015

09:00- 01:00pm
Sala 1
Room 1

Mini-curso / Short course

ARRANJO DE MICROFONES UTILIZANDO ARDUINO
Thiago de Moura Prego - CEFET - RJ

02:00-6:00pm
Auditório A
Lecture Room A

Mini-curso / Short course

DESENVOLVIMENTO DE APLICATIVOS PARA DISPOSITIVOS MÓVEIS NA PLATAFORMA ANDROID
Francisco Henrique de Freitas Vianna - CEFET -RJ

SUMÁRIO / *CONTENTS*

COMITÊ DE ORGANIZAÇÃO / *ORGANIZING COMMITTEE* III

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

EDITORIAL V

PROGRAMAÇÃO / *SCHEDULE* VI

PALESTRAS & MINICURSOS

PALESTRA: LEIS BÁSICAS DA PSICOFÍSICA E PSICOACÚSTICA
Stephan Paul 2

PALESTRA: HIGH-RESOLUTION SOUND FIELD ANALYSIS-SYNTHESIS FOR 3D AUDIO APPLICATIONS
Olivier Warusfel 14

PALESTRA: PASSADO, PRESENTE DOS CODIFICADORES PERCEPTUAIS DE ÁUDIO
karlheinz Brandenburg 15

PALESTRA: O COMPOSITOR E O SOM
Rodrigo Cicchelli Velloso 16

MINI-CURSO: **ARRANJO DE MICROFONES UTILIZANDO ARDUINO**
Thiago de Moura Prego 28

MINI-CURSO: **DESENVOLVIMENTO DE APLICATIVOS PARA DISPOSITIVOS MÓVEIS NA PLATAFORMA ANDROID**
Francisco Henrique de Freitas 29

ARTIGO CONVIDADO / INVITATED PAPER

STUDY ON SPATIAL COHERENCE OF MOVING AUDIO-VISUAL OBJECTS
Thomas Sporer, Judith Liebetrau, David Goecke, Karlheinz Brandenburg 31

ARTIGOS CURTOS / SHORT PAPERS

SENSITIVITY ANALYSIS OF AN ACOUSTIC SENSOR LOCALIZATION TECHNIQUE
Diego Haddad, Luiz Biscainho, Wallace Martins, Markus Lima 42

IMPLEMENTAÇÃO DE UM OITAVADOR COM BAIXA DISTORÇÃO
Bruno Griep Fernandes, Sidnei Noceti Filho, ... 47

ANÁLISE E REPROJETO DE MESA DE ALTA QUALIDADE PARA GRAVAÇÃO ACÚSTICA
Marcelo Johann, Andrei Yefinczuk, Rodolfo Pirotti, Marcio Chiaramonte, ... 52

ESTUDO DO DESENVOLVIMENTO DE UM SISTEMA COMPUTACIONAL PARA TAQUIGRAFIA MUSICAL
Antonio Penteado, Jose Fornari 57

SUMÁRIO / *CONTENTS*

ARTIGOS COMPLETOS / *FULL PAPERS*

FAN CHIRP TRANSFORM WITH NONLINEAR TIME WARPING
Isabela Apolinário, Luiz Biscainho, Martín Rocamora, Pablo Cancela 62

TEMPOS DE REVERBERAÇÃO EM SALAS COMUNS DE RESIDÊNCIAS TÍPICAS NO SUL DO BRASIL: RESULTADOS DE MEDIÇÕES IN-SITU
Michael Klein, Stephan Paul 69

PROJETO E CONSTRUÇÃO DE UMA INTERFACE GESTUAL PARA IMPROVISAÇÃO LIVRE
Eduardo Meneses, Jose Fornari, Marcelo Wanderley 75

AQUISIÇÃO SIMULTÂNEA MULTICANAL PARA APLICAÇÕES EM ARRANJOS DE MICROFONES
Marcelo Costa, Ivandro Sanches 82

ANÁLISE DA FATORAÇÃO DE MATRIZES NÃO-NEGATIVAS PARA A SEPARAÇÃO DE FONTES ACÚSTICAS VIA DISTÂNCIA EUCLIDIANA QUADRÁTICA E DIVERGÊNCIA DE KULLBACK-LEIBLER
Marcelo Santos, Flávia Freitas, Zelia Peixoto 91

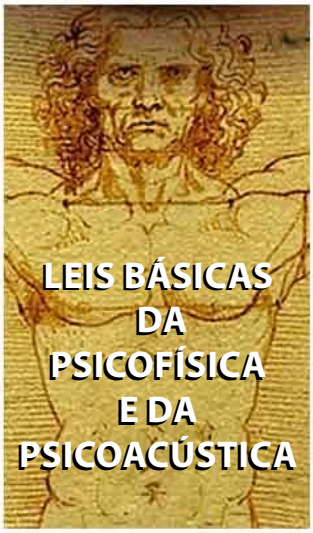
ÍNDICE DE AUTORES / AUTHORS INDEX 99

13º CONGRESSO DE ENGENHARIA DE ÁUDIO

PALESTRA

1:30 -3:00pm
Auditório D
Lecture Room D

Terça-feira
Tuesday
Maio / May
26, 2015



LEIS BÁSICAS DA PSICOFÍSICA E DA PSICOACÚSTICA

13º CONGRESSO DE ENGENHARIA DE ÁUDIO

STEPHAN PAUL
UFSC, BRASIL

Sociedade de Engenharia de Áudio
Palestra em Congresso

Ministrada no 13º Congresso de Engenharia de Áudio
19ª Convenção Nacional da AES Brasil
25 a 28 de Maio de 2015, 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.

Leis básicas da psicofísica e psicoacústica

Stephan Paul[1]

[1] Universidade Federal de Santa Catarina, Centro de Engenharias da Mobilidade e Laboratório de Vibrações e Acústica,
Joinville, SC, 89218-035, e Flornaópolis, SC, 88040-900, Brasil

stephan.paul@ufsc.br

RESUMO

A psicoacústica é a área da psicofísica que investiga a relação entre características físicas do som e as sensações auditivas produzidas. A partir da análise da fisiologia do sistema auditivo e a partir de medições das sensações auditivas a psicoacústica estabelece leis que descrevem a relação entre o mundo físico e as sensações auditivas. A presente palestra pretende introduzir as leis da psicofísica que são válidas também na psicoacústica e mostrar quais são as implicações destas leis para a quantificação de sensações auditivas, por exemplo, da sensação de volume sonoro através de grandezas logarítmicas como o nível de pressão sonora, o nível de *loudness* ou o *loudness*.

0 INTRODUÇÃO

A psicofísica trata, de forma geral, das relações entre estímulos físicos e sensações correspondentes, tanto para descrever esta relação como para saber quais são os processos fisiológicos que transformam estímulos físicos em sensações. Observando que há vários sistemas sensoriais, tais como audição, olfato e tato, entre outros, a psicofísica pode ser subdividida em áreas específicas, tais como psicoacústica, psicofísica da visão, etc.

De forma geral, e independente do sistema sensorial, a psicofísica também pode ser subdividida em duas áreas: a psicofísica externa e a psicofísica interna.

A psicofísica externa, que também é chamada de psicofísica clássica e cujas leis básicas são assunto do presente artigo, pesquisa a relação entre a quantidade de excitação e a intensidade da sensação que caracteriza o evento sensorial produzido pela excitação. A relação é expressa comumente através de uma função, tal como o exemplo da Figura 2. Para achar essa relação é preciso quantificar tanto a característica física do estímulo como também a sensação provocada pelo estímulo. Enquanto a quantificação da característica física normalmente é mais fácil, a quantificação da sensação corresponde é mais difícil; e conforme mencionado, essa dificuldade já fora reconhecida por Fechner. De forma geral, é possível quantificar a sensação através da medição

de uma representação fisiológica, por exemplo das características dos impulsos nervosos e ondas cerebrais gerados pelo sistema sensorial em função da exposição ao estímulo (Figura 1a).

A outro possibilidade, usada na área da psicoacústica com mais frequência, é a medição psicométrica, representada para a área específica da psicoacústica na Figura 1b. Nela mede-se a magnitude da sensação através de um ensaio psicométrico no qual a pessoa exposta a estimulo é questionada sobre a magnitude da sensação associada ao evento sensorial produzido pelo estimulo, conforme exemplificado na Figura 1b. Pelo fato de a pessoa precisar responder, há inevitáveis fatores cognitivos e emocionais, ou seja psicológicos, envolvidos, mesmo que isso possa distorcer o resultado e seja indesejado. Como Fechner utilizou principalmente ensaios psicométricos, conforme mencionado acima, ele chamou toda a ciência que pesquisa a relação entre estímulo e sensação de psicofísica.

(a) Medição de representação fisiológica

(b) Medição psicométrica

Figura 1: Medição das magnitudes das sensações que caracterizam um evento sensorial.

Se a informação sobre a magnitude da sensação for obtida através de ensaios psicométricos, a função que relaciona a magnitude da característica do estímulo e a magnitude da sensação é chamada de função psicométrica (Figura 2).

A psicofísica interna pesquisa sobre os mecanismos envolvidos na transformação de um estímulo físico em evento sensorial caracterizado por sensações, conforme mostra o esquema genérico da Figura 3.

No presente trabalho o foco está na psicofísica externa, ou seja, nas relações entre a quantidade de excitação e a intensidade da sensação que caracteriza o evento sensorial produzido pela excitação do sistema sensorial em questão.

Essas relações são descritas pelas leis da psicofísica, que são universais no sentido de valer para diferentes sistemas sensoriais e sensações e ao mesmo tempo específicas para determinadas sub-áreas da psicofísica, como por exemplo a psicoacústica, nos detalhes.

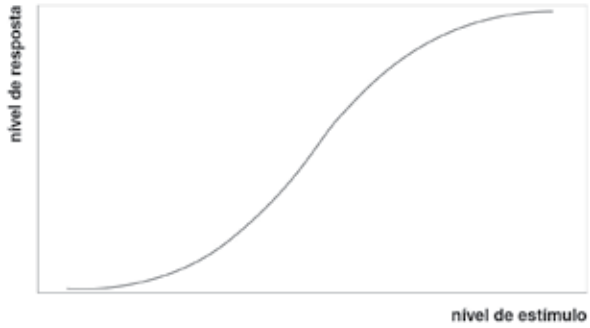


Figura 2: Função psicofísica ou função psicométrica: relação entre a magnitude de uma característica do estímulo e a magnitude de uma sensação produzida pelo estimulo.

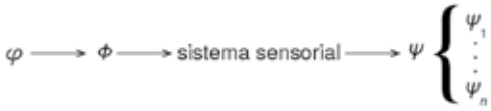


Figura 3: Modelo de transmissão que descreve, de forma geral, a transformação do evento físico φ em estímulo ϕ e este em evento sensorial Ψ caracterizado por sensações Ψ_i.

1 LEIS BÁSICAS DA PSICOFÍSICA

As relações entre característica do estímulo e sensação são, de forma geral, descritas por leis psicofísicas. Entre estas há:

- a lei de sensibilidade diferencial, que na sua forma clássica do século XIX é a lei de Weber [1, p.xvi];
- a lei de Fechner;
- a lei de Weber-Fechner;
- a lei de potência de Stevens;
- a lei de linearidade assintótica, introduzida por Zwislocki [1];
- e a lei de aditividade [1].

Quais destas são consideradas leis básicas da psicofísica depende um pouco da literatura consultada. Zwislocki [1], por exemplo, considera a lei de potência de Stevens como a primeira lei da psicofísica. Além disso, as leis são relacionadas ou substituem uma à outra em determinados casos específicos. A lei de linearidade assintótica, por exemplo, substitui a lei de potência de Stevens perto do limiar absoluto inferior da detectabilidade.

1.1 Lei de Weber

Ernst Heinrich Weber realizou estudos sobre a diferença no limiar do observável entre dois estímulos similares a partir dos experimentos nos quais ele pesquisou a relação entre vibrações e sensação tátil, e entre massa e sensação de peso. Weber teorizou nos anos 1830 a existência de um "limiar de sensação" que deveria ser alcançado antes que alguma mudança de intensidade em um estímulo pudesse ser realmente percebida, ou seja antes que a magnitude da sensação mudasse.

Seus resultados foram publicados em 1834 e 1851, sugerindo que a relação entre a diferença da magnitude de um estímulo ΔE e o estímulo inicial E poderia ser representado por uma constante k_{W} que variaria para cada sistema psicofísico:

$$\frac{\Delta E}{E} = k_{\mathrm{W}}, \tag{1}$$

sendo

ΔE : a diferença de magnitude entre dois estímulos similares;

E : a magnitude do estímulo inicial;

k_{W} : a constante do sistema psicofísico, também chamada de razão de Weber (*Weber fraction*).

A razão de Weber k_{W} seria portanto a diferença no limiar do observável para o estímulo e o sistema sensorial em questão. A partir dos experimentos de Weber e outros determinaram-se as seguintes razões de Weber

- sensação tátil: $k_{\mathrm{W}} \approx 0{,}03$ sendo E a pressão [Pa] sobre a pele;
- sabor: $k_{\mathrm{W}} \approx 0{,}1$ a $0{,}2$ sendo E a concentração do agente;
- peso: $k_{\mathrm{W}} \approx 0{,}02$ sendo E a massa segurada pela mão (que tenha a liberdade de se movimentar).

Mais tarde mostrou-se que a razão de Weber para a grande maioria dos sistemas psicofísicos e sensações na verdade não é constante. Esta constatação, feita por Riesz em 1928, é chamada de *near miss* da lei de Weber. Para algumas sensações a razão de Weber é aproximadamente constante se o estímulo tiver, por exemplo, uma magnitude mínima. No caso da sensação de peso, a magnitude mínima do estímulo seria de 50 gramas para valer a lei de Weber.

1.2 Lei de Fechner

Gustav Theodor Fechner utilizou os resultados de Weber associando-os com variações nas sensações. A partir de experimentos que procuraram pela menor diferença perceptível na sensação Fechner formulou a lei que tem o seu nome:

$$\Delta S = k_{\mathrm{F}} \frac{\Delta E}{E} = k_{\mathrm{F}} r_{\mathrm{F}} \tag{2}$$

sendo ΔS a variação da sensação, ΔE a variação na magnitude do estímulo e $\frac{\Delta E}{E}$ a razão de Fechner r_{F}, não mais considerada constante como tinha assumido Weber.

Para diferenças muito pequenas, as variações na sensação e no estímulo poderiam ser representadas na sua forma diferencial:

$$\mathrm{d}S = k_{\mathrm{F}} \frac{\mathrm{d}E}{E}, \tag{3}$$

sendo S a magnitude da sensação e E a magnitude do estímulo. A Eq. (3) representa então a lei de Fechner na sua forma diferencial.

1.3 Lei de Weber-Fechner

A integração da forma diferencial da lei de Fechner (Eq. (3)) resulta na expressão conhecida como lei de Weber-Fechner (Eq. (4))[1]:

$$\begin{aligned} \int \mathrm{d}S &= \int k_{\mathrm{F}} \frac{1}{E} \mathrm{d}E \\ S - \underbrace{S_0}_{=0} &= k \ln\left(\frac{E}{E_0}\right) \\ S &= \underbrace{2{,}3\, k_{\mathrm{F}}}_{=c} \log\left(\frac{E}{E_0}\right) \end{aligned}$$

$$S = c \log\left(\frac{E}{E_0}\right) \tag{4}$$

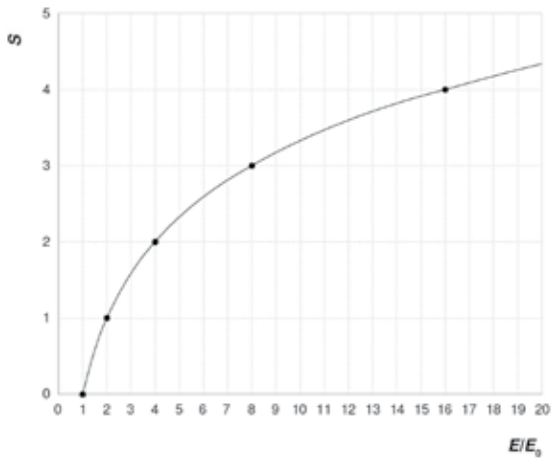


Figura 4: Relação logarítmica entre sensação S e estímulo normalizado E/E_0.

Fechner observou que a função log descrita pela Eq. (4) e exemplificada na Figura 4 é consistente com os resultados de experimentos de Weber que determinaram diferenças no limiar do observável [1].

[1] Zwislocki [1, p.2] considera a Eq. (4) como fórmula (lei) de Fechner, enquanto o próprio Fechner chamou a Eq. (4) lei de Weber. Porém, a maioria dos livros considera a Eq. (4) como sendo a lei de Weber-Fechner.

Esta observação, e vários experimentos no campo da sensação de saturação de cores ou luminosidade mostraram resultados compatíveis com a lei de Weber-Fechner e a consagraram por quase um século como a lei geral da psicofísica [1], apesar de críticas existentes.

Mas a lei de Weber-Fechner, expressa pela Eq. (4), tem uma antecedente importante. Já em 1738 o matemático famoso D. Bernoulli formulara a hipótese que o valor subjetivo do dinheiro, a utilidade do dinheiro, aumenta com o logaritmo da quantidade real de dinheiro através de uma função $\log$ muito comprimida [2, 1], expressando através de uma relação logarítmica uma solução para o modelo de utilidade de Cramer como solução do Paradoxo de São Petersburgo. A função $\log$ está sendo utilizada até hoje para expressar, por exemplo, a utilidade que dinheiro ou *commodities* têm e explica porque em aplicações sem risco o retorno diminui.

A função logarítmica, (ou a escala logarítmica, também chamada escala decibel), portanto, não é exclusiva da acústica ou da psicofísica. A escala logarítmica e o indicador decibel (dB) foram introduzidos, ainda sem este nome, pelos engenheiros dos laboratórios Bell para descrever a perda de energia do sinal elétrico em cabos telefônicos muito longos através da grandeza *transmission unit (TU)*. O nome decibel surgiu seguindo uma proposta de Purves de usar o nome "bell" para lembrar de Alexander G. Bell. Porém, Purves sugeriu que fosse usando o "bel" em vez de "bell" pois "bell" seria um "*everyday English word*" e "bel" ainda iria combinar as variáveis β, ϵ e l que são muito utilizadas na teoria de transmissão de sinais [3, p. 222].

Apesar de ser considerada por muito tempo a lei geral da psicofísica, já desde a virada do século XIX a lei de Fechner (Eq. (2)), e subsequentemente a lei de Weber-Fechner, tem sido criticada. Mostrou-se que ela somente é valida em circunstâncias específicas, normalmente apenas para uma faixa muito limitada de "intensidade" do estímulo. Ainda, para sensações qualitativas, como por exemplo a sensação de altura de um tom, a lei de Weber-Fechner também não se aplica de forma adequada.

1.4 Lei de potência de Stevens

Stanley Smith Stevens[2] descobriu em 1957, ou seja, depois da publicação do *Handbook of Experimental Psychology* [4], que a expansão da lei de Weber para a lei de Weber-Fechner é genérica demais e se aplica apenas para pequenas faixas de intensidade do estímulo.

Os experimentos realizados por Stevens, principalmente na área de psicoacústica e na área da psicofísica da visão [2], mostravam que a relação entre estímulo e sensação pode ser descrita de forma mais geral pela função:

$$S = k_S \cdot (E - E_0)^n \tag{5}$$

sendo

S : a magnitude da sensação

k_S : uma constante de proporcionalidade

E : a intensidade do estímulo

E_0 : a intensidade mínima do estímulo para produzir uma sensação

n : uma potência específica para cada sistema sensorial e para cada tipo de sensação, conforme Tabela 1.

Se $n = 1$, tem-se proporcionalidade entre a magnitude da característica do estímulo E e a magnitude da sensação S, como é o caso para a sensação de comprimento de uma linha reta.[3] Se $n > 1$, a amplitude da magnitude da sensação cresce mais rapidamente que a magnitude da característica do estímulo E, o que é verdade para a sensação térmica. Isso significa que a sensação térmica cresce muito rapidamente com a temperatura, que é responsável pelo desconforto térmico em climas quentes mas também proporciona proteção contra queimaduras pois a reação ao incremento exponencial da sensação térmica é retirar a parte do corpo exposta ao estímulo térmico. Se $n < 1$, a amplitude da magnitude de sensação S cresce mais lentamente que a magnitude da característica do estímulo E, o que é verdade para a sensação de volume sonoro.

A Eq. (5) ficou conhecida como a Lei de potência de Stevens, mas também tem uma antecedente específica: Em 1728 Cramer, matemático suiço, propusera a função de potência $U = kD^{0,5}$ para expressar a relação entre quantidade de dinheiro D e utilidade U do dinheiro. A lei de Stevens, mais universal que a relação proposta por Cramer, é comumente expressa na nomenclatura original de Stevens:

$$\psi = a \cdot \phi^n, \tag{6}$$

sendo ψ a magnitude da sensação, a uma constante, ϕ a magnitude da característica física do estímulo e n a potência característica do sistema psicofísico que depende ainda de ϕ e pode variar um pouco com aspectos experimentais.

De forma gráfica, a "intensidade" do estímulo e a sensação estão relacionadas conforme mostra a Figura 5. No caso de se usar escalas logarítmicas tanto para o estímulo E como para a sensação S, tem-se que n é a inclinação da reta no gráfico que relaciona $\log E$ e $\log S$.

[2] Professor de Psicologia em Harvard. Chegou a ser nomeado como primeiro professor de psicofísica do mundo em 1962. Ele criou, junto com L. L. Beranek, um conjunto de laboratórios de psicoacústica e eletroacústica para estudar os efeitos do ruído no interior de aeronaves de combate e medidas para reduzir ruído e seus efeitos negativos. Mais tarde Békésy chegou em Harvard, e durante décadas vários pesquisadores importantes na área de psicoacústica passaram pelos laboratórios fundados por Stevens e Beranek, como por exemplo Zwicker.

[3] Tal fato faz com que uma escala em linha seja um instrumento fiel de medição em psicometria.

Tabela 1: Alguns expoentes da lei de potência de Stevens para diferentes sistemas sensoriais e sensações, exceto o sensorial acústico, a partir de [1] e [2].

sensação	exponente medido	condição
vibração (*vibration*)	0,95	excitação do dedo com sinal de 60 Hz
vibração (*vibration*)	0,6	excitação do dedo com sinal de 250 Hz
claridade (*clarity*)	0,5	fonte pontual de luz
claridade (*clarity*)	0,5	*flash* de curta duração
claridade (*clarity*)	1,0	fonte pontual de luz emitindo *flash*
(*lightness*)	1,2	reflectância de papel cinza (*reflectance of gray papers*)
comprimento (*length*)	1	linha reta
área (*area*)	0,7	retângulo
saturação de vermelho (*redness (saturation)*)	1.7	mistura das cores vermelho e cinza
sabor (*taste*)	1,3	açucar
sabor (*taste*)	1,4	sal
sabor (*taste*)	0,8	sacarina
cheiro (*smell*)	0,6	heptano
frio (*cold*)	1	peça metálica em contato com o braço
calor (*warmth*)	1,6	peça metálica em contato com o braço
calor (*warmth*)	1,3	irradiação sobre área pequena na pele
calor (*warmth*)	0,7	irradiação sobre área grande na pele
pressão [na palma de mão] (*pressure on palm*)	1,1	força estática aplicada na pele
peso (*heaviness*)	1,45	levantando massas diferentes
viscosidade (*viscosity*)	0,42	ao mexer fluidos à base de silicone
choque elétrico (*electric shock*)	3,5	corrente elétrica passando pelos dedos

A grande diferença entre a lei de potência de Stevens e a lei de Weber-Fechner reside na fundamentação teórica e no método experimental para se obter informações sobre a magnitude das sensações [2]. Enquanto Fechner, da mesma forma como Weber, determinou diferenças no limiar do observável e construiu uma relação entre estímulo e sensação acrescentando diferenças no limiar do observável, Stevens utilizou um método de estimativa de magnitude na qual as pessoas quantificavam a magnitude da sensação atribuindo um número à mesma e ainda tinham uma referência, tanto

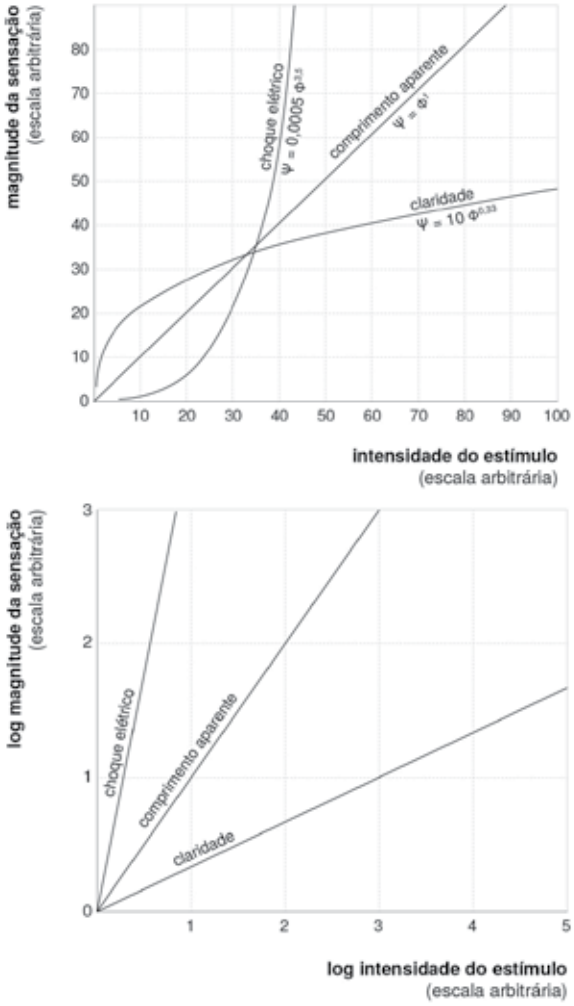


Figura 5: Funções de potência de Stevens para três sensações em escala linear e loglog para E e S. Fonte: [2].

em termos de sensação como em termos de magnitude da sensação. E, o que era impossível para Fechner, a quantificação das magnitudes de sensação por medição de uma grandeza física como representação do processo fisiológico também foi possível. Além de deixar as pessoas estimarem a magnitude de sensação de acidez de soluções de açúcar e limão em relação a uma solução padrão, empregou-se também a medição dos potenciais evocados[4] pelas soluções e transmitidos nos nervos que pertencem ao *nervus facialis*[5]. Observou-se que as magnitudes de sensação obtidas pelo método de estimativa de magnitude e pelo método de medição dos potenciais evocados eram as mesmas, comprovando desta forma a validade dos dois métodos de medição.

Naturalmente, a lei de Stevens também tem limitações. Uma limitação é que ela não é válida para magnitudes do estímulo perto do limiar inferior. Nesta

[4] Potenciais evocados são os potenciais elétricos evocados no sistema sensorial pelo estímulo.

[5] Acessível durante uma operação na orelha média com anestesia local.

região a lei de potência de Stevens é substituída pela lei de linearidade assintótica. Além disso, mesmo na região da validade fisiológica da lei de Stevens, ela sofre, bem como qualquer outra lei psicofísica, com a influência do contexto quando a interface psicológica está sendo utilizada na medição das sensações. Isso é o caso de quase todos os métodos de medição psicométrica.

1.5 A lei de linearidade assintótica

A lei de Weber-Fechner e a lei de potência de Stevens se aplicam apenas para determinadas faixas da intensidade do estímulo, conforme explicado na seção anterior.

Para estímulos com intensidades perto do limiar inferior da detetabilidade, observou-se que as magnitudes de sensação crescem de forma linear com a intensidade do estímulo.

Este fato é mostrado, por exemplo, para o caso da sensação de claridade (*brightness*) através da Figura 6, que relaciona a sensação de claridade com a intensidade da luz, usando para ambos uma escala logarítmica.

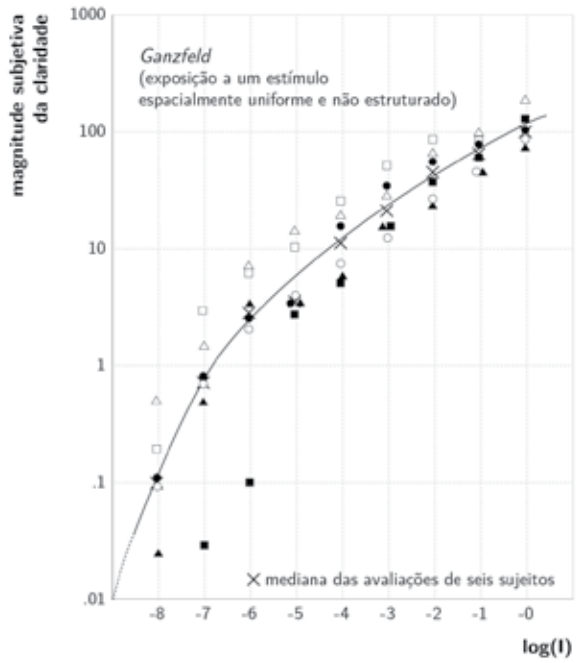


Figura 6: Claridade subjetiva de um flash com duração de 1 segundo produzido por uma lâmpada de tungstênio de 750 W no *Ganzfeld*, que é um estímulo espacialmente uniforme e não estruturado. A claridade foi medida usando o método de estimativa absoluta de magnitude em função da intensidade física da luz. Participaram seis pessoas da medição cujas avaliações da claridade são representados pelos diferentes símbolos. As cruzes indicam as medianas. A curva sólida foi ajustada manualmente e foi extrapolada de forma linear através de linha pontilhada. Fonte: [1].

Evidentemente, a lei de linearidade assintótica se aplica apenas a sensações de intensidade.

1.6 Lei da aditividade

Em alguns casos, as magnitudes de sensação podem ser somadas apesar de os estímulos terem sido processados de forma separada até a etapa de adição. Tal fato pode ser descrito pela Lei da aditividade.

Para o caso da sensação de volume sonoro, a Lei da aditividade se aplica aproximadamente, e dentro de certos limites, por exemplo para a sensação de volume biauricular quando comparada com a sensação de volume sonoro monauricular.

Também para a sensação de volume sonoro causada pela combinação de dois tons com frequências diferentes e que separadamente causam sensações de volume sonoro iguais, a Lei da aditividade é válida conforme mostravam Fletcher e Munson. Com isso, pode-se dizer que a Lei da aditividade é uma condição para a lei de potência de Stevens.

2 PSICOACÚSTICA

A psicoacústica preocupa-se com a transformação do estímulo sonoro em um evento auditivo caracterizado por sensações auditivas. Enquanto a psicoacústica externa tem como objeto de estudo o mecanismo responsável pela transformação (Figura 7), a psicoacústica interna estuda a relação entre magnitude da característico física do estímulo e a magnitude da sensação auditiva correspondente.

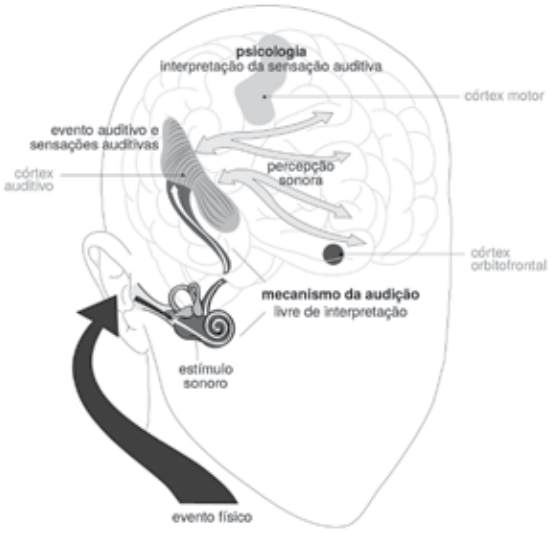


Figura 7: Transformação do evento físico som (evento sonoro) em estímulo sonoro e evento auditivo caracterizado por sensações auditivas e processamento do evento auditivo para dar origem à percepção sonora.

Algumas das leis da psicofísica descritas na seção anterior têm sua aplicação direta na psicoacústica e na acústica em geral. A sensação de altura de um tom depende, dentro de uma determinada faixa de frequências, de forma logarítmica da frequência fundamental. A oi-

tava descreve esta relação. Com relação à sensação auditiva de volume sonoro, a situação é mais complexa. A sensação de volume sonoro para tons puros pode ser, aproximadamente e dentro de uma faixa limitada, descrita como uma função do logaritmo da energia do estímulo.

Ensaios psicométricos com estímulos sonoros foram bastante importantes para elaborar várias leis da psicofísica. O fato de a razão $\frac{\Delta E}{E}$ não ser constante para muitos sistemas sensoriais, estímulos e sensações foi sugerido e comprovado em vários experimentos, por exemplo para a sensação de volume sonoro provocada por tons puros (Figura 8, [5] e Riesz em 1928 para o caso simples da variação da sensação de volume sonoro para tons puros).

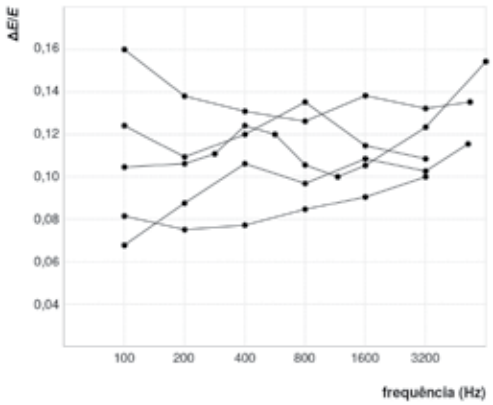


Figura 8: Resultados da medição da razão de Fechner r_{F} para a variação da sensação de volume sonoro para tons puros considerando, entre outras coisas, que todos os tons produzam a mesma sensação de volume sonoro. A partir de [5].

Já para a sensação de volume sonoro causado por sinais de ruído branco de banda larga, a lei de Weber é aproximadamente válida dentro de uma determinada faixa de "intensidade" do estímulo.

Sem dúvida, as leis da psicofísica mais importantes na psicoacústica e na acústica em geral são a Lei de Weber-Fechner, a Lei de potência de Stevens, a lei da linearidade assintótica e a lei de aditividade. Enquanto na acústica geral apenas a Lei de Weber-Fechner encontra aplicação, as demais leis formam a base sólida da psicoacústica e devem ser consideradas também em áreas como áudio, acústica subjetiva ou acústica ambiental.

2.1 A lei de Weber-Fechner na psicoacústica: Níveis sonoros

Na área da psicoacústica e da acústica em geral, a lei de Weber-Fechner (4) origina diretamente grandezas como o nível de pressão sonora para descrever, de forma aproximada, e apenas para uma determinada faixa da magnitude da grandeza física que descreve o estímulo, a sensação de volume sonoro. E de forma indireta, origina também as demais grandezas logarítmicas usadas na acústica, tais como nível de potência sonora, nível de intensidade sonora, nível de exposição sonora, nível de sensação, etc.

Substituindo-se na lei de Weber-Fechner (4) a grandeza E, que expressa a magnitude da característica de interesse do estímulo, pela pressão sonora eficaz ao quadrado p_{ef}^2, obtém-se diretamente:

$$S = c \log \left(\frac{p_{\mathrm{ef}}^2}{p_{0,\mathrm{ef}}^2} \right), \tag{7}$$

ou seja, uma relação logaritmica entre a sensação S e a característica física p_{ef}^2 do estímulo sonoro. Substituindo-se c por 10 obtém-se a definição da grandeza base da acústica, do nível de pressão sonora:

$$L_p = 10 \log \left(\frac{p_{\mathrm{ef}}}{p_{0,\,\mathrm{ef}}} \right)^2 \quad [\mathrm{dB}], \tag{8}$$

sendo L_p o nível de pressão sonora, p_{ef} a pressão sonora eficaz e $p_{0,\mathrm{ef}}^2$ a pressão sonora (eficaz) de referência ($p_{0,\mathrm{ef}} = 2 \cdot 10^{-5}\,[\mathrm{Pa}]$).

Observa-se que o nível de pressão sonora é então uma grandeza psicoacústica, ou mais especificamente, a aplicação da lei de Weber-Fechner para a sensação de volume sonoro provocada por tons puros.

Se o estímulo sonoro tiver pressão sonora eficaz igual a $2 \cdot 10^{-5}\,[\mathrm{Pa}]$ têm-se que

$$L_{p_{\mathrm{ef}}=2\cdot 10^{-5}\,[\mathrm{Pa}]} = 10 \log \left(\frac{2 \cdot 10^{-5}\,[\mathrm{Pa}]}{2 \cdot 10^{-5}\,[\mathrm{Pa}]} \right)^2 = 0\,\mathrm{dB},$$

ou seja, a magnitude da sensação de volume sonoro é zero. Evidentemente, valores negativos do nível de pressão sonora são possíveis, e corretos, sempre que $p_{\mathrm{ef}} < 2 \cdot 10^{-5}\,[\mathrm{Pa}]$. Nestes casos existe o fenômeno físico som, com $p_{\mathrm{ef}} < 2 \cdot 10^{-5}\,[\mathrm{Pa}]$, mas existe sensação auditiva apenas para tons puros com frequências f nas quais o limiar verdadeiro for menor que $p_{0,\mathrm{ef}} = 2 \cdot 10^{-5}\,[\mathrm{Pa}]$, por exemplo frequências entre $\approx 2,5$ e 5 kHz. Para as demais frequências não há sensação auditiva se $p_{\mathrm{ef}} < 2 \cdot 10^{-5}\,[\mathrm{Pa}]$, apesar da existência física do som.

Tendo isso em vista, fica bem evidente que é errado definir som como um fenômeno auditivo, erro que muitos autores cometem [6]. Independente disso, a definição do nível de pressão sonora como grandeza para representar a sensação de volume sonoro, a partir da Lei de Weber-Fechner, causa também outros problemas.

A partir da definição do nível de pressão sonora, é evidente que os limiares inferiores e superiores da audição podem ser descritos tanto em termos do nível de pressão sonora como em termos da pressão sonora eficaz, apesar de o primeiro ser mais comum. Medindo-se os limiares auditivos para diferentes frequências usando algum procedimento psicométrico ou físico (para medir uma representação fisiológica) obtêm-se os limiares que podem ser representados como curvas em

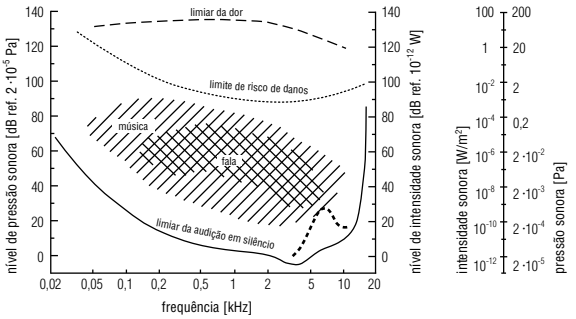


Figura 9: Limiar inferior e superior da audição delimitando a área da audição.

um diagrama e delimitam então a área da audição (Figura 9).

Da mesma forma como ocorreu na psicofísica, a lei de Weber-Fechner consagrou-se como lei universal na acústica, tanto que muitos nem conhecem as sérias limitações que ela tem. Ainda, para sensações qualitativas, como por exemplo a sensação de altura de um tom, a lei de Weber-Fechner também não aplica de forma adequada. A limitação mais óbvia pode ser deduzida da comparação da Eq. 8, assumindo $p_{\text{ef}} = 2 \cdot 10^{-5}\,[\text{Pa}]$, ou seja, o limiar inferior da audição, com a curva do limiar inferior da audição, conforme Figura 9. Enquanto o limiar definido pela Eq. 8 não considera de forma alguma a frequência do estímulo, a curva do limiar inferior que se encontra na prática depende muito evidentemente da frequência, conforme se verifica na Figura 9. Essa limitação da definição do nível de pressão sonora pode ser contornada parcialmente através da definição de curvas de ponderação em frequência, sendo que a curva de ponderação mais conhecida e aplicada, até de forma errada, é a curva de ponderação A.

A partir do nível de pressão sonora são definidas outras grandezas logarítmicas típicas da acústica, psicoacústica a audiologia, tais como nível de potência sonora, nível de intensidade sonora, nível de exposição sonora, nível de sensação etc. Todos estes níveis são de forma genérica níveis sonoros, definidos por:

$$L = 10 \log \left(\frac{X}{X_0} \right) \quad [\text{dB}], \tag{9}$$

sendo X a grandeza que caracteriza a quantidade física em questão e L o nível sonoro, podendo ser o nível de pressão sonora conforme e Eq. (8), o nível de potência sonora $L_W = 10 \log (W/W_0)$ [dB] com $W_0 = 10^{-12}\,\text{W}$, o nível de intensidade sonora $L_I = 10 \log (I/I_0)$ [dB] com $I_0 = 10^{-12}\,\text{W/m}^2$, e assim por diante. Na acústica a o indicador submúltiplo dB da escala log foi introduzido apenas quando engenheiros dos laboratórios Bell sugeriram usá-lo para descrever a sensação de volume sonoro evocadas por sinais de telecomunicação [3, p.219–221].

2.1.1 Escala e grandeza nível de loudness

Barkhausen, nos anos de 1920, introduziu a medição da sensação de volume sonoro por comparação direta entre as sensações de volume sonoro produzidas por um estímulo de referência e o estímulo de interesse. Para tanto um sujeito escutou, através do fone monauricular do aparelho da Figura 10, um tom puro com amplitude ajustável. A tarefa consistia em ajustar a amplitude do tom até que a sensação de volume sonoro provocada pelo tom fosse igual à sensação de volume sonoro produzida pelo ruído em questão, que era escutado com a outra orelha não obstruída pelo fone monauricular.

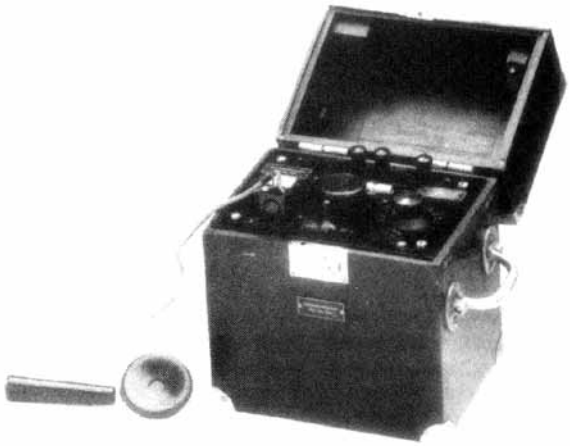


Figura 10: Medidor da sensação de volume sonoro de Barkhausen. Fonte: [7].

A partir dos ensaios foi desenvolvida a escala de nível de *loudness*[6], uma escala logarítmica, que usa o símbolo L_{N} e a pseudo-unidade phon. Com a escala foi definida também uma referência: O nível de *loudness* L_{N} [phon] de um som é o nível de pressão sonora de um sinal harmônico de 1 kHz na condição de onda plana e incidência frontal que produz a mesma sensação de volume sonoro que o som. Assim, um tom puro de 1 kHz e $L_p = 40\,\text{dB}$ tem o nível de *loudness* $L_{\text{N}} = 40\,\text{phon}$. Qualquer som que produza a mesma sensação de volume sonoro terá o mesmo nível de *loudness*. O nível de *loudness* pode ser determinado, na teoria, para qualquer som, mas é mais fácil e conhecido para tons puros.

2.1.2 Curvas isofônicas e curvas de ponderação em frequência

A determinação dos níveis de *loudness* para diferentes níveis de pressão sonora de tons puros com diferentes frequências, dá origem às curvas isofônicas, ou seja, curvas que representam para tons com diferentes frequências igual sensação de volume sonoro.

Ao longo de quase cem anos, inúmeros estudos para determinar as curvas isofônicas foram realizados, sendo

[6] Em português muitas vezes chamado de nível de audibilidade. Porém, o autor do presente artigo discorda do uso desse termo pois audibilidade é um conceito dicotômico, um som pode ser audível ou não, mas não há escala associada.

que os estudos de Kingsbury, Fletcher e Munson, Churcher e King são provavelmente os mais conhecidos. As curvas isofônicas levantadas nos estudos diferem consideravelmente entre si, principalmente em baixas frequências [8].

Essa variação se deve a vários fatores, entre eles técnicos (e.g. dificuldade de se produzir tons puros de baixa frequência com elevado nível de pressão sonora sem distorção harmônica), fatores humanos, e questões experimentais como a forma de apresentação dos tons (via fones de ouvido ou via alto-falantes em campo livre; apresentação monótica ou diótica, método de avaliação).

A apresentação dos estímulos através de fones de ouvido a uma orelha só (apresentação monótica), resulta em um limiar inferior mais elevado e também curvas isofônicas mais elevadas, se comparada à apresentação às duas orelhas ao mesmo tempo (apresentação diótica), pois neste último caso os impulsos nervosos gerados nas duas orelhas contribuem para a sensação de volume sonoro gerado nos córtices auditivos.

O limiar inferior da audição representativo para a condição monótica é chamado de pressão mínima audível[7] (*minimum audible pressure*) (MAP) enquanto o limiar inferior na condição diótica é chamado de campo sonoro mínimo audível (*minimum audible field*) (MAF).

Para baixos níveis de pressão sonora, as curvas isofônicas são muito similares às curvas que representam os limiares absolutos, na condição monótica ou diótica respectivamente. Para níveis de pressão sonora mais elevados, as curvas isofônicas ficam menos curvadas o que faz com que uma curva não é representativa da outra.

Além das curvas isofônicas para tons puros, foram determinadas também curvas isofônicas para ruídos de banda estreita. Entretanto, são muito pouco utilizadas.

As curvas isofônicas indicam então, pelo menos para tons puros e ruídos de banda estreita, a mesma sensação de volume sonoro, e desta forma representam um aspecto importante da audição humana, a sensibilidade diferente do sistema auditivo para diferentes frequências.

Atualmente, as curvas isofônicas estão padronizadas na ISO 226:2003 *Acoustics – Normal equal-loudness-level contours.* A partir das curvas isofônicas determinam-se também algumas das curvas de ponderação em frequência, mais especificamente as curvas de ponderação A, B e C, que são utilizados para corrigir o nível de pressão sonora de tal forma que ele corresponda melhor à sensação de volume sonoro. As curvas de ponderação A, B e C são baseadas na inversão e no "alisamento" das curvas isofônicas de 40, 70 e 100 phon, conforme sugere a Figura 11, que compara a curva isofônica de 40 phon invertida, porém não alisada, com a curva de ponderação A.

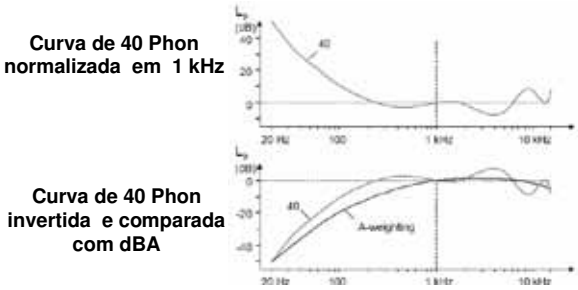


Figura 11: Esquema da transformação da isofônica de 40 phon em curva de ponderação em frequência A.

É possível representar o inverso das curvas de ponderação junto às isofônicas, como feito na Figura 12.

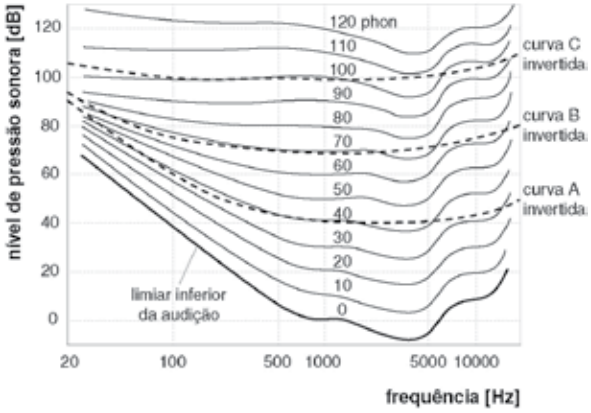


Figura 12: Curvas isofônicas e curvas de ponderação em frequência invertidas.

Como as curvas isofônicas não são exatamente paralelas, pois devido aos mecanismos ativos na cóclea e na orelha média o sistema auditivo não é um sistema linear, as curvas de ponderação A, B e C também tem curvaturas diferentes. Por isso, cada uma delas teria uma faixa de aplicação específica. A curva de ponderação A deveria ser utilizada apenas se o nível de pressão sonora global, e sem ponderação, do ruído em questão for menor que 55 dB. Infelizmente isso não acontece, e quase todas as normas técnicas e leis que dizem respeito à medição acústica em geral ou à avaliação da poluição sonora etc estabelecem o uso da ponderação A.

É muito importante notar também que as curvas de ponderação em frequência A, B e C são baseadas em versões mais antigas das curvas isofônicas e não naquelas padronizadas na ISO 226:2003.

[7] Mais correto seria "mínima pressão sonora eficaz audível", pois a sensação de volume sonoro depende da pressão sonora eficaz.

2.2 A lei de potência de Stevens e a lei da linearidade assintótica na psicoacústica

Stevens realizou experimentos para determinar a relação entre estímulo e sensação, principalmente na área de psicoacústica e na área da psicofísica da visão. Usando o método de estimativa e produção de magnitude, ou seja métodos psicométricos, e de forma mais limitada ensaios para medição de grandezas fisiológicas, ele mostrou que a relação entre estímulo e sensação pode ser descrita de forma mais geral pela função dada pela Lei de Potência de Stevens (Eq. 5), desde que a sensação seja uma sensação de intensidade (e.g. *loudness*) e não de qualidade (sensação de altura de um tom). Para algumas sensações de interesse na psicoacústica e acústica, a Tabela 2 fornece os expoentes obtidos em vários ensaios.

Tabela 2: Exponentes da lei de potência de Stevens para o sistema sensorial acústico, a partir de [1].

sensação	exponente medido	condição
volume sonoro (*loudness*)	0,67	pressão sonora eficaz de um tom de 3000 Hz
volume sonoro (*loudness*)	0,33	intensidade sonora de um tom puro
esforço vocal (*vocal effort*)	1,1	pressão sonora da voz
duração de estímulo sonoro (*duration*)	1,1	estímulo de ruído branco

Observe na Tabela 2 que a potência k para a mesma sensação de volume sonoro é 0,67 quando a magnitude do estímulo for dada em termos da pressão sonora eficaz, mas 0,33 se for dada em termos da intensidade sonora. Portanto, é importante diferenciar entre pressão sonora e intensidade sonora, pois como sabemos da Acústica física são conceitos diferentes, embora relacionados.

2.2.1 A lei de potência de Stevens e a sensação de volume sonoro: Escala e grandeza *loudness*

O nível de *loudness* L_N introduzido na Seção 2.1.1 não é uma medida direta ou modelo da SVS no sentido de não proporcionar informação de como a SVS depende das características do estímulo, pois a escala é obtida por comparação direta com a sensação de volume sonoro provocada por outro som.

Diferentemente do método de comparação utilizado por Barkhausen, Stevens utilizou a técnica de estimativa e produção de razão (*ratio-scaling*) para medir a sensação de volume sonoro produzida por diferentes sons de forma direta. A maioria dos trabalhos subsequentes para determinar a sensação de volume sonoro para os mais variados estímulos também utilizou o mesmo paradigma[8].

A partir das medições diretas da sensação de volume sonoro, Stevens desenvolveu uma nova escala, a escala *loudness* que usa a grandeza do mesmo nome e a unidade sone. Ela se relaciona de forma linear com a sensação de volume sonoro.

Sendo a sensação de volume sonoro uma sensação de intensidade, a mesma pode ser descrita pela sua forma específica da lei de potência de Stevens:

$$N = k_1 L^{k_2} = k_3 I^{k_4} \tag{10}$$

As constantes k_i foram determinadas em vários estudos, sendo que a maioria dos estudos determinou $k_2 = 0,6$ ou $k_4 = 0,3$, o que corresponde à duplicação do *loudness* N se I aumentar em $10\,\mathrm{dB}$. Stevens determinou $k_2 = 0,67$, e neste caso um aumento de $9\,\mathrm{dB}$ corresponde à duplicação de N. Os resultados de outros estudos resultaram em uma constante média de $\bar{k_2} = 0,77$.

Na condição de campo livre, e assumindo uma onda plana, tem-se que $L_p = L_I$ e um tom puro de 1 kHz e $L_p = 40\,\mathrm{dB}$ produz a sensação de volume sonoro de $1\,\mathrm{sone}$, o que equivale para este som específico a $L_N = 40\,\mathrm{phon}$.

Para tons puros existe, então, uma relação simples entre o nível de *loudness* e o *loudness*, conforme mostra a Figura 13. A norma alemã DIN 45631 traz uma relação entre *loudness* N e nível de *loudness* L_N:

$$N = \begin{cases} \left(\frac{L_N}{40}\right)^{(1/0,35)} - 0,0005 & L_N < 40\ \mathrm{phon} \\ 2^{0,1(L_N-40)} & L_N \geq 40\ \mathrm{phon}. \end{cases} \tag{11}$$

Em uma faixa dinâmica considerável, a relação entre L_N e N é descrita pela lei de potência de Stevens e no gráfico da Figura 13 por uma linha reta. Observa-se na Figura 13 e na Eq. (11) que para níveis de pressão sonora inferiores a $40\,\mathrm{dB}$ a sensação de volume sonoro, quantificada na escala *loudness*, cresce mais rápido com o nível de pressão sonora do tom puro, ou seja, o nível de *loudness* deste tom. Já para tons puros com níveis de pressão sonora inferiores a 40 dB, ou seja níveis de *loudness* inferiores a 40 phon, a lei de Stevens não é válida, e nesta região vale a lei da linearidade assintótica que foi introduzida na Seção 1.5.

Existindo uma relação entre o nível de *loudness* (ou o nível de pressão sonora para tons puros) e o *loudness*, é possível representar as curvas isofônicas também como curvas de igual *loudness* conforme feito na Figura 14, pois tanto o nível de *loudness* como o *loudness* são grandezas que representam a sensação de volume sonoro. Da mesma forma, é possível expressar o limiar da audição em termos do nível de *loudness* ou em termos do *loudness*. Expressando-o desta forma, o limiar não depende mais da frequência, pois as

[8] Apesar de ser criticado recentemente na área de psicofísica.

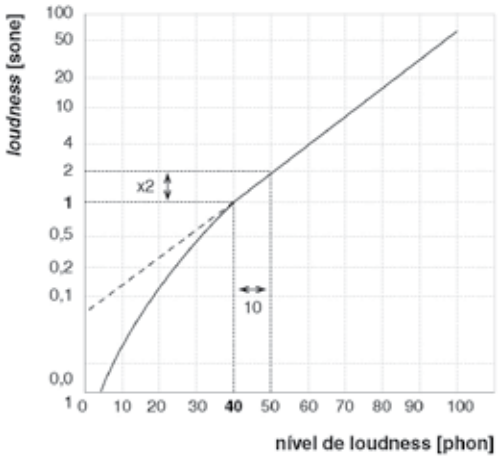


Figura 13: Relação entre *loudness* [sone] e nível de *loudness* [phon] (ou NPS de tom puro de 1 kHz). Fonte: [9].

próprias grandezas nível de *loudness* e *loudness* já consideram, pela forma como foram desenvolvidas, a sensibilidade diferente do sistema auditivo para diferentes frequências. Conforme ANSI S3.4–2007, o limiar da audição corresponde a 2 phon e 0,003 sone.

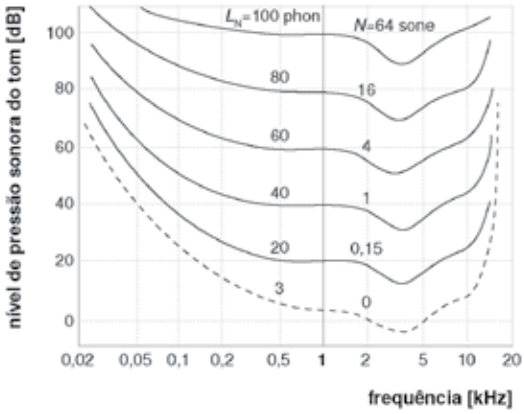


Figura 14: Curvas isofônicas expressas em *loudness* [sone] e nível de *loudness* [phon] (ou NPS de tom puro de 1 kHz)

3 CONCLUSÕES

No presente artigo de revisão, introduziu-se o leitor à psicofísica e suas leis básicas em geral, e sua aplicação na área da acústica e psicoacústica. Discutiu-se a divisão entre psicofísica externa e interna e mostrou-se, de forma sucinta, como na psicofísica externa são obtidas as relações entre magnitude da característica física do estímulo e a magnitude da sensação. A partir dessa introdução foram introduzidas as leis básicas da psicofísica que representam essas relações de forma genérica. Atenção especial foi dada para a Lei de Weber-Fechner, que consagrou-se como lei universal da psicofísica— mesmo tendo sérias limitações —, e a Lei de potência de Stevens, que é a versão mais moderna da lei universal da psicofísica. Para o caso específico da psicoacústica, procurou-se demonstrar a validade e as limitações das leis básicas da psicofísica, em especial da Lei de Weber-Fechner e da Lei de potência de Stevens, para a sensação de volume sonoro. Assim, mostrou-se que o nível de pressão sonora é uma implementação específica da Lei de Weber-Fechner para descrever a sensação de volume sonoro, que é sujeita a sérias limitações. Discutiram-se brevemente as curvas de ponderação em frequência para remediar uma das limitações da grandeza nível de pressão sonora e neste contexto procedeu-se à introdução de outra grandeza logarítmica, do nível de *loudness*, e das curvas isofônicas. Finalmente analisou-se a aplicação da Lei de Potência de Stevens na psicoacústica para o caso da sensação de volume sonoro. Mostrou-se que tal aplicação dá origem a uma nova escala e grandeza para medir a sensação de volume sonoro: o *loudness*.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Jozef F. Zwislocki, *Sensory Neuroscience: Four Laws of Psychophysics*, Springer, 2009.

[2] S.S. Stevens, *Psychophysics: Introduction to Its Perceptual, Neural, and Social Prospects*, Wiley, 1975 (reprint 2000).

[3] Robert T. Beyer, *Sounds of Our Times - Two Hundred Years of Acoustics*, AIP Press Springer, 1999.

[4] S.S. Stevens, Ed., *Handbook of experimental psychology*, Wiley and Sons, 1951.

[5] Vern. O Knudsen, "The sensibility of the ear to small differences of intensity and frequency," *Physical Review*, vol. 21(1), pp. 84–102, 1923.

[6] Stephan Paul, "Som e ruído - releituras críticas de textos brasileiros," in *Anais do XXIII Encontro da Sociedade Brasileira de Acústica*, 2010.

[7] Rudolf Martin, "History of sound evaluation," in *Contributions to Psychological Acoustics - Results of the Fifth Oldenburg Symposium on Psychological Acoustics*, 1991, pp. 11–29.

[8] Y. Suzuki and H. Takeshima, "Equal-loudness-level contours for pure tones," *#JASA#*, vol. 116, pp. 918, 2004.

[9] Stanley A. Gelfand, *Hearing - An Introduction to Psychological and Physiological Acoustics*, Informa Healthcare, 5th edition, 2010.

PALESTRA

3:30-5:00Ppm
Auditório A
Lecture Room A

Terça-feira
Tuesday
Maio / May
26, 2015

HIGH-RESOLUTION SOUND FIELD ANALYSIS-SYNTHESIS FOR 3D AUDIO APPLICATIONS

INTERCÂMBIO DE PESQUISA FRANÇA E BRASIL VIA FAPESP

13º CONGRESSO DE ENGENHARIA DE ÁUDIO

OLIVIER WARUSFEL - IRCAM/CNRS/ UNIVERSIDADE PARIS VI - FRANÇA

PALESTRA

03:30 - 5:00pm
Auditório A
Lecture Room A

Quarta-feira
Wednesday
Maio / May
27, 2015

PASSADO PRESENTE FUTURO DOS CODIFICADORES PERCEPTUAIS DE ÁUDIO

13º CONGRESSO DE ENGENHARIA DE ÁUDIO

KARLHEINZ BRANDENBURG

IDMT FRAUNHOFER

ALEMANHA

PALESTRA

01:30- 03:00pm
Auditório C
Lecture Room

Quarta-feira
Wednesday
Maio / May
27, 2015

O COMPOSITOR E O SOM

13º CONGRESSO DE ENGENHARIA DE ÁUDIO

RODRIGO CICCHELLI VELLOSO
UFRJ

Sociedade de Engenharia de Áudio
Palestra em Congresso

Ministrada no 13º Congresso de Engenharia de Áudio
19ª Convenção Nacional da AES Brasil
25 a 28 de Maio de 2015, 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.

O compositor e o som

Rodrigo Cicchelli Velloso
Universidade Federal do Rio de Janeiro
Rio de Janeiro, RJ, 22421-000, Brasil
rodcv@acd.ufrj.br

RESUMO

Panorama da criação contemporânea, recheado de exemplos musicais, em que analisamos as diversas estratégias de manipulação sonora. Tendências históricas e manifestações recentes são abordadas, nas músicas instrumentais e eletroacústicas: Musique Concrète, Elektronische Musik e Computer Music; Serialismo e Espectralismo; Livre Improvisação e Artes Sonoras.

0. INTRODUÇÃO

No presente texto, discutimos a relação do compositor contemporâneo com o som. Nossa reflexão parte, antes de tudo, de nossa experiência em composição musical. Não se trata, portanto, de um apanhado histórico e/ou musicológico distanciado e exaustivo.

Neste sentido, a menção a tendências e personagens históricos, a discussão de suas ideias e atitudes em relação ao som, e a utilização de exemplos musicais estão necessariamente vinculadas àquelas correntes e àqueles criadores de quem mais nos aproximamos ao longo de nossa própria trajetória - seja porque com eles guardamos alguma afinidade ou, pelo motivo oposto, porque deles nos sentimos estranhos. Ambos nos alimentam e influenciam.

Não queremos dizer, com isso, que nos sintamos tributários ou filiados a esta ou àquela corrente, ou que sejamos contrários a esta ou àquela escola. Trata-se, portanto, de reconhecer que grande parte de nossa reflexão tem sido estimulada por aquilo que mais ou menos nos cai fortuitamente na mão, e que porventura podemos chegar a amar ou rejeitar. Com este legado, temos tentado realizar uma discussão que nos alimenta criativamente, e assim vamos procurando forjar nossa própria ação, mesmo quando ela resulta em silêncio.

Também o professor universitário terá voz neste texto. Sob este aspecto, escritos produzidos anteriormente e as notas de aula de quase duas décadas como docente da Escola de Música da UFRJ, lecionando disciplinas que vão da Composição à Tecnologia Musical, da Harmonia Vocal-Instrumental à Música Eletroacústica, deixarão marcas neste estudo. Observar a reação de meus alunos – muitas vezes estupefata – quando confrontados com esta ou aquela peça de música, com a leitura deste ou daquele texto, certamente influenciará a construção do presente artigo, que tem em mente o público do Congresso de Engenharia de Áudio.

Por fim, encerramos esta introdução esclarecendo de qual música estamos falando. Ainda que menções a outros gêneros possam pipocar aqui e ali, sobretudo nas seções finais do texto, estaremos tratando

daquela música conhecida como clássica, erudita, ou de concerto, para citar algumas das adjetivações mais ou menos correntes em nosso país. Se assim procedemos, não é por desmerecer outras formas de expressão musical, mas porque não somos capazes de tudo abarcar, nem tanto pela falta de espaço, mas por nossas próprias limitações.

Feitas estas ressalvas e esclarecimentos preliminares, comecemos a traçar nosso panorama.

1. O RUIDOSO SÉCULO XX

1.1 Primeira metade

Acredita-se que Claude Debussy (1862-1918) teria dito o seguinte em 1913: "Não será nosso dever encontrar os meios sinfônicos de expressar nosso tempo, meios que evoquem o progresso, o arrojo e as vitórias do mundo moderno? O século do avião merece sua própria música." [1] Para muitos, a nova música, a música pós-tonal, tem início com Debussy e o *L'après midi d'un faune* (1894). Os compassos iniciais

Ex. 1
https://www.youtube.com/watch?v=FcYd3bC3usc

estão construídos sobre uma melodia de flauta com âmbito de trítono – do Dó# ao Sol e daí de volta ao Dó# – realizando um movimento pendular em que diferentes padrões escalares cromáticos não nos permitem identificar um centro tonal preciso. Além disso, a melodia é entoada primeiramente sem acompanhamento, não havendo uma base harmônica que eventualmente explique ou revele se se trata de uma peça tonal. Impossível subestimar a influência de Debussy na música da primeira metade do século XX e além.

Antes de tratar de sua influência sobre outros autores, retornemos, porém, a seu chamado – "O século do avião merece sua própria música." Que música seria essa e que meios teriam que ser inventados para fazê-la eclodir? A frase de Debussy deixa transparecer uma firme e positiva crença nas "vitórias" da tecnologia e das invenções que estavam transformando o mundo ao começo do século XX – ao menos as coisas do mundo. Este mundo transformado necessitaria de novos meios para expressar-se musicalmente.

Outro célebre compositor francês, Édgar Varèse (1883-1965), sintetizou bem a angústia que muitos sentiam em procurar expandir os recursos disponíveis para a produção sonora ao dizer: "Tornei-me uma espécie de Parsifal diabólico à procura, não do Santo Graal, mas da bomba que explodisse o mundo musical, deixando entrar pela brecha todos os sons – sons que na época eram chamados, e algumas vezes continuam sendo hoje, de ruídos." [2]

A distinção entre Som (musical) e Ruído remonta a Grécia Antiga. De acordo com Hugues Dufourt (1943-), os Gregos distinguiam o som musical da desordem dos ruídos, descrevendo as características de uniformidade e regularidade que opõem um aos outros. [3] Este legado – alimentado pela descoberta de Pitágoras (500 AC) que, observando a correspondência entre o comprimento de uma corda em vibração e a nota por ela produzida, estabeleceu as razões aritméticas entre os intervalos harmônicos – foi fundamental para o desenvolvimento da música no Ocidente. Os Gregos, porém, não desenvolveram uma teoria geral sobre a frequência da vibração. Tal empreendimento teria que aguardar o século XVII e os trabalhos de Galileo Galilei (1564-1642), Marin Mersenne (1588-1648) e outros. [4] A Modernidade também testemunhou uma explosão do timbre. Dufourt argumenta que a questão principal do século XVII é a integração entre altura e timbre. A linguagem instrumental, o nascimento da ópera, o temperamento igual são contemporâneos das primeiras conquistas da ciência moderna. O pensamento moderno descarta fatos estranhos que não possam ser assimilados ao tecido da nova visão científica e mecanicista do mundo. Foi apenas quando o som passou a ser encarado como uma espécie de autômato (outra obsessão do século XVII) que se começou a entender as engrenagens que "determinam simultaneamente altura e timbre." [5] O timbre passa a ser compreendido como a sobreposição de harmônicos. Hermann von Helmholtz (1821-1894), com meios limitados de análise mas consciente de que certas características dos sons instrumentais dependem da forma como começam ou terminam, somente estudou o que há de regular e contínuo neles, acreditando que estas características determinam a qualidade musical do som. [6] Helmholtz é o pai da acústica e da psicoacústica contemporânea, mas foi apenas com o advento do sonógrafo e da tecnologia computacional que o som pôde ser analisado em seus menores detalhes. A ideia de que a identidade do timbre é devida a uma invariância física e a concepção do som musical como absolutamente periódico é suplantada pela compreensão do som em termos de ruptura, irregularidade e evolução dinâmica no tempo. Jean-Claude Risset (1938-) chega a afirmar: "Um som musical não está reduzido a um som periódico e poder-se-ia até dizer que ele só é musical se não for periódico." [7]

O chamado de Debussy e a procura de Varèse estavam condicionados, na aurora do século XX e ao longo de sua primeira metade, aos recursos tecnológicos então disponíveis. Teríamos que aguardar a segunda metade do século XX até que a revolução eletroeletrônica produzisse efeitos marcantes no meio musical. Um compositor profundamente influenciado pela linguagem de Debussy, numa de suas páginas mais célebres, reflete

a concepção helmholtziana de timbre. No *Bolero* (1928) de Maurice Ravel (1875-1937), onde um único tema é reorquestrado continuamente, a passagem entre os compassos 137 e 153,

Ex. 2
https://www.youtube.com/watch?v=KK23BhEQVyU

em que os flautins harmonizam em paralelo o tema principal, ilustra talvez, mais que do que um exemplo de politonalidade, um caso inaugural de escrita espectral, em que os parciais superiores da melodia são empregados como efeito de variação de timbre/harmonia, técnica que será tão cara aos compositores de Música Espectral de que falaremos mais adiante.

As inovações de Varèse em direção ao ruído estiveram restritas aos meios instrumentais acústicos, recaindo, sobretudo, na ampliação do aparato percussivo. Uma das primeiras obras de seu curto catálogo, *Amériques* (1921-1923, revista em 1927), para orquestra sinfônica, requer nada menos que treze percussionistas, executando os mais variados instrumentos.

Ex. 3
https://www.youtube.com/watch?v=4hupNWVMuPI

Trata-se de uma partitura que enaltece o mundo moderno tal como vivenciado pelo compositor na Nova Iorque dos anos 1920 e que também mostra sua dívida para com a fragmentação motívica de Debussy e os blocos sonoros e rítmica irregular do Igor Stravinsky (1882-1971) da *Sagração da Primavera* (1913).

Ex. 4
https://www.youtube.com/watch?v=5UJOaGIhG7A

Estas influências são levadas a outro patamar de radicalidade, chegando o compositor do *som organizado* a afirmar: "Digo às pessoas que não sou músico: emprego os ritmos, as frequências e as intensidades." [8]

Varèse não foi o único a buscar na percussão a ampliação do instrumental sonoro. Nos EUA, onde se radicou por grande parte de sua vida, outro compositor-chave do século XX, John Cage (1912-1992) também se notabilizou pela exploração do instrumental percussivo como meio para a obtenção de novas sonoridades. O piano, talvez o instrumento mais emblemático da tradição ocidental, passa a ser preparado numa série de formas inusitadas, com o emprego de pregos, borrachas, cartolinas e outros materiais nas cordas, para produzir sons igualmente inusitados que salientam seu caráter de instrumento percussivo, numa tradição que teve em Henry Cowell (1897–1965), Maurice Delage (1879–1961) e Heitor Villa-Lobos (1887-1959) alguns dos precursores. Do piano preparado inventado por Cage ouvem-se sons ora reminiscentes do gamelão balinês, ora sugestivos de tambores ou outros instrumentos de percussão. *Bacchanale* (1940)

Ex. 5
https://www.youtube.com/watch?v=VZnc1BlUx7Y

é um dos primeiros exemplos. Para um vídeo ilustrativo da preparação de um piano, ver:

Ex. 6
https://www.youtube.com/watch?v=kc3-C7Lnzh0

Esta atitude exploratória em busca de novas sonoridades deve ser encarada como o espírito da época. A ampliação do instrumental percussivo deixou marcas também na música de Villa-Lobos. Aqui destacamos o *Choros No. 8* (1925), que contém dois pianos solistas (um deles com preparações) e grande orquestra, em que se destaca o aparato percussivo.

Ex. 7
https://www.youtube.com/watch?v=TzQh9tujhn4

Para Villa-Lobos, um grande amigo de Varèse, o ritmo será talvez o motor principal da música, mas será outro latino-americano, o cubano Amadeo Roldán (1900-1939), o primeiro compositor no ocidente a escrever peças para um grupo contendo apenas instrumentos de percussão, como é o caso de suas *Rítmicas V e VI* (1930).

Ex. 8
https://vimeo.com/13979484

Os Futuristas italianos, liderados pelo pintor Luigi Russolo (1885-1947), dedicaram-se a desenvolver, já na segunda década do século XX, uma luteria inovadora e puseram-se a construir engenhocas mecânicas que produzissem toda sorte de ruídos. No manifesto *L'arte dei rumori*, uma carta escrita em 1913 ao músico Francesco Balilla Pratella (1880-1955) e transformada em livro em 1916, Russolo clama em favor de uma música de ruídos, e não de notas ("É preciso que se rompa com este círculo restrito de sons puros e que se conquiste a variedade infinita dos 'sons-ruídos'."); por uma música de ruídos relacionados à experiência quotidiana ("Nós queremos entoar e regular harmônica e ritmicamente estes variadíssimos ruídos."); e por uma orquestra futurista com "6 famílias de ruídos" produzidos mecanicamente. [9] Os instrumentos criados são denominados *intonarumori* (entoadores de ruídos) e receberam nomes tão curiosos como *scoppiatore* (explodidor), *ronzatore* (zumbidor), *crepitatore* (crepitador), *stropicciatore* (esfregador), *ululatore* (ululador), *sibilatore* (sibilador), *gracidatore* (coaxador), *gorgogliatore* (gorgolejador), e outros.

"*Intonarumori* são instrumentos musicais monocórdios com extensão de uma oitava, idealizados pelo pintor futurista Luigi Russolo que, em colaboração com o pintor Ugo Piatti (1888-1953), em 1913, realizou alguns exemplares no estúdio da via Stoppani em Milão. (...) O primeiro intonarumore, um scoppiatore, foi apresentado numa noite futurista no Teatro Storchi em Modena em 2 de junho de 1913, enquanto os outros 15 foram

apresentados à imprensa na Casa Marinetti em 30 de setembro. Memoráveis os seus concertos: em 1914 em Milão, Gênova e Londres, enquanto "Intonarumori com orquestra" no Teatro Champs-Elysée de Paris em 1921." [10]

Na plateia parisiense estavam presentes alguns dos personagens aqui mencionados, como Ravel, Stravinsky, Varèse (que se tornou um entusiasta e defensor dos novos instrumentos) e muitos outros criadores em busca da renovação da linguagem musical. Nada resta dos *intonarumori* originais e muito pouco nos chegou da música por eles produzida. *Risveglio di uma città*, para uma orquestra de *intonarumori*, é uma das poucas exceções.

Ex. 9
https://www.youtube.com/watch?v=IC3KMbSkYNI

A luteria de Russolo era exclusivamente mecânica. A primeira metade do século XX testemunhou também a invenção de instrumentos musicais movidos a eletricidade. Foge ao escopo deste texto realizar um inventário de tais inventos. No entanto, cabe observar que os instrumentos elétricos de síntese sonora criados desde fins do século XIX, tais como o Telharmonium (1897), tinham por princípio emitir sons em linha com a noção de nota (musical) e não visavam a romper com a música até então produzida.

Russolo talvez tenha sido o primeiro a tentar uma sistematização dos ruídos, propondo sua classificação em 6 famílias: 1) Estrondos, trovões, explosões, rajadas de sons, quedas, ribombos; 2) Silvos, sibilos, sopros; 3) Cochichos, murmúrios, sussurros, cicios, borbotões; 4) Rangidos, estalidos, roçaduras, zumbidos, crepitações, fricções; 5) Ruídos obtidos com percussão sobre metais, madeiras, peles, pedras, terracotas etc.; 6) Vozes de animais e de homens, gritos, berros, gemidos, bramidos, risadas, estertores, soluços. Por esta razão, ele é considerado uma espécie de precursor da música eletroacústica, sobretudo aquela de inspiração concreta. [11]

A incorporação do ruído ao tecido musical não era, porém, um valor universal, compartilhado por todos os compositores. Mesmo um revolucionário como Stravinsky, que em 1913 produziu talvez uma das obras mais célebres e influentes de todo o século XX, restringiu o universo sonoro com o qual se propunha a trabalhar. Numa de suas seis lições sobre a Poética da Música (1940), refletindo sobre o terror que tomava sua conta ao se deparar com a infinidade de possibilidades que se apresentam ao criador ao iniciar um trabalho, declarou tranquilizar-se pelo pensamento de que tem à sua disposição "as sete notas da escala e seus intervalos cromáticos", dizendo-se plenamente convencido de que as combinações permitidas pelos "doze sons em cada oitava" e por todas as possíveis variedades rítmicas jamais seriam exauridas pelo gênio humano. [12]

Uma visão consolidada da Nova Música da primeira metade do século XX [13] opunha Stravinsky a Arnold Schoenberg (1874-1951), criador do dodecafonismo. Sem romper com o paradigma do som musical baseado na nota, mas procurando uma solução de encadeamento das alturas que rompesse com o sistema tonal, Schoenberg desenvolveu um método em que as notas eram dispostas em sequências (séries), que por sua vez eram submetidas a operações de transposição, inversão e retrogradação, de forma a produzir o vocabulário fundamental à disposição do compositor. Para ele, continuava-se a compor como antes, mas com as novas séries dodecafônicas orientando o encadeamento das alturas. [14] O caminho de Schoenberg em direção ao dodecafonismo passou por fases (e obras-primas) célebres, da tonalidade estendida das primeiras peças ao atonalismo livre de obras como o *Pierrot Lunaire* (1913), talvez sua composição mais conhecida. (Stravinsky gostava de se referir a Schoenberg como o "compositor do *Pierrot*".) Há nesta obra uma ambientação sonora inaudita, que, mesmo hoje, surpreende seus primeiros ouvintes. Neste sentido, o Sprechgesang (ou Sprechstimme), que evoca o canto-falado dos cabarés germânicos da época, tem um papel importante para criar a atmosfera expressionista, ao lado das harmonias dissonantes que nunca repousam em algum centro tonal, configurando-se numa das composições mais emblemáticas do expressionismo musical.

Ex. 10
https://www.youtube.com/watch?v=veUJxETj7-c&list=PLVI9eMoGlM0TbBOT09UmqV21lOc-zRPw_

Encerramos este breve sobrevoo de tendências e peças da primeira metade do século XX com a audição de um trecho do *Concerto para Piano e Orquestra* (1942) de Schoenberg, obra dodecafônica madura.

Ex. 11
https://www.youtube.com/watch?v=A-fyWc6Mpd8

1.2 Segunda metade

Conforme discutimos em outro lugar [15], o pensamento musical da segunda metade do século XX foi profundamente marcado pela polarização estabelecida por Theodor W. Adorno (1903-1969) em sua *Philosophie der neuen Musik*. Reunião de dois estudos produzidos separadamente, nela Adorno opôs dois dos grandes compositores da primeira metade do século: Schoenberg e Stravinsky. Ex-aluno de Alban Berg (1885-1935), discípulo de Schoenberg, Adorno irá cunhar nos títulos destes dois trabalhos a senha que influenciará toda uma geração de criadores do pós-guerra: Schoenberg e o Progresso, Stravinsky e a Restauração. Signo de um tempo de (o)posições radicalizadas, de luta entre

sistemas antagônicos, o texto de Adorno funcionará como um divisor de águas para os praticantes da nova música. Para a nova geração, era preciso optar: de um lado o progresso, a racionalização, a emancipação, o dodecafonismo, a complexidade, representados por Schoenberg e seus discípulos, de outro a restauração estilística, a reação, a arbitrariedade, a tonalidade, a simplificação, empreendidas por Stravinsky (e seus discípulos). A necessidade histórica contra o desejo subjetivo, a severa autoevolução da substância e a organização integral (flexível) em oposição à severidade do fenômeno e à reconstrução da música em sua autenticidade (rígida).

Ambos, porém, reduziam o som musical às doze notas cromáticas fornecidas pela oitava do sistema temperado. O movimento em direção à plena incorporação do ruído no tecido musical encontrou em Pierre Schaeffer (1910-1995) e na sua *musique concrète* a motivação estética, os recursos tecnológicos e uma agenda de investigação que iriam produzir consequências fecundas. Ao final dos anos 1940, trabalhando como engenheiro na Rádio France, Schaeffer irá inventar uma nova forma de produzir música, que ele batizou de Música Concreta. A expressão foi cunhada em 1948 para designar uma música composta a partir de sons gravados em suporte. Esta abordagem se distingue da concepção tradicional de música, em que o compositor imagina sons em sua cabeça e, por meio de notação simbólica, grafa uma partitura que será posteriormente lida por intérpretes que produzem os sons com seus instrumentos. Na música concreta, o ato de composição se dá concretamente, pela manipulação e organização sobre um suporte, tais como discos, fita magnética etc. A obra existe concretamente, como substância audível fixada no meio material. Os sons podem ser de qualquer origem – instrumental, sintética ou ambiental. O termo concreto não designa uma fonte sonora, referindo-se ao ato de composição. Mas a tradição do gênero recairá na escolha de materiais sonoros gravados, chegando a gerar a curiosa e errônea denominação de "som concreto" (em oposição a som eletrônico).

Em 1948, Schaeffer difunde pela rádio os seus *Cinq études de bruits*. O *Étude aux chemins de fer*, produzido integralmente pela manipulação de sons de locomotivas, é uma das peças que compõem o ciclo.

Ex. 12
https://www.youtube.com/watch?v=N9pOq8u6-bA

O pensamento de Schaeffer irá partir de alguns conceitos e achados fundamentais, dos quais destacamos a situação *acusmática* de escuta, que designa a audição de um som sem ver sua causa material, por meio de alto-falantes, por exemplo, como na escuta do rádio. O termo *acusmático* remonta a Pitágoras (500 a.C.), que ensinava seus discípulos oculto por detrás de uma cortina. Assim, seus alunos não poderiam se distrair com o aspecto visual do mestre, concentrando-se em suas palavras. Isso terá consequências importantes para Schaeffer, que passará a realizar todo um trabalho que reavaliação do som musical – um novo solfejo, culminando na publicação do *Traité des objets musicaux* [16]. Para ele, muito do que se acreditava ouvir seria informado pelo aspecto visual. Ao desconectar um som de sua fonte, o que ouvimos? Como descrever estes sons e como incorporá-los na produção de uma nova música? Schaeffer irá tomar emprestados alguns conceitos da Fenomenologia (*époché* e redução fenomenológica) para desenvolver uma série de categorias (escuta reduzida, objeto sonoro) e critérios de classificação baseados na percepção auditiva e nas qualidades sensíveis dos objetos sonoros (Tipo-Morfologia), visando à incorporação de todos os sons no tecido musical. "A música mais inclusiva que pode haver", teria dito Olivier Messiaen (1908-1992).

Schaeffer era um tradicionalista e tinha horror ao dodecafonismo de Schoenberg. Numa entrevista concedida ao final da vida, declarará que os franceses tinham sido bem sucedidos em expulsar a dominação territorial alemã, mas que a ocupação cultural germânica permanecia, com a predominância da técnica dodecafônica. Acreditava que talvez ele, não sendo músico, pudesse fazer emergir uma nova maneira de compor utilizando os aparatos tecnológicos então à disposição. [17]

A oposição franco-germânica perdurará nos anos 1950 no domínio das novas músicas produzidas com novos aparatos tecnológicos. De um lado a *musique concrète* francesa, de outro a *Elektronische Musik* alemã. Esta polarização guarda pontos de contato com o antagonismo adorniano: o dodecafonismo e seu sucessor direto, mais racional e objetivo, o serialismo integral, estavam aí implicados, com entusiasmo (pelos eletrônicos) e com horror (pelos concretos). Tratava-se também de uma oposição de escolhas tecnológicas e de materiais, de sons e sonoridades, de procedimentos e técnicas e de atitudes de escuta.

O termo *Elektronische Musik* foi cunhado por Werner Meyer-Eppler (1913-1960) em 1949 para designar uma música feita a partir de equipamentos eletrônicos de produção sonora. As experiências de Herbert Eimert (1897-1972) e Robert Beyer (1901-1989) na rádio NWDR de Colônia visavam ao "racionalismo e abstração" na criação de novos sons por meio da síntese analógica, que viria também a ser conhecida como Escola Senoidal, devido à rejeição de sintetizadores com timbres pré-programados em favor de osciladores de voltagem. A Elektronische Musik logo se confundirá com música serial. Para tal, concorre a contribuição decisiva de um jovem

compositor, KarlHeinz Stockhausen (1928-2007), que irá fornecer as bases técnico-composicionais e estéticas para o desenvolvimento desta nova música eletrônica. Para Stockhausen, o princípio serial trabalha com um número limitado de grandezas distintas. Estas grandezas devem ser proporcionais e estruturadas de acordo com sequências e intervalos determinados. A escolha serial se aplica a todos os elementos de uma composição. As séries originárias geram séries derivadas e as proporções das séries são um princípio estrutural totalizante, em busca da "necessária consequência formal." [18] Neste sentido, o som em si será resultado de um número determinado de operações seriais que coordenam atributos como frequências, durações, intensidades e envelopes dinâmicos, e a quantidade e disposição de parciais. Ainda que Stockhausen estivesse consciente de que os chamados "quatro parâmetros do som" (altura, duração, intensidade e timbre, tal como definidos classicamente no serialismo musical) [19] não fossem atributos da mesma ordem, a aplicação de fórmulas seriais (numéricas) tendia a assim homogeneizá-los. Os *Estudos* do começo dos anos 1950 são exemplos da aplicação exaustiva de métodos seriais. Ouçamos o primeiro deles, composto em 1953.

Ex. 13
https://www.youtube.com/watch?v=1GoUzk6fQAA

Como bem sintetizou o pai da Música Concreta, em texto publicado em meados dos anos 1960 [20], a dicotomia Concreto x Eletrônico representava, porém, uma simplificação insustentável. De um lado estariam os "dissecadores de cadáveres" (os concretos), de outro os "construtores de robôs" (os eletrônicos). [21] "A música viva estava em outro lugar e só se revelaria àqueles que soubessem escapar destes modelos simplistas." [22] Schaeffer preferiu abandonar o termo concreto em 1958 em favor de experimental, mas o território da música produzida por meios eletroeletrônicos passou a ser genericamente denominado música eletroacústica. Um dos principais clássicos do gênero, *Gesang der Jünglinge* (1955-1956), de Stockhausen, ao incorporar a voz de um menino aos sons eletrônicos sintéticos, contribuiu para a percepção de que o campo eletroacústico finalmente realizara uma obra de síntese, com o perdão do trocadilho. Mas não se iludam os leitores/ouvintes, o *Gesang* é composição de inspiração e aplicação serial meticulosa, ainda que parte de sua realização tenha sido seja deixada a cargo dos intérpretes/técnicos no estúdio, seguindo orientações gráficas esboçadas por Stockhausen (e sob sua "batuta"), aflito com a lenta evolução do trabalho.[23]

Ex. 14
https://www.youtube.com/watch?v=nffOJXcJCDg

Posteriormente, o campo eletroacústico, como gênero alargado da expressão musical contemporânea, passará a abarcar também a música produzida por meios computacionais, a Computer Music surgida nos EUA a partir de finais dos anos 1950.

O que a acústica contemporânea irá inaugurar, como enunciado anteriormente, é a compreensão do som em termos de ruptura, irregularidade e evolução dinâmica no tempo, não sendo possível reduzi-lo a modelos simplistas. A intervenção da tecnologia computacional para a sua análise foi um fator determinante neste sentido. Ao considerar as teorias da percepção psicofisiológica correntes ao final do século XX, a matéria se torna ainda mais complexa. Sumariamente, não há uma associação simples dos parâmetros de estímulo físico com as respostas da percepção auditiva: a altura não é uma função unidimensional da frequência; a intensidade (*loudness*) não apenas depende da amplitude, mas também da frequência e do timbre; a duração rítmica não é seguida pela percepção como subdivisão cronométrica do tempo, mas depende do contexto musical; e o timbre é o resultado perceptivo de interferências de elementos físicos, que dependem não somente das propriedades espectrais dos harmônicos de um som (número, natureza frequencial, intensidade, distribuição da energia), mas de variações temporais (frequências formantes, ataque, evolução dinâmica, modificação do envelope espectral etc.). [24] O som deixa de ser dividido em diferentes categorias, das quais a altura seria a principal?

Esta agenda foi incorporada nos esforços de criação de novos modelos que orientassem a síntese de sons por meios digitais. O trabalho de Jean-Claude Risset em seu catálogo de 1969 [25] leva em consideração muitos dos aspectos mencionados acima e forneceu o marco (e os códigos!) a partir de onde os criadores poderiam desenvolver novos sons. Trabalhando em diversas oportunidades nos Laboratórios da Bell Telephone Company ao lado do pioneiro da criação de programas computacionais para a síntese digital, Max Mathews (1926-2011), Risset irá em seguida colaborar em centros de pesquisa associando música e ciência, como o IRCAM - Institut de Recherche et Coordination Acoustique-Musique, fundado por Pierre Boulez (1925-) em 1975. *Songes* (1979) é uma obra que incorpora os frutos de suas pesquisas e experimentos e teve grande impacto no desenvolvimento da música eletroacústica produzida por meios computacionais.

Ex. 15
https://www.youtube.com/watch?v=8fSKk4OqZp0

Debrucemo-nos agora sobre duas tendências composicionais que emergiram no último quarto de século XX e que jogam nova luz sobre a questão, renovando o debate acerca do que é afinal um som musical. De um lado, os compositores do movimento denominado Música Espectral e de outro os compositores da Nova Complexidade.

Conforme discutimos em outro texto [26], em diferentes países da Europa ao final dos anos 1960 e começo dos anos 1970, jovens compositores passaram a utilizar análises espectrográficas do som como substrato para a criação de estruturas harmônicas instrumentais. Na França, dois deles, Tristan Murail (1947-) e Gérard Grisey (1946-1998), tornaram-se alguns de seus mais conhecidos e influentes praticantes. Em grande parte, o ímpeto inicial destes compositores se produziu como uma reação ao serialismo. Muitos dos jovens criadores não se sentiam à vontade com os grilhões conceituais, estilísticos e técnicos da música serial. Sensíveis a manifestações alternativas – músicas eletroacústicas, as obras de György Ligeti (1923-2006), Giacinto Scelsi (1905-1988), Iannis Xenakis (1922-2001), dentre outros, e interessados em acústica e psicoacústica – passaram a trocar experiências e reunir esforços na produção de uma nova música.

Murail já descreveu o espectralismo como uma estética e não como um conjunto de técnicas fechadas. [27] As escolhas técnicas que caracterizam o espectralismo se pautam tanto por uma atitude aberta e abrangente em relação ao material sonoro e musical quanto por uma reação à combinatória e à escolástica serial. Optam pelo som, e não pela nota (o símbolo grafado em uma partitura). O termo espectralismo é, para alguns, inapropriado e portador de um reducionismo enganoso. [28] De fato, a abertura mencionada anteriormente diz respeito à concepção alargada do fenômeno musical, entendido, no limite, como som evoluindo no tempo. As estratégias de confecção desta evolução do som no tempo estando sujeitas à exploração das fronteiras entre timbre e harmonia e ao livre curso da fantasia criativa. Os dogmas seriais, suas interdições e (de)limitações conceituais e sonoras representando camisas-de-força intoleráveis para a criação de uma nova música que ousasse trabalhar na amplificação do sonoro para tocar (e estender) as fronteiras do inaudito. Numa declaração célebre, Grisey afirmou: "Somos músicos e nosso modelo é o som e não a literatura, o som e não as matemáticas, o som e não o teatro, as artes plásticas, a física quântica, a geologia, a astrologia ou a acupuntura." [29] Usando o som como modelo, o compositor espectral o analisará, (res)sintetizará e submeterá a uma série de transformações, controladas pela (e destinadas à) percepção auditiva. O compositor espectral rejeita, assim, a adoção de modelos extramusicais, tão comum na música da segunda metade do século XX, e se distancia daquilo que não seja perceptível ao ouvido. Numa das obras marcantes da fase inicial do Espectralismo, *Désintegrations* (1982-1983) de Tristan Murail, amalgama sons de síntese digital (aditiva), com um conjunto instrumental alargado.

Ex. 16
https://www.youtube.com/watch?v=HRuxHVWfQtA

Um dos críticos do Espectralismo, no entanto, qualificou esta abordagem como altamente problemática por "ausência de dialética" e "carência de ritmo" e chegou à conclusão de que ela rejeita o conceito de *écriture*. [30] *Écriture* foi definida por Pierre Boulez como uma espécie de combinatória de objetos, no sentido mais amplo do termo, ou seja, colocar objetos em jogo e em relação, interconectando-os num contexto criado para isso. "A natureza do objeto e a lógica de encadeamento estando, então, ligadas uma à outra por um conjunto praticamente ilimitado de modos de emprego, não derivando forçosamente de uma lógica racional." [31] Para os músicos da tradição serial, o valor de uma composição musical está ligado a e depende deste conjunto de técnicas legadas ao longo da história. No entanto, para que a combinatória possa operar de forma plenamente satisfatória, é preciso proceder a uma redução do fenômeno sonoro, tomando-o não na totalidade de sua complexidade: "Na minha abordagem do objeto musical, eu o considerei exclusivamente sob o signo da altura." [32] Assim, paradoxalmente, para os compositores da tradição serial é necessário proceder a uma simplificação do fenômeno sonoro para que a composição musical floresça em toda a sua complexidade.

Firmemente calcada na tradição serial, encontra-se a tendência denominada Nova Complexidade. O termo nova complexidade surge nos anos 1980 para designar uma abordagem composicional que coloca em valor e radicaliza aspectos relacionados à combinatória de objetos notacionais em complexas tramas evolutivas (processos) que afetam durações, alturas, dinâmicas e timbre. [33] Duas de suas figuras de proa, os compositores britânicos Brian Ferneyhough (1943-) e James Dillon (1950-), exerceram grande influência internacional, notadamente o primeiro, em virtude de sua intensa atuação pedagógica em diversas instituições de ensino e em cursos internacionais dedicados à música contemporânea. A nova complexidade é herdeira direta do expressionismo dodecafônico e do serialismo que o sucede, constituindo a corrente pós-serial mais influente dos dias atuais.

O aspecto superficial mais marcante desta corrente é a produção de partituras de extraordinária complexidade e sofisticação notacional, saturada de complicações rítmicas (uso permanente de quiálteras, ritmos e fórmulas de compassos "irracionais"), melódicas (uso intenso de microtons, afinações alternativas, saltos angulosos) e expressivas (frequentes e abruptas mudanças de andamento e intensidade), fazendo apelo a um vasto repertório de técnicas instrumentais estendidas. A dificuldade de realização destas obras atingiu reputação inaudita, e apenas intérpretes de virtuosidade destemida têm se aventurado em sua execução. Pelas características da

complexidade notacional, avaliamos que as obras mais bem-sucedidas da Nova Complexidade são aquelas que trabalham com conjuntos instrumentais de timbre homogêneo. Para fazer emergir um valor complexo, será necessário atenuar ou simplificar o objeto colocado em movimento? O *3º Quarteto de Cordas* (1987) de Brian Ferneyhough é um bom exemplo.

Ex. 17
https://www.youtube.com/watch?v=hBlbFHl74Xw

A despeito dos muitos aspectos que as distinguem e mesmo as opõem, ambas as correntes fazem apelo à ideia de processo, conceito que esteve inicialmente associado não só a compositores do minimalismo norte-americano – Steve Reich (1936) e sua 'Music as a Gradual Process' [34], mas também a alguns dos protagonistas do universo vanguardista europeu. De forma geral, processo em música está relacionado a uma ação objetiva que gera uma transformação qualquer no material musical. No caso de *Reed Phase* (1966) de Reich, o processo é de natureza repetitiva, fazendo com que um fragmento rítmico-melódico seja defasado de si mesmo, gerando novos padrões.

Ex. 18
https://www.youtube.com/watch?v=Gw7jxjGXOI8

Os processos podem também ser fruto de: operações do acaso (chance), como em Cage; aplicações de fórmulas matemáticas, como em Iannis Xenakis (1922-2001), sendo *Metastaseis* (1953-1954) um caso emblemático;

Ex. 19
https://www.youtube.com/watch?v=SZazYFchLRI

ou graduais não repetitivos, fazendo com que as composições evoluam de um ponto ou estado de partida 'A' ao ponto de destino 'B'. Muitas das peças de Ligeti assim procedem, como é o caso de *Lux Aeterna* (1966).

Ex. 20
https://www.youtube.com/watch?v=Rgg0zh_X-9k

Tanto o espectralismo quanto a nova complexidade irão reivindicar processo como elemento fundamental de suas poéticas. No entanto, haverá diferença no escopo e no grau de atuação destes processos: para os espectrais, sua audibilidade é fundamental, enquanto os compositores da nova complexidade farão uso de intrincadas sobreposições de processos, tornando-os de difícil identificação pelo ouvido. Assim, no espectralismo os processos são aliados e reforçam a noção de direcionalidade do discurso musical e de certa narratividade que se desvela gradualmente. Há evidentes pontos de contato aqui com a música de Ligeti, com os pontos focais de transformação e mudanças de estado graduais. É importante, porém, distinguir estas mudanças graduais do gradualismo minimalista, que, pela mudança de pequeno escopo, enfatiza, sobretudo, a repetição e a *stasis*. Na nova complexidade, diversos processos são sobrepostos em cascatas. Muitas vezes, mais de cinco processos ocorrem simultaneamente, tornando a superfície musical um intricado tecido em contínua transformação, exacerbando os aspectos texturais da música serial.

2. TENDÊNCIAS RECENTES

Todas as tendências mencionadas anteriormente sobrevivem e coexistem no mundo contemporâneo. Gostaríamos de destacar, antes de concluir este texto, duas outras manifestações que vêm marcando a contemporaneidade e com as quais temos guardado uma relação relutante, ora de aproximação, ora de afastamento.

A primeira delas faz da livre-improvisação a base de sua manifestação. Seja ela realizada com recursos puramente instrumentais (muitas vezes enfatizando as chamadas técnicas instrumentais estendidas), puramente eletroacústicos ou mistos, a livre-improvisação tem raízes tanto no mundo clássico (Stockhausen e Cage são exemplos lapidares de compositores que, com interesses e pressupostos muito diferentes, lançaram-se à exploração de obras calcadas na livre improvisação dos intérpretes) quanto no mundo do jazz e da música popular. Os músicos que se lançam hoje à livre-improvisação chamada "não-idiomática" [35] estão imbuídos do desejo de produzir uma música com certo tipo de sonoridade que, guardando pontos de contato com a produção experimental que a antecede, no entanto a ela reagem, sobretudo em relação à Nova Complexidade. Por meios "informais" e distintos, chegam a resultados muitas vezes semelhantes.

Tal é o caso, por exemplo, do projeto de livre-improvisação *The Horrors of Darmstadt* (1994). O título seria referência irônica a um programa da BBC sobre a cidade alemã de Darmstadt que, desde fins dos anos 1940, abriga um curso de férias que foi, por décadas, a Meca da música de vanguarda na Europa.

Ex. 21
https://www.youtube.com/watch?v=-NEMaUiUWco

Destaquemos, ainda, outros exemplos, como o do quarteto de cordas australiano The Noise String Quartet

Ex. 22
http://music.thenoise.com.au/album/spontaneous-improvisation

e o trabalho desenvolvido no Sonic Arts Research Centre da Queen's University de Belfast, como é o caso do projeto *Disparate Bodies v2.0*, exemplo de uma *performance* em rede alimentada por partituras gráficas, com músicos atuando em Belfast, Hamburgo e Graz.

Ex. 23
https://www.youtube.com/watch?v=k7qqCPh0HEw

Sonic Art é um termo cunhado pelo compositor inglês Trevor Wishart (1946-) que, cansado de ouvir algumas de suas peças classificadas como não sendo música, cria uma definição capaz de "abranger as artes de organizar eventos sonoros no tempo." [36] Para Wishart, estas artes incluem, por exemplo, a música, a música eletroacústica, e também *text-sound* [37] ou mesmo os efeitos sonoros. Retornando aos primórdios de nossa narrativa, mencionamos que Varèse, desejoso de deligar-se da noção de música legada pelo passado romântico, preferia usar o termo *Som Organizado*. Quaisquer que sejam as designações que queiramos dar a esta arte profundamente transformada pela (possibilidade de) incorporação de todos aos sons, temos assistido nos últimos anos ao surgimento e à disseminação de formas de expressão sonora que escapam ao domínio da sala de concertos e passam a ocupar as galerias de artes plásticas, por meio de instalações. O termo genérico que as designa é Arte Sonora e a sua história precisa ainda ser escrita.

Limitar-nos-emos, aqui, a mencionar alguns aspectos de um projeto em que estivemos envolvidos no ano de 2014 e que, por suas características, exemplifica muitos dos elementos contidos em outras iniciativas ao redor do mundo. Trata-se do projeto participativo de arte sonora *Som da Maré*, liderado pelo compositor português radicado no Reino Unido, Pedro Rebelo (1972-), que é docente do Sonic Arts Research Centre da Queen's University de Belfast e atuou em 2014 como professor visitante junto à Escola de Música da UFRJ em convênio com a Escola de Belas Artes. O projeto resultou, após quatro meses de oficinas e trabalhos de campo na favela da Maré, em duas intervenções culturais: uma exposição de arte sonora no Museu da Maré, e passeios sonoros guiados no Parque do Flamengo, zona sul do Rio de Janeiro. A exposição incluiu três instalações que refletem as vivências sonoras, o cotidiano e as reflexões sobre o futuro de moradores da Maré e utilizam recursos sonoros, visuais e interativos que ocuparam três galerias no Museu da Maré. Já nos passeios sonoros, o público era guiado por membros da Cia Marginal, grupo teatral da Maré, fazendo uso de *headphones* por onde se ouviam depoimentos, ambientes sonoros e gravações de campo, em contraponto com as intervenções dos atores/guias, percorrendo diversos ambientes do parque. Para mais informações sobre o projeto e exemplos audiovisuais, ver https://somdamare.wordpress.com/

Ainda que seja perigoso generalizar em virtude da variedade de situações possíveis em projetos de arte sonora e instalações, nestas obras o som é, mais que um fim em si mesmo, um meio que, ao lado de outros elementos, coloca em valor uma percepção diferente do espaço e do tempo, muitas vezes de forma dinâmica e não linear. Numa instalação, o público tem em geral liberdade de mover-se, de aproximar-se ou afastar-se das fontes sonoras, entrar ou sair do espaço de exposição e mesmo emitir sons, sem que isso seja considerado um problema. Nada mais distante de uma sala convencional de concertos, onde a emissão de ruídos e a movimentação do público enquanto os músicos estão tocando são consideradas lesivas ao espetáculo!

3. CONSIDERAÇÕES FINAIS

Já definimos referenciais teóricos "como o conjunto de conceitos que amparam, orientam ou dialogam com os compositores, auxiliando-os na criação das condições de possibilidade de suas poéticas particulares, estando assim adstritos àquilo que classicamente se tem denominado de nível *poiético*." [38] Há, no momento da criação, contato com conceitos que amparam, dialogam e alimentam o ímpeto criativo. Os instrumentais técnico e tecnológico, além de motivações psicológicas e emotivas, apresentam-se como constituintes do nível *poiético*. Desta forma, os referenciais teóricos apresentam-se como alimentadores da *poiesis*, embora não sejam os seus únicos motivadores. São como âncoras conceituais que nos permitem fazer.

Estes referenciais não precisam refletir convicções universalmente aceitas e compartilhadas. Não importa, por exemplo, se o conceito que os compositores têm do que é som está correto e reflete o conhecimento científico consolidado de sua época. Um artista não é um cientista produzindo experimentos, ainda que muito da música do século XX tenha buscado na ciência não só uma fonte de inspiração, mas também um modelo de organização. Para Schaeffer, isto apenas reflete a crise de uma arte que, insegura de seus pressupostos e não encontrando em si mesma o motor de sua própria existência, buscará em outras disciplinas o rigor e a confiança perdidos, tomando de empréstimo seus procedimentos. [39]

Em nosso breve e incompleto panorama, utilizamos uma série de obras musicais que exemplificam diferentes concepções de som. Indo da música instrumental pura, passando pelas músicas eletroacústicas e chegando à arte sonora, em que a própria categoria "música" é colocada em cheque, este panorama busca evidenciar que o que se entende como som e como som musical tem variado de acordo com a época, as filiações estéticas dos criadores e por seus referenciais teóricos, deixando marcas profundas nas músicas de nosso tempo.

REFERÊNCIAS BIBLIOGRÁFICAS E NOTAS

[1] Griffiths, Paul. *A Música Moderna, uma história concisa de Debussy a Boulez*. Rio de Janeiro: Zahar. 1987. p. 97

[2] Moraes, J. Jota de. *Música da Modernidade, origens da música de nosso tempo*. São Paulo: Brasiliense. 1983. p. 132

[3] Dufourt, Hugues. Hauteur et timbre. In Marc Jimenez (ed.). *InHarmoniques 3, Musique et Perception*. Paris: IRCAM. 1988. p. 46

[4] Para uma descrição técnica, ver: Pierce, John R. *The Science of Musical Sound*. New York: W. H. Freeman. 1992. pp. 60 e 61

[5] Dufourt 1988, op. cit., p. 61

[6] Risset, jean-Claude e Wessel, David. Exploration du timbre par analyse et synthèse. In Barrière , Jean-Baptiste (ed.) : *Le timbre, métaphore pour la compostion*. Paris : Christian Bourgois Éditeur. 1991. p. 104

[7] Risset citado em Dufourt 1988, op. cit., p. 67

[8] Moraes 1983, op. cit., p. 131

[9] Para a tradução brasileira da carta, ver: Russolo, Luigi. A Arte dos Ruídos, Manifesto Futurista. In Menezes, Flo. *Música Eletroacústica: história e estéticas*. São Paulo: Edusp. 1996. pp. 51 a 55

[10] http://www.rossanamaggia.it/intonarumori-setF.html

[11] Russolo in Menezes 1996, op. cit., p. 54

[12] Stravinsky, Igor. *Poetics of Music in the form of six lessons*. Cambridge: Harvard University Press. 1994. pp. 63 e 64

[13] Adorno, Theodor W. *Philosophie der neun Musik*. Frankfurt: Europaische Verlagsanstalt GmbH. 1958. Edição brasileira: Adorno, Theodor W. *Filosofia da Nova Música*. São Paulo: Perspectiva. 1974

[14] Schoenberg produziu farta obra teórica. Para uma discussão do próprio autor sobre o método dodecafônico, ver Schoenberg, Arnold. 'Composition with Twelve Tones'. 1941. in Stein, Leonard (ed.). *Style and idea, selected writings of Arnold Schoenberg*. London: Faber and Faber. 1975. pp. 214 a 245

[15] Velloso, Rodrigo Cicchelli. A Permanência das Polarizações. *Humanidades*, n. 58. Brasília: UnB. 2011. pp. 22-28

[16] Schaeffer, Pierre. *Traité des objets musicaux*. Paris: Éditions du Seuil. 1966.

[17] Hodgkinson, Tim. 'Pierre Schaeffer interviewd by Tim Hodgkinson'. In Chris Cutler (ed.). *ReRecords Quarterly Magazine*. Volume 2 no 1. London: November Books. 1987.

[18] Stockhausen, Karlheinz. 'Da Situação do *Metier*: Composição do Som (*Klangkomposition*)'. 1953. In Menezes, Flo. *Música Eletroacústica: história e estéticas*. São Paulo: Edusp. 1996. pp. 59 a 72

[19] Krenek, Ernst. 'O que é e como surge a Música Eletrônica?'. 1956. . In Menezes, Flo. *Música Eletroacústica: história e estéticas*. São Paulo: Edusp. 1996. pp. 97 a 104

[20] Schaeffer, Pierre. *Traité des objets musicaux*. Paris: Éditions du Seuil. 1966

[21] A ironia de Schaeffer faz referência às fontes sonoras preferenciais dos compositores de música concreta (sons pré-gravados de origem ambiental e instrumental, que, isolados, transformados e editados por meios eletroacústicos, tornam-se "cadáveres" sendo "dissecados") e dos compositores de música eletrônica (sons de síntese senoidal, que, superpostos a partir de critérios seriais, tornam-se os tais "robôs").

[22] Schaeffer 1966, op. cit., p. 61

[23] Para uma descrição aprofundada dos procedimentos empregados na criação de suas composições eletrônicas dos anos 1950, ver: Stockhausen, Karlheinz. *Elektronische Musik 1952-1960*. Kürten: Stockhausen Verlag. 2001.

[24] Duchez. Marie-Elisabeth. 'L'évolution scientifique de la notion de matériau musical'. In Barrière, Jean-Baptiste (ed.). *Le timbre, métaphore pour la composition*. Paris : Christian Bougois Éditeur. 1991. pp. 50 e 51

[25] Risset, Jean-Claude. 'An Introductory Catalogue of Computer Synthesized Sounds'. 1969. In Goebel, Johannes (prod.). *The Historical CD of Digital Sound Synthesis. Computer Music Currents 13*. Mainz: Wergo. 1995. pp. 109 a 254

[26] Velloso, Rodrigo Cicchelli. A Permanência das Polarizações. *Humanidades*, n. 58. Brasília: UnB. 2011. pp. 22-28

[27] Fineberg, Joshua. 'Spectral music'. *Contemporary Music Review*. 19: 2, 1-5. London: Routledge. 2000.

[28] Anderson, Julian. 'A Provisional History of Spectral Music'. *Contemporary Music Review*, 19: 2, 7-22. London: Routledge. 2000.

[29] Gérard Grisey citado em Hurel, Philippe. 'Le phénomène sonore, un modèle pour la composition'. in Barrière , Jean-Baptiste (ed.). *Le timbre, Métaphore pour la composition*. Paris: Christian Bourgois Éditeur. 1991. p.261

[30] Ver Bonnet, Antoine. 'La part de l'insaisissable'. in Barrière, Jean-Baptiste (ed.). *Le timbre, Métaphore pour la composition*. Paris: Christian Bourgois Éditeur. 1991.pp. 338 e 339

[31] Boulez, Pierre. 'Entre ordre et chaos'. in Marc Jimenez (ed.) *InHarmoniques 3, Musique et perception*. Paris: IRCAM. 1988. p.112

[32] Ibid. p.112

[33] Fox, Christopher. 'New Complexity'. *The New Grove Dictionary of Music and Musicians*. Second edition. Edited by Stanley Sadie and John Tyrrell. London: Macmillan Publishers. 2001.

[34] Reich, Steve. *Writings about Music*. Halifax: Press of the Nova Scotia College of Art and Design. 1974.
[35] Por "não-idiomático" os praticantes entendem ser um tipo de livre-improvisação que evita procedimentos "idiomáticos" jazzísticos ou de outras práticas musicais. No entanto, o "não-idiomatismo" acaba configurando um novo idioamtismo, pelo recurso a determinadas técnicas de execução instrumental (técnicas estendidas, *scribling*, recusa de padrões, etc.).
[36] Wishart, Trevor. *On Sonic Art*. Amsterdam: Harwood Academic Publishers. 1996. p. 4
[37] Não há uma boa tradução desta expressão, que visa designar uma forma de expressão sonora que explora o texto e a vocalidade em sentido alargado.
[38] Velloso, Rodrigo Cicchelli et alii. 'Referenciais Teóricos da Música Eletroacústica Brasileira Contemporânea: acerca de um questionário'. In: CONGRESSO DA ASSOCIAÇÃO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM MÚSICA, XIX, 2009, Curitiba. *Anais*. Curitiba: ANPPOM. 2009. pp. 501-504.
[39] Hodgkinson, Tim. 'Pierre Schaeffer interviewd by Tim Hodgkinson'. In Chris Cutler (ed.). *ReRecords Quarterly Magazine*. Volume 2 no 1. London: November Books. 1987.

MINI CURSO

09:00 - 01:00pm
Sala 1
Room 1

Quinta-feira
Thursday
Maio / May
28, 2015

ARRANJO DE MICROFONES UTILIZANDO ARDUINO

13º CONGRESSO DE ENGENHARIA DE ÁUDIO

THIAGO DE MOURA PREGO
CEFET - RJ

MINI CURSO

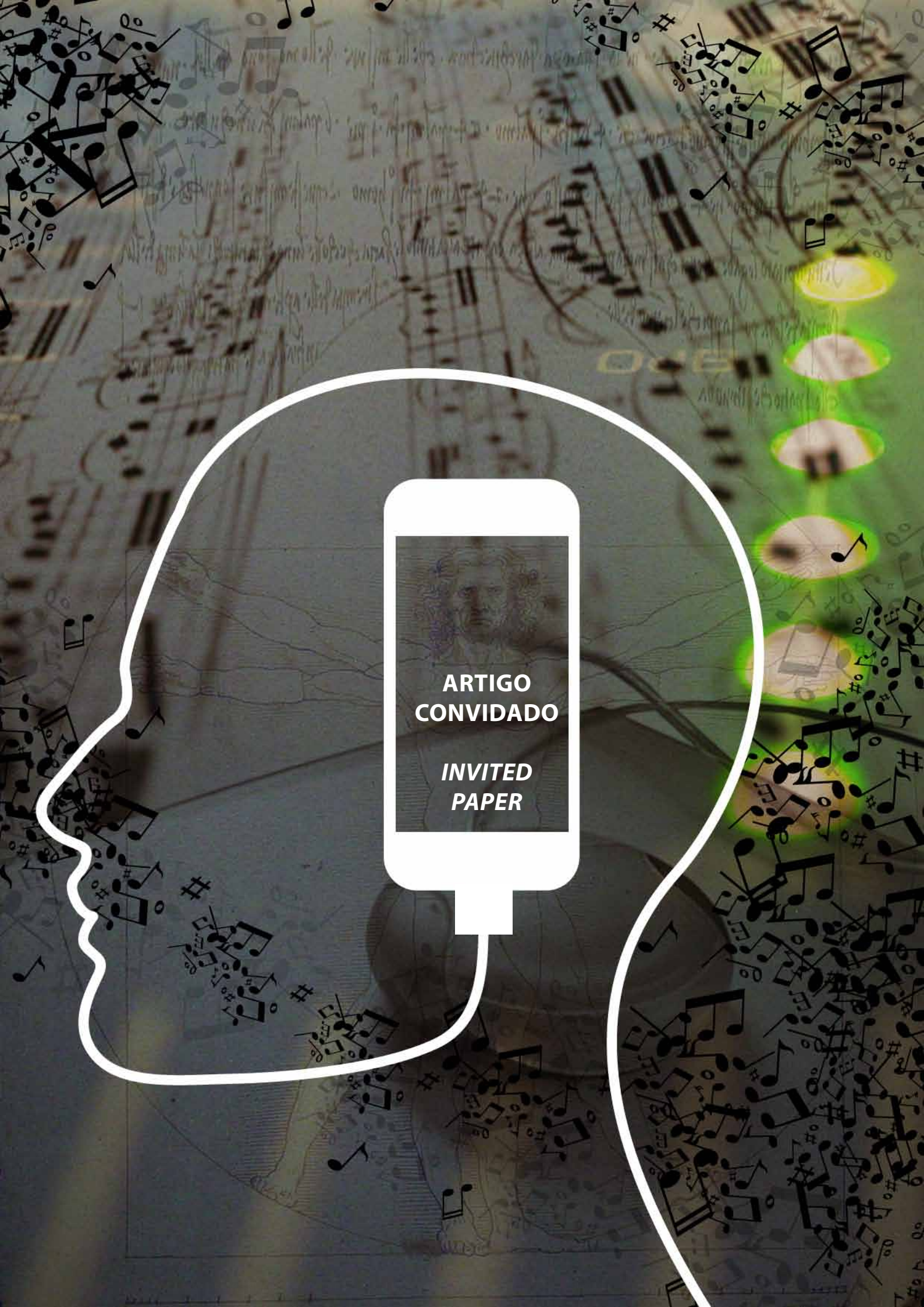
02:00 - 06:00pm
Auditório A
Lecture Room A

Quinta-feira
Thursday
Maio / May
28, 2015

DESENVOLVIMENTO DE APLICATIVOS PARA DISPOSITIVOS MÓVEIS NA PLATAFORMA ANDROID

13º CONGRESSO DE ENGENHARIA DE ÁUDIO

FRANCISCO HENRIQUE DE FREITAS VIANNA - CEFET -RJ


ARTIGO
CONVIDADO
INVITED
PAPER

Sociedade de Engenharia de Áudio
Artigo de Congresso

Apresentado no 13º Congresso de Engenharia de Áudio
19ª Convenção Nacional da AES Brasil
25 a 28 de Maio de 2015, 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.

Study on Spatial Coherence of Moving Audio-Visual Objects

Thomas Sporer,[1] Judith Liebetrau,[1] David Goecke,[1,2] and Karlheinz Brandenburg[1,2]

[1] Fraunhofer IDMT, Ehrenbergstr. 31, 98693 Ilmenau, Germany

[2] Technische Universität Ilmenau, Ehrenbergstraße 29, 98693 Ilmenau,Germany

spo@idmt.fraunhofer.de, ltu@idmt.fraunhofer.de, davidgoecke@gmx.de, bdg@idmt.fraunhofer.de

ABSTRACT

In media technology video and audio coding are often applied together. The quality of experience (QoE) in such audio-visual systems is influenced by several factors. Besides the dominant factors pure audio and video quality, the temporal coherence of the auditory and visual events plays a role. For common two-dimensional applications theses factors are well investigated, understood and most data formats in coding and reproduction system are optimized to meet the perceptual requirements. More sophisticated media technology systems enable the reproduction of 3D audio and video. In such systems the spatial audiovisual coherence is of significant interest. While spatial coherence is well understood when audio and video are on fixed positions, hardly no studies exist for moving objects.

In this paper perceptual tests are presented trying to investigate the perceptual threshold for moving audio-visual objects in the horizontal plane. An offset of less than 2° was detectable for trained expert listeners. For these the result is independent from the sound stimulus used (speech, pink noise) and from the offset direction (audio in front of video versus video in front of audio). However a small difference of the threshold dependent on the direction of movement (left-to-right versus right-to-left) was observed.

0 Introduction

Audio coding is often used together with video coding. The quality of experience in audio-visual systems is dependent on a number of factors. The dominant factors are pure audio quality and pure visual quality. Another important factor is temporal coherence of auditory and visual events. All these factors are well understood and most data formats in coding and reproduction systems are optimized to meet the perceptual requirements.

In the past audio reproduction has been based on

only a small number of loudspeakers for the front (”on screen”). On such systems audio objects usually are rather wide, and the overall audio image is rather fuzzy.

Newer reproduction systems are using more loudspeakers and provide a more defined localization of audio objects. While spatial coherence of audio and video for static sources is well understood hardly no studies exist for moving objects. Especially such moving objects are important in object-based audio coding schemes like MPEG-H: the rate of position updates must be high enough to enable smooth movements and inaudibility of position miss-match between audio and video but should be as low as possible to reduce bit-rate.

In the first chapter the major effects for spatial perception of audio and video, and their relation are summarized. In addition some background information about spatial sound formats and their capabilities to provide proper spatial information are discussed. Chapter 2 explains the test design including test method, stimuli and room acoustics. In chapter 3 the results of the experiments are presented.

1 State of the Art

The research concerning spatial coherence of moving audiovisual stimuli and its application is an interdisciplinary task. Several research fields have to be taken into consideration. In the following the basics of auditory, visual and audiovisual perception will be explained. An overview about related research, targeting on the identification of a perceptual threshold between audio and visual stimuli, is presented. Finally, knowledge about object oriented audio reproduction as well as insight into the new MPEG-H standard, applicable for 3D video and audio compression, will be given.

1.1 Perception

1.1.1 Hearing

In order to introduce spatial perception, we first introduce some basics about the working principles of the ears and the brain.

Listening is influenced by

- The form of the head, torso and pinna
- The middle and inner ear with the ossicles and the cochlea
- The hair cells leading to nerve firing

All the influencing factors, potential as well as limits of the auditory system can be mainly explained by mechanical models. Masking effects, auditory thresholds and such are well understood. [1]

To understand how we perceive sound in complex environments including our abilities to detect the position of a sound source, we focus at monaural and binaural effects.

Monaural spatial hearing can be explained via the so called Head Related Transfer Functions (HRTF). Depending on the incoming elevation angle and direction of sound, we can use a frequency dependent function as a model for the overall transfer function from the source to the ear drum. This function is very much dependent on the individual pinna and head geometry. As humans, we learn how sounds get modified when coming from different directions and use this to estimate the direction.

The better known cues to derive the position of a sound source depend on binaural processing in the brain. The signals from both ears arrive in a common region in the brain. Again, depending on frequencies, our brain derives the most important cues from either time (or phase) differences between the signal at both ears or from level differences. These binaural cues are called ITD (Interaural Time Differences) and ILD (Interaural Level Differences). To estimate the incident angle of a sound source ITDs are more important at lower frequencies and ILDs at higher frequencies. There is an abundance of literature explaining spatial hearing in much more detail, see e.g. [1].

Localization accuracy is defined as the smallest difference in angle or position of a sound source which can be perceived as coming from a different direction. Over the years there have been many publications on localization accuracy in hearing, mainly for static (non moving) sound sources, see e.g. [2, 3]. It depends on the position of the sound source. The smallest perceived differences are smaller for frontal sound sources and larger for side or even back locations. Looking into more details, it has been noted that localization accuracy becomes better (smaller) if the sound is known to the listener and if the listener is allowed to move the head. The latter observation can at least in part be traced to the information gathered from different sounds arriving at the ear drum when moving the head (see the explanation of HRTF).

To formalize the knowledge about localization accuracy, we use the terms MAA (Minimum Audible Angle) and MAMA (Minimum Audible Movement Angle, see [2]). Earlier research suggests that the MAMA for frontal and slowly moving sources is nearly equal to the MAA while it becomes larger (less accuracy) for other directions and increasing velocity of the sound source [4].

1.1.2 Visual Perception

The processing and interpretation of information that is contained in visible light can be roughly defined as visual perception. The visual system contains of different parts: Seeing starts when the lens of the eye focuses an image of its surroundings onto the retina. This light-sensitive membrane in the back of the eye converts patterns of light into neuronal signals. These neural impulses are processed in a hierarchical fashion by different parts of the brain. Varying kinds of signal pro-

cessing in the brain are involved in the recognition of objects or motion perception. As the latter is of interest for this paper, it will be explained shortly.

The estimation of the speed and direction of objects in a scene is defined as motion perception. First research concerning motion perception started in the late 18th [5] and early 19th [6] century. The persistence of vision was found to be a reason for motion perception. These findings were improved by Wertheimer [7] in 1912. He described 'beta movement' as an optical illusion, where a series of static images create an impression of continuous motion at a frame rate of 10-12 single images per second. The phi phenomenon is based on the same findings. In contrast to the beta movement, the phi phenomenon is caused by luminous impulses in sequence. Beta movement results from luminous stationary impulses. It can be concluded that first-order motion perception occurs, when a change in luminance at one point on the retina is detected and correlated with a change in luminance at a neighbouring point on the retina after a short delay.

1.1.3 Audiovisual (AV) Perception

Human vision perception is influenced by auditory perception and vice versa. A well known effect for perceptual fusion of audio and visual stimuli is the ventriloquism-effect [8, 9]. A visual cue is used to distract the location of an simultaneous audio stimulus off from its origin. In cases where the local discrepancy is large enough, both stimuli will be perceived as two discrete sources. Decreasing the AV offset results in an increasing influence of the visual stimuli. It pulls the audio stimulus until a certain threshold where perceptual fusion is reached. In this case both stimuli will be perceived as a single one. Lewald and Guski [10] showed that fusion can be reached if a maximum AV offset of 12° is present. The thresholds of perceptual fusion and localisation accuracy of audio and video stimulus is dependent on several factors. Besides the spatial difference between the test items, also time alignment plays an important role.

1.2 Studies Concerning AV Perception and Fusion

Several studies were conducted in the past to quantify influencing factors as well as AV perception thresholds. That the localisation accuracy of sound sources is influenced by the extent of the AV-offset is one major finding. In their study Bertelson and Radeau [8] used loudspeakers and red flashing LEDs as stimuli, which were placed in the horizontal plane. The location of the stimuli were rated through hand pointing on a transparent screen, where a degree graduation was printed on. The observations showed that the audio stimuli localization accuracy can be changed about 4° for 7° AV offset, 6.3° for 15° and 8.2° for 25° difference between both stimuli. Also they mentioned that the audio stimuli localisation is more distracted by the visual stim- 0.6°). Besides the spatial AV offset of simultaneous stimuli, the time offset plays a role. Perceptual fusion decreases with larger time difference [11, 12]. An AV offset of 100 ms results in 73% fusion, 200 ms in 38% and 300 ms only in 7%. Fusion is more often reached when the visual stimulus preceded the audio stimulus. Investigation with moving visual and static audio stimuli were conducted by Lewald, Ehrenstein and Guski [13]. Periodic short pulses of sine tones and continuous 1 kHz tones were used in addition to a projected red light spot which moved on the screen in the horizontal plane. Assessors were asked to press a key as soon as the sound and the laser spot were perceived to be in spatial alignment. They found out that a spatial AV offset of 3° maximum results in fusion for a time difference of 100 ms. The influence of the distance between the assessor and the stimuli on the fusion perception was investigated by Cote [14] and Corrigan [15]. Cote used a virtual environment to ask test participants to localize virtual AV objects in distance. Female and male speech were reproduced through headphone (binaural rendering) as stimuli. A virtual loudspeaker was visualized on a stereoscopic passive screen, 2m in front of the observer. Assessors were asked to report their egocentric distance judgment by using a keypad. For an observer distance of 2m fusion is reached for depth offset between 1 m - 2 m. These results were supported by Corrigan. The differences in depth between audio and visual stimuli in stereoscopic-3D environments was assessed here by presenting pink noise and female speech over headphones using binaural renderings and a stereoscopic display. The stimuli were generated at different depths from the viewer. The assessor rated if the audio stimuli were nearer, further away or at the same depth as the visual stimuli. The study indicated that the range of fusion increases as the depth of the visual stimulus increases.

Several experiments showed that the end-positions of both, auditory and visual event, are perceived distorted. The representational momentum for visual objects was studied by Fröhlich [16], whereas Getzmann [10] did investigations in the case of auditory events. In his study the end position of moving pink noise had to be indicated via hand pointing. In further investigations [17], [18] short noise pulses played randomly at a peripheral or a central azimuthal position and moved from left to right or vice versa along the frontal horizontal plane. A 2-AFC method was used to compared the onset position of the target to a visual reference stimulus positioned left or right of the auditory onset position. Result showed that the onset positions of moving targets were distracted in the direction of motion. When the visual reference stimulus was presented after the auditory motion, the largest displacement occurred. In [19] the localization of a moving sound at different points in time was observed. Applying either a relative-judgment or pointing task, the assessors indicated the

dition a moving visual stimulus (red LEDs) was presented either in-line with the auditory event or shifted to the left or right. Depending on the location of the stimuli their position was perceived to be displaced in a range of -1° to 4°. Getzmann concluded that in spatial hearing the representational momentum is related to the final point of motion. A definite threshold was not obtained.

Kunze investigated [20] spatial congruency of 3D audio and video objects. 3D video material of a jazz concert was reproduced via stereoscopic display while the according audio content was played back on a wave field syntheses system. Different degrees of AV offsets were introduced. Absolute category rating was applied to identify the horizontal displacement angle for the AV offset. The results showed that a displacements of more than 5° were noticeable and annoying above 10°. A similar experiment was conducted by André [21]. A video of a person was presented together with auditive stimuli, which had different AV offsets. The assessors had to identify the displacement. Results showed a larger threshold for audiovisual fusion (19°).

1.3 Fileformats

Modern reproduction approaches for realistic spatial audio and video, promise more realistic audiovisual experience for the recipient. This includes proper perception of direction, distance, and elevation of the audio and video objects as well as their correct spatial and temporal alignment. Although manifold research was conducted to enable realistic audiovisual experience, there is still place for improvement on the production and reproduction site.

To enable sound scene reproduction of spatial audio as intended during production, the position of the loudspeakers at the reporduction site must be known. There are two principal approaches for the sound engineer: either mixing content in a standardized environment with a well defined position (and properties) of loudspeakers, or he produces the mix in a virtual sound scene, and the mapping of audio to loudspeakers is done via a renderer at the reproduction site.

In the first approach, which usually is called the channel-based approach, it is necessary that in each reproduction room the loudspeakers have to be exactly at the predefined positions as during the mixing. In the second approach the number and position of loudspeakers are flexible on the reproduction site. The adaptation of the mix to the loudspeaker is done through sophisticated calculation (rendering). This of course requires a higher amount of data, which have to be transferred to the reproduction site and the rendering device. There are different alternatives for the storage of the content and additional data: In the **object-based** paradigm the audio scene is decomposed into audio objects. Each audio object is defined by the audio essence (e.g. a pcm time series) and some metadata describing position and properties of the object. Rendering algorithms for object based might be vector based amplitude panning (VBAP) [22] or wave field synthesis (WFS) [23]. VBAP and WFS calculate the driving signals for each loudspeaker dependent on source position and loudspeaker positions. In general every sound object is reproduced by a multitude of loudspeakers. If a sufficient number of loudspeakers is available the perceived position of an audio object is constant in the whole reproduction space. There is even the possibility to place sound objects in front of the loudspeakers. Due to the fact that always many loudspeakers contribute to the overall sound field, no change of coloration appear with different position of a sound object. If sound objects are narrow (or focussed) an increase of localization accuracy could be observed.

Another paradigm is called higher order ambisonics (HOA) [24]. In HOA the target sound-field is described by spherical harmonics at a reference positions. With HOA these components are mapped to all available loudspeakers at the reproduction site.

Channel-based formats with all loudspeakers in the horizontal plane, like mono, stereo and 5.1 have been used for quite some time. Improved formats like Auro3D use in addition elevated loudspeakers. Nevertheless, all of these formats use a small number of loudspeakers and therefore the created soundfield is not like a natural soundfield: Positions of sound objects between loudspeakers are only possible with the concept of phantom sources. From a perceptual point of view, phantom sources have three major disadvantages:

- They are rather wide, meaning not having a well-defined spatial position.
- They are only stable if the listener is near the reference position of the mix.
- They have a different sound coloration compared to the direct loudspeaker signals.

The region where phantom sources work quite well is called the sweet-spot. The size of the sweet-spot gets larger when using more loudspeakers, but in the channel-based paradigm these additional loudspeakers must be foreseen by the sound engineer during production. Due to width of phantom sources it was not necessary to analyze the miss-match of visual and auditory position in detail for channel based reproduction. However, the position of loudspeakers had been optimized in perceptual experiments to provide plausible localization of audio stimuli on the screen. In music production the center channel is often not used because of different coloration compared to sound directions between left and right speaker. In cinema production the dialogs are often mixed only to the center to avoid unwantes "out-of-sweet-spot" distortions.

Due to it's advantages and flexibility the object-based paradigm is used in new formats like IOSONO [25], MPEG-H 3D-Audio [26, 27] and Dolby Atmos

[28]. All these formats can also transport channel-based content or a combination of channel-based content enriched by objects. This enables sound engineers to use their traditional work-flow in production for recording and mixing. For objects the position is encoded in the metadata.

In MPEG-H also HOA content can be used, either alone or in combination with objects and/or channels. Internally the HOA content is decomposed into "predominant sources" and ambiance sound. The predominant sounds are in fact very similar to objects in the object-based paradigm. The ambiance sound is the residual when the predominance is subtracted from the HOA data. Usually it is sufficient to encode this residual signal using zero and first order coefficients.

MPEG-H is a standardized format where internal details, like coding of object positions, are openly available. IOSONO and Atmos are proprietary formats, and therefore such information is not accessible to the public. In MPEG-H the positions of objects are updated at a fixed rate. The reproduction usually uses interpolation between such positions. The precision of the transmitted azimuth is $1.5°$ and of the elevation is $3°$. The rate of position updates is adapted by the encoder depending on the speed of moving audio objects. The fastest possible update rate is 64 samples, which is equivalent to 1.3 ms at a sampling rate of 48 kHz.

Although the technical parameters for AV content coding, transportation and reproduction are already standardized, an improvement of signal processing based on findings concerning perception thresholds would be beneficial. The consideration of threshold and perceptual effects could lead to an simplification of coding as well as lower cost at the reproduction site. Therefore further experiments to detect useful thresholds have to be performed.

2 Test Design

As shown in the state of the art, different investigations were performed to investigate the perceptual AV offset, but the issue of perceptual thresholds for concurrent movement of audio and video has not been cover in detail. Within our work we try to improve the investigations done by [19] and [21]. The focus of our work lays therefore on the synchronous horizontal movement of audio and visual stimuli on a silver screen. It will be investigated, if a minimum perceptive movement distance (MPMD) between audio and video stimuli can be observed.

2.1 Test Method

As the assessment aims on the detection of AV perception thresholds, a method for threshold estimation should be applied. One representative of an adaptive method is the N-alternative forced choice test (N-AFC) as described by Békésy [29]. The assessor is provided simultaneously with n-stimuli in random order. Only one stimulus changes its characteristic across trials, whereas the others are fixed. The difference between the fixed and variable stimuli ranges from below to above the expected detection or discrimination threshold. Assessors are simply asked to detect the stimulus with different characteristics. This way the discrimination threshold can be detected which represents the minimum perceivable difference between two stimuli. The test method is usually quite powerful and reliable. Due to the long duration of the assessment task lapses of attention can occur. Assessors might give the wrong response to trials that are largely over threshold due to fatigue. A modified version of N-AFC is the 3-AFC test with a 2-down, 1-up approach. A test participant is presented with 3 stimuli, where only one has a different characteristic. The assessor has to indicate the differing stimulus. If the right stimulus was selected, the same stimuli combination is presented again but in random order. When the right stimulus was identified once again, another stimuli combination is presented. The characteristic of the variable stimulus is changed towards the direction of the assumed threshold (2-down). If at any time the wrong stimulus is selected, the varying stimulus is changed contrary to the assumed threshold (1-up). This procedure is started with a clearly audible difference between stimuli. It approaches the threshold, and after reaching the threshold it starts to oscillate around the perception threshold. The threshold is calculated from the last six inflection points at least. In a 3-AFC-2-down test the probability of two correct answers in a row while guessing (below threshold) is 9%. Using a N-AFC test with larger N would reduce the probability of correct answers below the threshold, but also increase the time for each trial. N equal 3 is therefore a good trade-off between test accuracy and time effort.

2.1.1 Test Procedure

The test method was slightly adopted to shorten the time duration. In a first step the change of the AV offset was large and was reduced after a certain amount of ratings. This should enable a fast and rough estimation of the threshold but also a fine adjustment. The assessment started with a horizontal AV-offset of 27 cm. The offset is changed by 5 cm for the next eight double-right decisions. The last AV offset is now assumed o be the starting point of the fine adjustment. In case of double correct identification the offset is reduced about 1.086 cm. If the wrong stimulus was selected, the offset is raised about 1.086 cm. After eight double correct identification the assessment tasked stopped.

The test is organized in two sessions. In the first session (experiment 1) the AV offset should be evaluated when the audio follows the visual stimulus. In the second session (experiment 2) the audio stimulus preceded the video stimulus. To avoid fatigue of the assessors, both session where performed on different days. The order of the experiments stayed constant for all asses-

sors.

2.1.2 Rating Device

As rating device a graphical user interface was implemented. The GUI consisted of three elements: A presentation area where the stimuli were displayed, three Buttons for selection of the test stimuli and a counter of correct ratings. With the help of a computer mouse the button had to be selected, which represents the stimulus with the differing characteristic compared to both other stimuli.

2.2 Test Stimuli

For the experiments to investigate the AV offset with moving audiovisual sources simple stimuli were selected. To avoid content based influences the AV-stimuli were meaningless. One criterion while choosing the auditory stimulus was how accurate it can be localized. Speech was chosen because of its familiarity to a listener in generally. Humans are familiar to localize speech in everyday life situations and are able to concentrate on one speech out of a group (Cocktail-Party Effect [30]). In earlier investigations pink noise (120 Hz - 18 kHz) was used as auditory stimulus. To compare the results and to use more features of interaural perception this broadband noise can be used. Both auditory stimuli (meaningless speech, pink noise) for experiments in terms of AV-Coherence are usable [31]. The visual stimulus is a white point with 10 cm diameter (2D). That way influences of color or depth on the MPMD could be avoided.

2.3 Test Room and Equipment

The experiments were conducted in the 3D presentation room of Fraunhofer IDMT. Besides its size this room meets all requirements for listening rooms as defined in Recommendation ITU-R BS.1116 [32]. The properties of the room are: size 78 m^2, volume 380 m^3, noise level NR20, and reverberation time 0.32 s. For audio reproduction a loudspeaker wall with 504 individually controlled loudspeakers (Scheck Audio, LSTG9) was used. The loudspeakers are controlled by time-aligned DA converters including an equalizer for each loudspeaker. The front of each loudspeaker is 10,2 cm by 10,2 cm leading to a total size of the loudspeaker wall of 3.6 m by 1.4 m. Distance of listener from the loudspeaker was 2 m. Due to the close distance to the wall there would be a perceivable change in sound coloration between the loudspeakers in center and near the edge of the wall. Therefore the frequency response of each loudspeaker at the listener position was measured with an omni microphone and compensated by an equalizer.

For visual presentation an acoustical transparent silver screen was placed in front of the loudspeakers. The size of the screen is 4 m by 3.2 m so that the loudspeakers were covered and invisible to the listeners. The projector (Christie Mirage S+6K) was operated at 1200 by 1050 pixels at a frame rate of 60 Hz. To avoid disturbance due to the noise of the projector, the device was placed in the neighboring room and the projection was done through a glass window.

Time-alignment of audio and video system was done via time-code and a common word clock. The audio and video reproduction systems were calibrated in a manner that the size of one loudspeaker is equivalent to 58 pixels of the projected image. The visual stimulus was moved pixel by pixel (no interpolation). For the interpolation of audio object position a truncated Gaussian distribution of levels was used. In all time instances exactly five loudspeakers were active (relative levels ranging from -30 dB to 0 dB). The update of levels was done sample by sample. This algorithm was chosen because it provides a constant total sound pressure level (deviation less than 0.1 dB) and no audible sound colorations.

3 Results

Before the real tests started, an informal pretest was conducted to identify possible weaknesses in test procedure, setup or playback. Four assessors participated in this phase. The key results can be summarized as follows:

- No problems with the test setup and playback device in particular were observed. The data collection and automatic storage of results was successful. Therefore the reproduction and signal processing chain remained unchanged.
- The assessors had sometimes trouble with the handling of the GUI. Some also complained about the visual distraction through the input device (mouse pointer). The graphical user interface was redesigned to avoid visual bias.
- As the test stimuli and the maximum AV offset were unknown to the participants of this pretest, a high error rate was observed. To enable familiarisation with the rating device as well as the stimuli under test, a training session was designed and had to be conducted before each experiment.

3.1 Experiment 1

In this test session the localization accuracy of moving audiovisual stimuli was assessed, where the audio stimulus run after the visual stimuli. The test panel consisted of 22 assessors (8 female, 14 male) in total with a mean age of 24.7 years. In principal both directions of the movement (left to right (LR) and vice versa) had to be evaluated by all test participants, but due to time constraints some assessors assessed only one direction. Eight assessors could be classified as experienced listeners, whereas the others can be considered as naïve listeners. The majority of test participants reported to have normal hearing. One assessor indicated slightly impaired hearing.

The results of the assessment are drawn into an coordinate system: On the x-axis the duration for the judgment is displayed and the y-axis represents the rated offset between audio and video stimuli. Red crosses indicate the infliction points of ratings. In figure 1 the individual rating curves for the test stimulus 'speech' are displayed. Figure 2 shows the individual offset curves in the case of 'pink noise'. Similar graphs were prepared for the direction RL to check the consistency of results. Due to the fact that they conform mainly to the tendencies displayed in LR, they are not printed here. Different observation are explained if necessary.

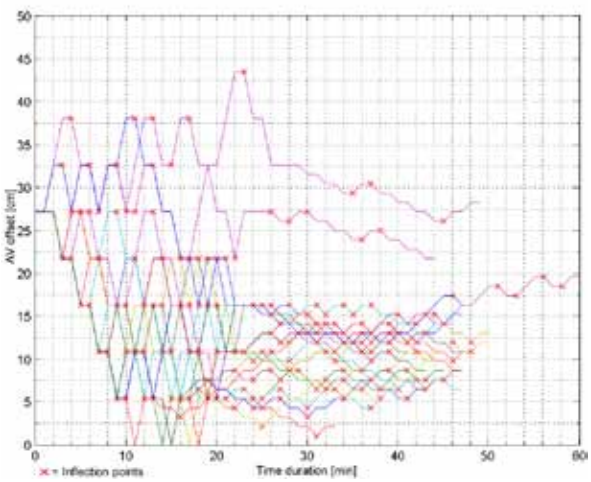


Figure 1: AV offset curves of 21 assessors for speech stimulus behind the visual stimulus, moving from left to right (LR)

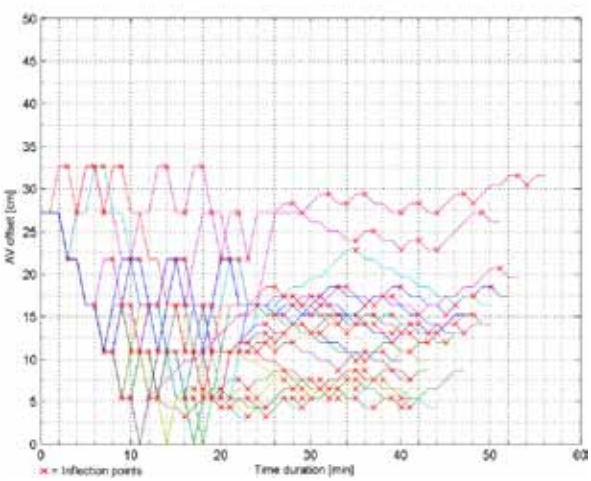


Figure 2: AV offset curves of 22 assessors for pink noise stimulus behind the visual stimulus, moving from left to right (LR)

For the calculation of the perceptual offset threshold, or as we call it here Minimum Perceptive Movement Distance (MPMD), between audio and visual stimuli, the last six inflection points of each individual curve are taken into consideration. Before averaging the individual MPMDs, outliers have to be detected first. Figure 1 indicates that at least two MPMD are different than the others. A post-screening method was applied to prove this assumption. The median of the individual MPMDs were calculated as well as the lower (Q_1) quartiles, upper quartiles (Q_3) and the interquartile ranges (IQR). Data was regarded as outlier when a MPMD falls below $Q_1 - 1.5(IQR)$ or above $Q_3 + 1.5(IQR)$. In case of the speech stimuli the data of two assessors (05, 19) were rejected. In case of pink noise data from two test participants (05 and 12) were not considered for further analysis. In figure 3 the averaged perceptual thresholds for both test stimuli are presented. In addition the standard deviation and the upper and lower 95% confidence intervals (CI) are printed. Only focussing on the last six inflection points,

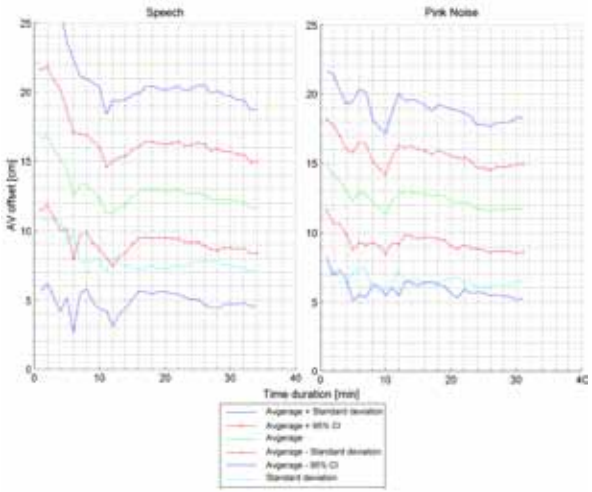


Figure 3: Average MPMD curve for speech and pink noise after post-screening

the MPMD for both test signals are in a range of 9 to 13 cm. An additional Shapiro-Wilk test showed that the individual MPMDs are normal distributed. Most of the individual MPMDs fall in a range of 12 cm. A two-sample t-test showed that here is no significant difference between the MPMDs of the different test signals.

3.2 Experiment 2

The noticeable offset between visual and audio stimuli, when the audio precede the visual stimulus was assessed here. The test panel consisted of the same but two assessors as in experiment 1 (8 female, 12 male). 19 out of 20 assessors evaluated both directions (LR and RL). Applying the same analysis methods as in experiment 1, figures 4 and 5 display the individual threshold curves of all participants. The graphs were used to check visually for outliers. Especially figure 5 suggested at least three potential outliers. Therefore the same post-screening method was applied as presented in experiment 1.

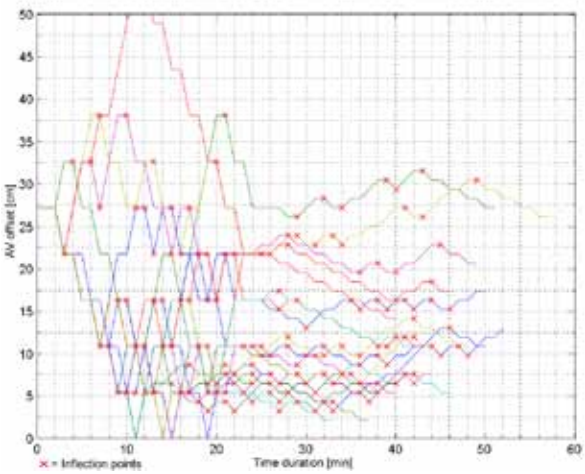


Figure 4: AV offset curves of 19 assessors for speech stimulus preceding the visual stimulus, moving from left to right

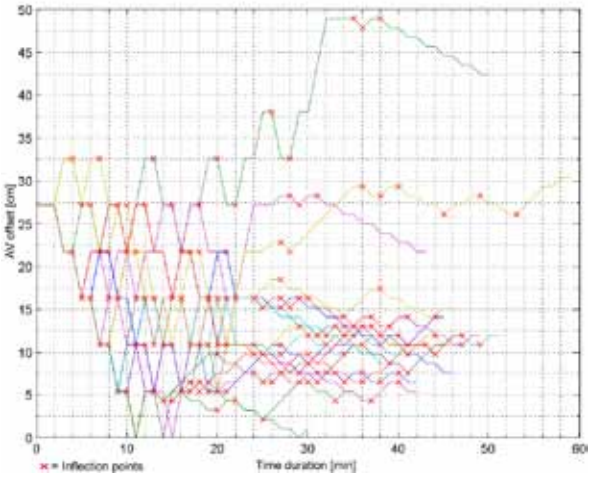


Figure 5: AV offset curves of 19 assessors for pink noise stimulus preceding the visual stimulus, moving from left to right

The analysis showed that in case of pink noise the data of only two assessors (02, 13) had to be rejected. For the speech signal, no outlier was detected. With this adjusted data the average MPMD curves for both test signals were calculated. As shown in figure 6 the average MPMD lays in the range of 8 to 15 cm. The Shapiro-Wilk test showed that the individual MPMDs are not normal distributed. Most of the thresholds are in a range of 5 to 10 cm. A higher number of thresholds fall in a range of 25 cm. This might indicate a bimodal distribution. Interestingly, most of the high thresholds are observed for the speech test item. Probably the test signal does influence in this case the perceived AV offset. This assumption was not further investigated.

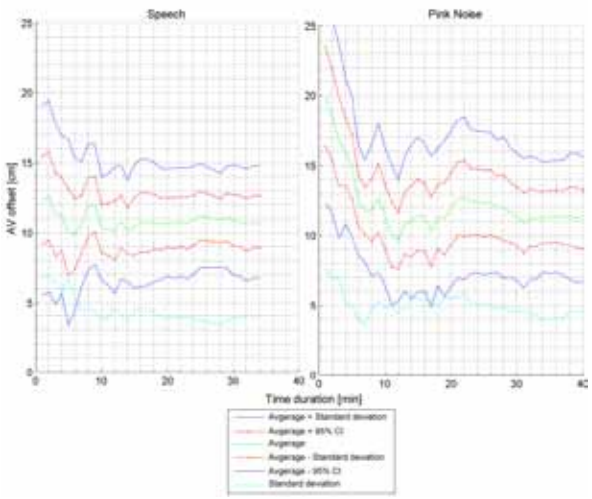


Figure 6: Average MPMD curve for speech and pink noise after post-screening

3.3 Comparison of Results

As the MPMDs do not show an entirely normal distribution, instead of comparing the mean values, the medians were used. The median value of the evaluated MPMDs for both directions (LR, RL), test signals (pink noise, speech) of both experiments were calculated and compared. The median as well as boxplots with upper and lower quartiles are displayed in figure 7.

General Observations

Under consideration of the median and the IQR, 7 cm represents the lowest MPMP (experiment 1, pink noise, LR and experiment 2, speech, LR). The maximum value lays at 19 cm (experiment 2, pink noise, RL). Comparing the single results of experiment 1 in the upper part of figure 7 to the values in the lower part (experiment 2), the relationship remain constant but the absolute values differ. A certain regularity in both experiments could be observed for threshold differences caused by the movement direction of the stimuli. Both observations will be discussed in the following.

Order of Test Stimuli

In the first experiment, where the audio stimulus follows the visual cue, a smaller deviation of the MPMDs could be observed. This is true for both test signals. The different deviations might result from the familiarity degree of the two test cases. Experiment one represents everyday life experience: The audio follows a visual event, due to the lower velocity of sound compared to light. In natural conditions the sound, reproduced by moving objects, arrives usually later at the observer. Audio preceding the visual cue as tested in experiment 2 is not as usual as the other way round. In addition the slightly larger median of the perception threshold of 12 cm (LR)/ 17 cm (RL) for averaged speech and pink noise in experiment 1 compared with 10 cm (LR)/ 14cm (RL) in test case 2, might also support this assumption. In the natural listening case the

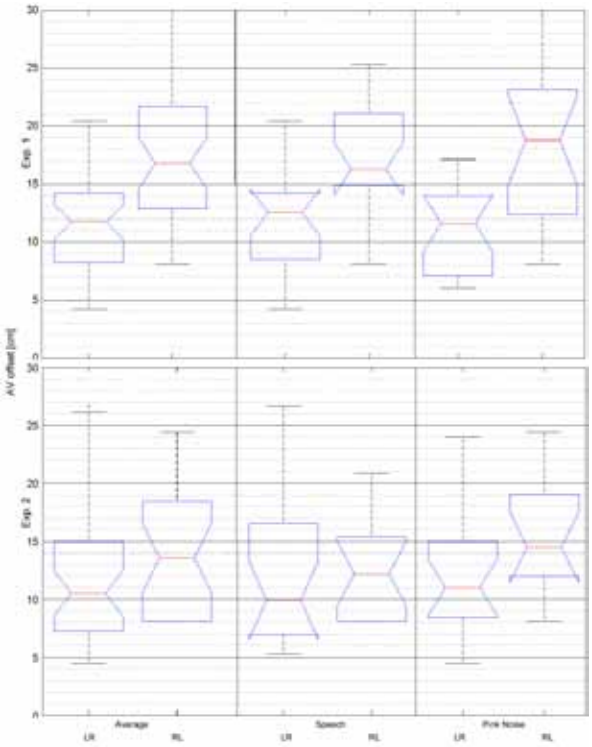


Figure 7: Median and boxplots of MPMDs for experiment 1 and 2, grouped according to the test signals (average of speech and pink noise, speech, pink noise) and presentation direction (LR and RL)

assessors are less sensitive to an offset between stimuli. In the more unnatural case (experiment 2) smaller AV offsets are plausible. Another reason for lower MPMD values in experiment 2 could have been resulted from the test design. All participants conducted the first experiment, followed by the second experiment. A training effect could have been occurred in the second experiment and therefore the assessors were more sensitive.

Moving Direction

The comparison of the moving direction shows an influence on the thresholds. The movement right to left results in a higher threshold in both experiments, independently from the test signal. Weather this effect is significant or not was not evaluated yet. One reason for the differences could be that less data points for movement direction RL were collected than for LR. The test procedure was interrupted after less inflection points, due to the original goal of this moving direction as a plausibility check. Hence, the resulting threshold for RL can only regarded as a rough approximation. A full test run might decrease the perceptual AV offset to the values resulting from moving direction LR.

4 Conclusions

Experiments to evaluate the perception thresholds for spatial audio-visual coherence have been conducted. There was a rather large spread in sensitivity between listeners which might be related to the fact that most listeners today are accustomed to huge differences in cinema and TV, and that only very well trained expert assessors are able to utilize the maximum possible performance of their audio-visual perception. The results show that trained expert listeners can perceive displacements as small as 2° (7 cm) for moving sources. The median of results was almost double that size. Due to the fact that after introduction of technical systems with better spatial resolution everybody will become a trained listener, the thresholds of the experts should be used when building technical systems.

No significant differences were observed when comparing the two stimuli speech and pink noise. There is a noticeable difference in perception whether audio is moving in front or after the video object. Due to differences in the speed of sound compared to the speed of light it is the normal situation that video is in front, and in quality tests it might be rated as unnatural if audio leads. There is also a noticeable difference in the thresholds depending on the direction of movement: Humans seem to be more sensitive for movements from left to right than vice versa. Whether this is a real effect or just a shortfall of the test procedure needs further assessment.

The results from this study are useful to create perceptual models to control the update rate of positions in encoders for spatial audio like MPEG-H 3D audio.

References

[1] H. Fastl and K. E. Zwicker, *Psychoacoustics: Facts and Models*, vol. 22 of *Springer Series in Information Sciences*, Springer, Berlin, 3rd edition, 2007.

[2] A. W. Mills, *The Minimum Audible Angle*, Harvard University, Harvard, 1958.

[3] D. R. Perrott and K. Saberi, "Minimum audible angle thresholds for sources varying in both elevation and azimuth," *Journal of the Acoustical Society of America*, vol. 87, pp. 1728–1731, 1990.

[4] J. D. Harris and R. L. Sergeant, "Monaural/binaural minimum audible angle for moving sound sources," *J. Speech Hear. Res.*, vol. 14, pp. 618–629, 1971.

[5] P. D'Arcy, *Histoire de l'Académie royale des sciences - Mémoire sur la durée de la sensation de la vue*, J. Boudot, Académie royale des sciences - Paris, 1765.

[6] P. Spindler, G. F. J. Meyer, and J. H. Meerburg, "Erklärung eines optischen Betruges bei Betrachtung der Speichen eines Rades durch vertikale Öffnungen," *Annalen der Physik*, vol. 81, pp. 93–104, 1825.

[7] M. Wertheimer, "Experimentelle Studien über das Sehen von Bewegung," *Psychologie und Physiologie der Sinnesorgane*, p. 161ff, 1912.

[8] P. Bertelson and M. Radeau, "Cross-modal bias and perceptual fusion with auditory-visual spatial discordance," *Perception and Psychophysics*, vol. 29, no. 6, pp. 578–584, 1981.

[9] Bernhard U Seeber and Hugo Fastl, "On auditory-visual interaction in real and virtual environments," *18. ICA Kyoto*, pp. 2293–2296, 2004.

[10] S. Getzmann, J. Lewald, and R. Guski, "Representational momentum in spatial hearing," *Perception*, vol. 33, no. 5, pp. 591–599, 2004.

[11] M. Radeau and Paul P. Bertelson, "Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations," *Perception and Psychophysics*, vol. 22, no. 2, pp. 137–146, 1977.

[12] C. E. Jack and W. R. Thurlow, "Effects of degree of visual association and angle of displacement on the ventriloquism effect," *Perceptual and Motor Skills*, vol. 37, pp. 967–979, 1973.

[13] J. Lewald, W. H. Ehrenstein, and R. Guski, "Spatio-temporal constraints for auditory–visual integration," *Behavioural Brain Research*, vol. 121, no. 1–2, pp. 69–79, 2001.

[14] N. Cote, V. Koehl, and M. Paquier, "Ventriloquism effect on distance auditory cues," in *Acoustics 2012*, Nantes, France, 2012, pp. 1063–1067.

[15] D. Corrigan, M. Gorzel, J. Squires, and F. Boland, "Depth perception of audio sources in stereo 3D environments," *Proc. SPIE*, vol. 8648, 2013.

[16] F. W. Fröhlich, "Über die Messung der Empfindungszeit [Measuring the time of sensation]," *Zeitschrift für Sinnesphysiologie*, vol. 54, no. 1–2, pp. 58–78, 1923.

[17] S. Getzmann, "Shifting the onset of a moving sound source: A Fröhlich effect in spatial hearing," *Hearing Research*, vol. 210, no. 1–2, pp. 104 – 111, 2005.

[18] S. Getzmann, "Representational momentum in spatial hearing does not depend on eye movements," *Experimental Brain Research*, vol. 165, pp. 229–238, 2005.

[19] S. Getzmann and J. Lewald, "Localization of moving sound," *Perception and Psychophysics*, pp. 1022–1034, 2007.

[20] K. Kunze, J. Liebetrau, and T. Korn, "Evaluating spatial congruency of 3d audio and video objects.," in *AES Convention: 132 (April 2012)*, 2012.

[21] C. André, *Audiovisual spatial congruence, and applications to 3D sound and stereoscopic video*, Dissertation, Université de Liège, 2014.

[22] Ville Pulkki, "Localization of amplitude-panned virtual sources ii: Two-and three-dimensional panning," *Journal of the Audio Engineering Society*, vol. 49, no. 9, pp. 753–767, 2001.

[23] Marinus M Boone, "Acoustic rendering with wave field synthesis," *Proceedings of the ACM Siggraph and Eurographics Campfire: Acoustic Rendering for Virtual Environments*, 2001.

[24] Markus Noisternig, Alois Sontacchi, Thomas Musil, and Robert Hóldrich, "A 3d ambisonic based binaural sound reproduction system," in *Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality*. Audio Engineering Society, 2003.

[25] Judith Liebetrau and Gabriel Gatzsche, "Digital cinema and object oriented sound reproduction," *Axmedis 2008*, vol. 45, pp. 138, 2008.

[26] Jürgen Herre, Johannes Hilpert, Achim Kuntz, and Jan Plogsties, "Mpeg-h audio—the new standard for universal spatial / 3d audio coding," in *Audio Engineering Society Convention 137*, Oct 2014.

[27] ISO/IEC JTC 1/SC29, "FDIS 23008-3: Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 3: 3D audio," 2015.

[28] Andrew Munro, "The design of dolby atmos post production theaters," in *Audio Engineering Society Convention 137*. Audio Engineering Society, 2014.

[29] G. Békésy, *Experiments in Hearing*, Acoustical Society of America through the American Institute of Physics by arrangement with McGraw-Hill Book Company, 1960.

[30] Harald Slatky, *Algorithms for direction specific Processing of Sound Signals-the Realization of a binaural i Cocktail-Party-Processor-System*, Ph.D. thesis, Dissertation, Ruhr-University Bochum, 1992.

[31] B. Seeber, *Untersuchung der auditiven Lokalisation mit einer Lichtzeigermethode*, Dissertation, TU München, München, 2010.

[32] ITU-R, "Recommendation BS.1116: Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems," 1997.

ARTIGOS
CURTOS
SHORT
PAPERS

Sociedade de Engenharia de Áudio
Artigo de Congresso

Apresentado no 13º Congresso de Engenharia de Áudio
19ª Convenção Nacional da AES Brasil
25 a 28 de Maio de 2015, 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.

Sensitivity Analysis of an Acoustic Sensor Localization Technique

Diego B. Haddad,[1] Luiz W. P. Biscainho,[2] Wallace A. Martins,[2] and Markus V. S. Lima.[2]

[1] CEFET-RJ/Nova Iguaçu, Telecommunications Coordination, Nova Iguaçu, RJ, 26041-271, Brazil.

[2] Universidade Federal do Rio de Janeiro, PEE/COPPE, Rio de Janeiro, RJ, 21941-972, Brazil.

diego@pads.ufrj.br, wagner@smt.ufrj.br, wallace.martins@smt.ufrj.br, markus.lima@ieee.org

ABSTRACT

From indoor navigation to surveillance, acoustic sensor localization (ASL) techniques find a wide range of applications. Yet, constrained to rely on inaccurate data such as time-of-flight (TOF) measurements, such techniques are prone to errors which can be quite difficult to predict. This paper takes as a case study a recently proposed ASL technique based on known acoustic signals emitted by loudspeakers at known positions, and proposes a sensitivity analysis relating the accuracy of location estimates to perturbations on the underlying mathematical model coefficients (especially those related to TOFs), which provides useful quantitative information to aid in design and performance assessment.

0 INTRODUCTION

Techniques for both sound source localization (SSL) and its dual problem, acoustic sensor localization (ASL), combine into an enabling technology [1] that have attracted substantial attention from the scientific community, due to their applicability to tasks as encompassing as equipment monitoring [2], target tracking [3], designing of efficient routing protocols [4], and surveillance [5], to mention just a few. In some applications, even to give meaning to the captured information, one must accurately know the sensor location [6]. Localization techniques often employ time-of-flight (TOF) [7], time-difference-of-flight (TDOF) [8], received-signal-strength [9], or angle-of-arrival [10] data/measurements to feed their related estimation models.

This paper focuses on the effect of inaccurate TOF estimates in the context of self-localization of an acoustic sensor. Blind *time-delay estimation* (TDE) [11] techniques that explore spatial diversity of sensors [12, 13] are of no use to perform TOF estimates when only information captured by the target sensor is available.

However, if several sound sources at known positions emit different previously known reference signals, one can use cross-correlation techniques to estimate the TOFs directly from the recorded signal. The weighted generalized cross-correlation (GCC) [14] and further improvements provide preferred solutions to this task. Unfortunately, estimated TOFs $\hat{t}$ may differ from their actual values t due to several issues/effects such as sampling, reverberation, interference, asynchronism, and clock skew[1] [15–18].

The present work resorts to a recently proposed ASL technique [17] for self-localization of portable devices based on the capture of reference signals emitted by loudspeakers as a case study to investigate how perturbations on TOF-dependent model coefficients influence localization accuracy. This sensitivity analysis can be useful to find loudspeaker locations that yield better performance during system design as well as to quantify in advance how reliable spatial coordinate estimates are.

This paper is organized as follows. Section 1 addresses the ASL technique under study. The related sensitivity analysis is described in Section 2. Simulation results are presented in Section 3, and conclusions are drawn in Section 4.

1 THE ASL TECHNIQUE

Let S be the number of loudspeakers and $\boldsymbol{s}_s \in \mathbb{R}^3$ be the 3D location of the sth loudspeaker. If the sensor is located at $\boldsymbol{m} \in \mathbb{R}^3$, one can define the TOF t_s between the sth loudspeaker and the sensor as:

$$t_s \triangleq \frac{\|\boldsymbol{s}_s - \boldsymbol{m}\|}{c}, \tag{1}$$

where c is the speed of sound and $\|\cdot\|$ stands for the Euclidean norm.

The loci of points that satisfy a specific TOF information is a sphere centered at the respective loudspeaker position with a radius of ct_s. The infinite number of points that adheres to a specific TOF constraint is not a problem when information arising from several loudspeakers is available. In this case, by gathering several TOF measurements, one is able to estimate the sensor location as the one that best matches the estimated TOFs. In mathematical terms, let $\overline{e}_s$ be the error associated with the sth TOF estimate $\hat{t}_s$ and with a candidate sensor position $\overline{\boldsymbol{m}}$:

$$\overline{e}_s \triangleq \hat{t}_s - \frac{\|\boldsymbol{s}_s - \overline{\boldsymbol{m}}\|}{\hat{c}}, \tag{2}$$

where $\hat{c}$ is an estimate[2] of c. Thus, the actual $\boldsymbol{m}$ could be estimated as

$$\hat{\boldsymbol{m}} \triangleq \underset{\overline{\boldsymbol{m}}}{\operatorname{argmin}} \|\overline{\boldsymbol{e}}\|_2, \tag{3}$$

where $\overline{\boldsymbol{e}} \triangleq \begin{bmatrix}\overline{e}_1 & \overline{e}_2 & \cdots & \overline{e}_S\end{bmatrix}^T$ and $\|\cdot\|_2$ stands for the l_2-norm. Although the optimization problem in (3) is quite intuitive, its nonlinear nature turns its solution very complicated. Hence, many approaches [19,20] alternatively minimize the norm of the error vector $\boldsymbol{e} \triangleq \begin{bmatrix}e_1 & e_2 & \ldots & e_S\end{bmatrix}^T$, where e_s is given by[3]:

$$e_s \triangleq \hat{t}_s^2 - \frac{\|\boldsymbol{s}_s - \overline{\boldsymbol{m}}\|^2}{\hat{c}^2}. \tag{4}$$

Since the sensor is not expected to have an embedded thermometer, the localization procedure must perform the estimation of c as a "nuisance parameter" [22], so that the overall procedure could be cast as the following minimization problem [20]:

$$\min_{\overline{\boldsymbol{\Theta}}} \|\boldsymbol{e}\|_2, \tag{5}$$

where $\boldsymbol{e} = \hat{\boldsymbol{H}}\overline{\boldsymbol{\Theta}} - \boldsymbol{g}$, with

$$\hat{\boldsymbol{H}} \triangleq \begin{bmatrix} 1 & 2\boldsymbol{s}_1^T & \hat{t}_1^2 \\ 1 & 2\boldsymbol{s}_2^T & \hat{t}_2^2 \\ \vdots & \ddots & \vdots \\ 1 & 2\boldsymbol{s}_S^T & \hat{t}_S^2 \end{bmatrix}, \tag{6}$$

$$\overline{\boldsymbol{\Theta}} \triangleq \begin{bmatrix}\|\overline{\boldsymbol{m}}\|^2 & \overline{\boldsymbol{m}} & \hat{c}^2\end{bmatrix}^T = \begin{bmatrix}\overline{\theta}_1 & \ldots & \overline{\theta}_5\end{bmatrix}^T, \tag{7}$$

$$\boldsymbol{g} \triangleq \begin{bmatrix}\|\boldsymbol{s}_1\|^2 & \|\boldsymbol{s}_2\|^2 & \ldots & \|\boldsymbol{s}_S\|^2\end{bmatrix}^T. \tag{8}$$

Note that, since $\overline{\theta}_1 = \overline{\theta}_2^2 + \overline{\theta}_3^2 + \overline{\theta}_4^2$, the elements of $\overline{\boldsymbol{\Theta}}$ are not independent, which calls for a constrained optimization procedure. Fortunately, ignoring such dependency does not significantly harm the location estimation [19]; thus, in a least-square sense, the actual $\boldsymbol{\Theta} \triangleq \begin{bmatrix}\theta_1 & \theta_2 & \ldots & \theta_5\end{bmatrix}^T$ (corresponding to the actual sensor location $\boldsymbol{m}$ and speed of sound c) can be estimated with reduced computational cost in closed-form[4] by the *unconstrained* estimation procedure [20,23]

$$\hat{\boldsymbol{\Theta}} \triangleq \left(\hat{\boldsymbol{H}}^T\hat{\boldsymbol{H}}\right)^{-1}\hat{\boldsymbol{H}}^T\boldsymbol{g}, \tag{9}$$

where only the second to fourth elements of $\hat{\boldsymbol{\Theta}}$, which contain the 3D sensor location estimate (see the parameter-vector structure in Eq. (7)), really matter.

2 SENSITIVITY ANALYSIS

The estimation procedure given in (9) finds the vector $\hat{\boldsymbol{\Theta}}$ that minimizes the l_2-norm of the residue of the following linear relation:

$$\hat{\boldsymbol{H}}\hat{\boldsymbol{\Theta}} \approx \boldsymbol{g}, \tag{10}$$

where the approximation symbol stresses the fact that in practice $\boldsymbol{g}$ is not an element of the column space of $\hat{\boldsymbol{H}}$.

[1] Clock skew may occur when loudspeaker and sensor do not share the same clock.

[2] The dependence of c with temperature can be modeled by $c \approx 331.3 + 0.606T$, with T in degrees Celsius.

[3] By redefining the error vector in terms of TDOF information only, a similar optimization procedure can be carried out by TDOF-based methods [21].

[4] If one assumes that $\hat{\boldsymbol{H}}$ is full rank.

The impact of possible errors in $\boldsymbol{g}$ on the estimation procedure can be mathematically modeled in a quite straightforward manner [23]. However, since $\boldsymbol{g}$ depends solely on the loudspeakers' locations, whose measurement can be made very accurately, this kind of error can be generally neglected.

On the other hand, matrix $\hat{\boldsymbol{H}}$ is more prone to errors due to its direct dependency on TOF estimates, which are often inaccurate, as discussed previously. Unfortunately, the influence of this matrix on the estimation process is much involved mathematically. A straightforward sensitivity analysis can provide a simple way to gather quantitative information linking TOF errors to localization inaccuracy. Thus, let $\hat{\boldsymbol{H}}$ be a corrupted version of the actual matrix $\boldsymbol{H}$:

$$\hat{\boldsymbol{H}} = \boldsymbol{H} + \boldsymbol{\Delta H}, \tag{11}$$

where $\boldsymbol{\Delta H}$ contains the estimation errors. In practice, $\boldsymbol{\Delta H}$ has a peculiar structure: its last column, related to the error-prone TOFs (see Eq. (6)), exhibits the largest entries. After a first-order analysis,[5] the impact of $\boldsymbol{\Delta\Theta}$ on the actual $\boldsymbol{\Theta}$ is encapsulated by the expression [24]:

$$\boldsymbol{\Delta\Theta} = \hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta} = -\left(\boldsymbol{H}^T\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T\boldsymbol{\Delta H}\boldsymbol{\Theta}, \tag{12}$$

from which the partial derivative w.r.t. each coefficient h_{ji} of matrix $\boldsymbol{H}$ can be computed [24] as

$$\frac{\partial\theta_k}{\partial h_{ji}} = -\left(\sum_{l=1}^{5}\left(\boldsymbol{H}^T\boldsymbol{H}\right)^{-1}_{kl} h_{jl}\right)\theta_i. \tag{13}$$

As long as $\frac{\partial\theta_k}{\partial h_{ji}}$ is known, one can evaluate the sensitivity $\mathcal{S}_{kji}$ of each coefficient θ_k to each coefficient h_{ji} by [25]:

$$\mathcal{S}_{kji} \triangleq \frac{h_{ji}}{\theta_k}\frac{\partial\theta_k}{\partial h_{ji}}, \tag{14}$$

whose absolute value can be used to predict how harmless to the localization procedure the misestimation of each parameter h_{ji} is.

In practice, the analysis can be restricted to $i = 5$, which corresponds to the last column of $\boldsymbol{\Delta H}$, where the largest errors are expected to occur.

3 SIMULATIONS

This section shows what kind of information one can gather from the sensitivity analysis described in Section 2.

The simulation setup was comprised of $S = 11$ loudspeakers located as listed in Table 1, and one microphone (sensor) positioned at the center of a room whose dimensions are $\{4.939 \times 7.345 \times 2.668\}$ m. The speed of sound was set at $c = 345$ m/s. In each experiment, the TOF corresponding to a single loudspeaker was corrupted by additive white Gaussian noise with standard deviation equal to 0.02 ms, which is compatible with TOF-estimation errors found in practice [17], and the sensor location was estimated according to Eq. (9). Discretization effects were not taken into account. The experiment was repeated 10^5 times for each choice of noisy-TOF loudspeaker, and the related mean-squared errors (MSE) were computed. The results corresponding to the 11 loudspeakers are depicted in Figures 1b to 5b for each component of $\hat{\boldsymbol{\Theta}}$, respectively.

Loudspeaker index	x (m)	y (m)	z (m)
1	0.3520	5.5430	2.2570
2	0.3880	5.5620	1.7065
3	0.6160	0.8275	2.0105
4	0.3120	2.8905	1.6430
5	0.3420	2.8725	0.9850
6	4.5720	3.0250	1.6020
7	4.5630	3.0185	1.0075
8	2.3315	0.5235	1.1710
9	4.5330	5.5830	2.2605
10	4.4740	5.5505	1.6875
11	4.1700	0.9000	2.0160

Table 1: Location of each loudspeaker.

For comparison, Figures 1a to 5a show respectively the sensitivities of each component of $\hat{\boldsymbol{\Theta}}$ w.r.t. the entry of the last column of $\boldsymbol{\Delta H}$ corresponding to the loudspeaker whose TOF has been perturbed. For a given TOF/loudspeaker index, as can be gathered from Figures 1 to 5 and confirmed by Table 2, the computed sensitivity is highly correlated with the MSE.

Such results demonstrate that a simple sensitivity computation can aid to predict the effects of badly estimated TOFs over the localization performance, and provide relevant information without resorting to extensive simulations—e.g., when estimating $\boldsymbol{m} = [x\ y\ z]^T$ in the present experiment, one expects to find larger errors in z than in x and y based on the proposed sensitivity analysis, which is corroborated by the simulated MSE results.

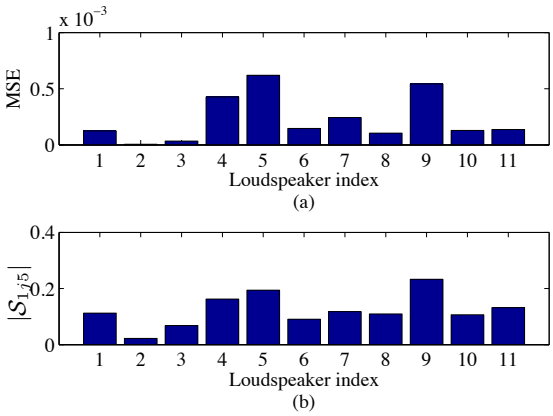


Figure 1: MSE and the absolute value of the sensitivity for θ_1, w.r.t. each loudspeaker.

[5] In this analysis, the identity $\boldsymbol{H\Theta} = \boldsymbol{g}$ holds.

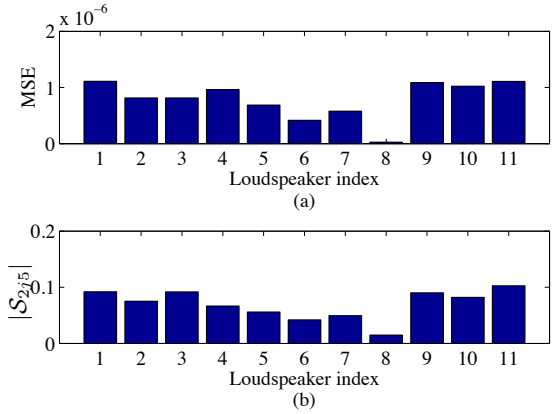


Figure 2: MSE and the absolute value of the sensitivity for θ_2, w.r.t. each loudspeaker.

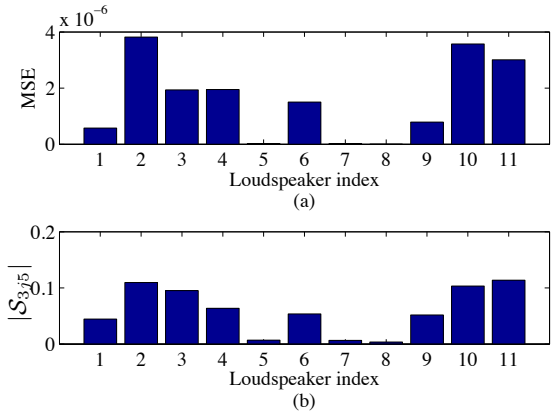


Figure 3: MSE and the absolute value of the sensitivity for θ_3, w.r.t. each loudspeaker.

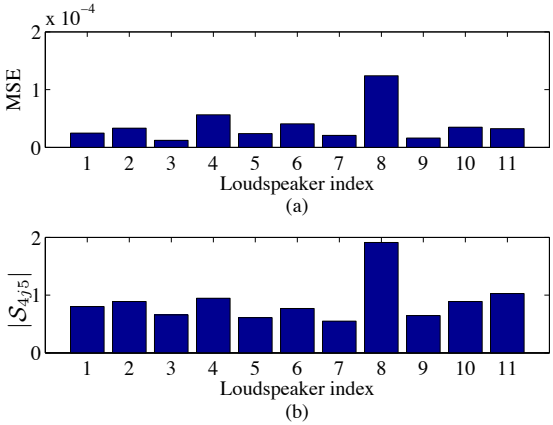


Figure 4: MSE and the absolute value of the sensitivity for θ_4, w.r.t. each loudspeaker.

4 CONCLUSIONS

Regarding self-localization of acoustic sensors, this paper started from a previously proposed ASL technique based on emissions of fixed sound sources (loudspeakers), and developed a sensitivity-based analysis that allows to predict the impact of deviations in their estimated TOFs to the sensor over the algorithm's localization performance, without resorting to extensive simulations. The main idea is to compute the sensitivity

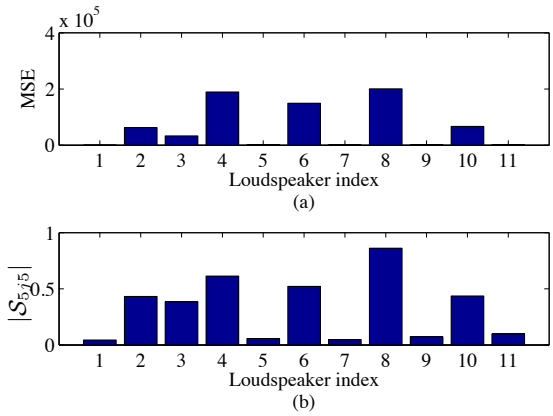


Figure 5: MSE and the absolute value of the sensitivity for θ_5, w.r.t. each loudspeaker.

Parameter	ρ
θ_1	0.83
θ_2	0.85
θ_3	0.86
θ_4	0.86
θ_5	0.85

Table 2: Correlation coefficient ρ between the MSE and the absolute value of the sensitivity (note that $|\rho| \leq 1$).

of the estimated sensor positions w.r.t. the coefficients related to TOFs in the mathematical model.

The analysis provides information that can be used to: compare and try to balance, according to the target application, the estimation errors in each spatial component; quantify the effect of each loudspeaker's information on the overall performance of the ASL algorithm; and, given the spatial region where the sensors are expected to occupy, use such data to better place the loudspeakers among viable positions within the room in order to minimize sensitivities.

REFERENCES

[1] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, R. L. Moses, and N. S. Correal, "Locating the Nodes: Cooperative Localization in Wireless Sensor Networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 54–69, Jul. 2005.

[2] K. H. Lee, C. H. Yu, J. W. Choi, and Y. B. Seo, "ToA based Sensor Localization in Underwater Wireless Sensor Networks," in *Proceedings of the International Conference on Instrumentaion, Control and Information Technology (SICE)*, Tokyo, Japan, Aug. 2008, SICE, pp. 1357–1361.

[3] X. Li, N. Mitton, I. Simplot-Ryl, and D. Simplot-Ryl, "Dynamic Beacon Mobility Scheduling for Sensor Localization," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 8, pp. 1439–1452, Aug. 2012.

[4] S. Das, I. Banerjee, and T. Samanta, "Sensor Localization and Obstacle Boundary Detection Algorithm in WSN," *Third International Conference on Advances in Computing and Communications (ICACC)*, pp. 412–415, Aug. 2013.

[5] R. V. Kulkarni and G. K. Venayagamoorthy, "Bio-inspired Algorithms for Autonomous Deployment and Localization of Sensor Nodes," *IEEE Transactions on Systems, Man, and Cybernetic - Part C: Applications and Reviews*, vol. 40, no. 6, pp. 663–675, Nov. 2010.

[6] S. Salari, S. Shahbazpanahi, and K. Ozdemir, "Mobility-Aided Wireless Sensor Network Localization via Semidefinite Programming," *IEEE Transactions on Wireless Communications*, vol. 12, no. 12, pp. 5966–5978, Dec. 2013.

[7] R. Dalce, A. van den Bossche, and T. Val, "Indoor Self-localization in a WSN, based on Time Of Flight: Propositions and Demonstrator," in *Proc. of International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Montbeliard-Belfort, France, Oct. 2013, IEEE, pp. 1–6.

[8] K. C. Ho, "Bias Reduction for an Explicit Solution of Source Localization Using TDOA," *IEEE Transactions on Signal Processing*, vol. 60, no. 5, pp. 2101–2114, May 2012.

[9] A. Edelstein, Xi Chen, Yunpeng Li, and M. Rabbat, "RSS-based Node Localization in the Presence of Attenuating Objects," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011, IEEE, pp. 3528–3531.

[10] H. Shao, X. Zhang, and Z. Wang, "Efficient Closed-Form Algorithms for AOA Based Self-Localization of Sensor Nodes Using Auxiliary Variables," *IEEE Transactions on Signal Processing*, vol. 62, no. 10, pp. 2580–2594, Mar. 2014.

[11] J. P. Ianniello, "Time Delay Estimation via Cross-correlation in the Presence of Large Estimation Errors," *IEEE Transactions on Acoustics, Speech and Audio Processing*, vol. 30, no. 6, pp. 998–1003, Dec. 1982.

[12] J. Benesty, "Adaptive Eigenvalue Decomposition Algorithm for Passive Acoustic Source Localization," *Journal of the Acoustical Society of America*, vol. 107, no. 1, pp. 384–391, Jan. 2000.

[13] Y. Huang and J. Benesty, "A Class of Frequency-domain Adaptive Approaches to Blind Multichannel Identification," *IEEE Transactions on Signal Processing*, vol. 51, no. 1, pp. 11–24, Jan. 2003.

[14] C. Knapp and G. Clifford Carter, "The Generalized Correlation Method for Estimation of Time Delay," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 24, no. 4, pp. 320–327, Aug. 1976.

[15] J. E. Salt and A. G. Wacker, "Optimistic and Pessimistic Approximations to Variance of Time Delay Estimators," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 5, pp. 634–641, May 1989.

[16] T. Gustafsson, B. D. Rao, and M. Trivedi, "Source Localization in Reverberant Environments: Modeling and Statistical Analysis," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 791–803, Nov. 2003.

[17] D. B. Haddad, W. A. Martins, L. W. P. Biscainho, M. do V. M. da Costa, and K. Kim, "Choosing Coherent Times of Flight for Improved Acoustic Sensor Localization," in *Proceedings of the International Telecommunications Symposium (ITS)*, São Paulo, Brazil, Aug. 2014, SBrT, pp. 1–5.

[18] J. Chen, J. Benesty, and Y. A. Huang, "Robust Time Delay Estimation Exploiting Redundancy Among Multiple Microphones," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 549–557, Nov. 2003.

[19] A. Beck, P. Stoica, and J. Li, "Exact and Approximate Solutions of Source Localization Problems," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1770–1778, May 2008.

[20] D. B. Haddad, L. O. Nunes, W. A. Martins, L. W. P. Biscainho, and B. Lee, "Closed-Form Solutions for Robust Acoustic Sensor Localization," in *Proceedings of the Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, USA, Oct. 2013, IEEE, pp. 1–5.

[21] P. Stoica and J. Li, "Lecture Notes - Source Localization from Range-Difference Measurements," *IEEE Signal Processing Magazine*, vol. 23, no. 6, pp. 63–66, Nov. 2006.

[22] J.-F. Cardoso, "Blind Signal Separation: Statistical Principles," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2009–2025, Oct. 1998.

[23] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*, Prentice Hall, Upper Saddle River, USA, 2000.

[24] S. D. Hodges and P. G. Moore, "Data Uncertainties and Least Squares Regression," *Journal of the Royal Statistical Society - Series C: Applied Statistics*, vol. 21, no. 2, pp. 185–195, 1972.

[25] G. Daryanani, *Principle of Active Networl Synthesis and Design*, John Wiley, Singapore, 1976.

Sociedade de Engenharia de Áudio

Artigo de Congresso

Apresentado no 13º Congresso de Engenharia de Áudio
19ª Convenção Nacional da AES Brasil
25 a 28 de Maio de 2015, 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.

Implementação de um Oitavador com Baixa Distorção

Bruno Griep Fernandes e Sidnei Noceti Filho
LINSE – Laboratório de Circuitos e Processamento de Sinais
Departamento de Engenharia Elétrica e Eletrônica, UFSC
Florianópolis, SC, 88040-900, Brasil
{brunogriep, sidnei}@linse.ufsc.br

RESUMO

Os oitavadores analógicos disponíveis no mercado proporcionam sinais com níveis relativamente altos de distorção. Neste trabalho, é proposto o projeto de um circuito que proporciona distorção relativamente baixa. A partir do sinal de uma guitarra, obtém-se uma onda quadrada cuja fundamental é a metade da frequência fundamental da corda percutida. A onda quadrada é aplicada simultaneamente em um Conversor Frequência-Tensão (FVC) (*Frequency-to-Voltage Converter)* e em um filtro passa-baixa de segunda ordem cuja frequência de corte é dependente da tensão na saída do FVC. Na saída do filtro, tem-se um sinal com nível de distorção relativamente baixo. O sinal oitavado e o sinal da guitarra são misturados no último estágio do circuito.

0. INTRODUÇÃO

Nos dias de hoje, é recorrente a discussão entre as vantagens e desvantagens de circuitos analógicos em relação a circuitos digitais, com relação às suas sonoridades. Não é escopo de este trabalho discutir este tema. Porém, pode-se dizer que diversas variáveis devem ser analisadas e, dentre elas, estão os níveis de ruído, a qualidade dos conversores A/D e D/A dos circuitos digitais e, por último, e talvez o mais importante, o gosto do ouvinte onde devem ser considerados efeitos psico-acústicos. Mas é inegável que para "hobistas" adeptos ao "Faça Você Mesmo" (em inglês os circuitos DIY (*Do It Yourself*)), apesar da grande difusão dos DSPs (de custos cada vez mais baixos), os circuitos analógicos são relativamente mais fáceis de construir. Estes fazem uso de componentes de fácil entendimento e são facilmente disponíveis, além de não exigirem conhecimento de programação.

Diversos pedais analógicos oitavadores são disponíveis no mercado. São encontrados modelos com os mais variados níveis de distorção. Os de mais alta distorção são conhecidos como *octafuzz*, e consistem em um oitavador com sonoridade do efeito *fuzz*. Como o objetivo deste trabalho é o projeto de um oitavador com baixo nível de distorção, foi realizada uma análise sonora de pedais com distorção relativamente menor, tais como os pedais das marcas Boss (modelo OC-3) [1], Electro Harmonix (modelo EHX Micro POG) [2], EBS (modelo Octabass) [3] e MXR (modelo Octave Deluxe) [4]. O pedal Boss OC-3 apresenta som com harmônicos ímpares, contendo opções de controle para aplicar efeito de distorção tanto no sinal oitavado quanto no sinal original. Os pedais EBS Octabass e EHX Micro POG apresentam um nível de distorção relativamente baixo, pois ambos apresentam controles de filtros onde o músico opta entre ressaltar as frequências graves, médias e/ou agudas. Desse modo, possibilita-se a diminuição ou o aumento da amplitude dos harmônicos, dando ao músico várias opções timbrais. O MXR Octave Deluxe apresenta a menor distorção entre os quatro modelos, tendo um controle que realça as médias frequências, o que leva o sinal oitavado a ter um timbre próximo ao timbre do instrumento.

1. FREQUÊNCIAS ENVOLVIDAS

A frequência fundamental produzida pela sexta corda de uma guitarra é 82,4Hz e pela primeira corda é 329,63Hz. Assim, a frequência mais baixa produzida pelo oitavador será $f_{min} = (82,4 \div 2) = 41,2\,\text{Hz}$. Se o 24º traste da primeira corda é percutido, a máxima frequência fundamental está duas oitavas acima de 329,63Hz, ou seja, 1318,52Hz. A divisão dessa frequência por dois resulta em $f_{max} = 659,26\,\text{Hz}$. Por simplicidade de cálculos, as frequências máxima e mínima consideradas serão $f_{min} = 40\,\text{Hz}$ e $f_{max} = 660\,\text{Hz}$. Isso resulta na razão $f_{max} / f_{min} = 16,5$.

2. DIAGRAMA EM BLOCOS SIMPLIFICADO

A Figura 1 mostra o diagrama de blocos do 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, com o intuito de minimizar a distorção, diversos procedimentos são realizados:

i) A onda quadrada é aplicada em um conversor frequência-tensão cuja saída é uma tensão contínua V_C' e, simultaneamente, em um filtro passa-baixa de segunda ordem cuja frequência de corte varia com uma tensão de controle V_B fornecida pelo bloco modificador de nível.

ii) O modificador de nível soma a tensão V_C' com outra tensão contínua produzindo o sinal de V_B que controla a frequência de corte do filtro passa-baixa.

iii) O filtro passa-baixa é excitado pela onda quadrada cuja frequência fundamental controla a frequência de corte. Os sinais harmônicos de ordem ímpar são filtrados com uma taxa de -12dB/o.

iv) Finalmente, o sinal filtrado e o sinal da guitarra são misturados no último bloco somador.

3. DESCRIÇÃO DOS BLOCOS

3.1 Bloco Conversor do Sinal da Guitarra em uma Onda Quadrada

A conversão de uma onda senoidal em uma onda quadrada pode ser feita com relativa facilidade por um circuito de ganho alto que apresente saturação na saída. Porém, 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.

Uma escolha criteriosa dentre todas as soluções apresentadas em sites que tratam de oitavadores, originou a escolha pelo circuito apresentado na Figura 2 [5].

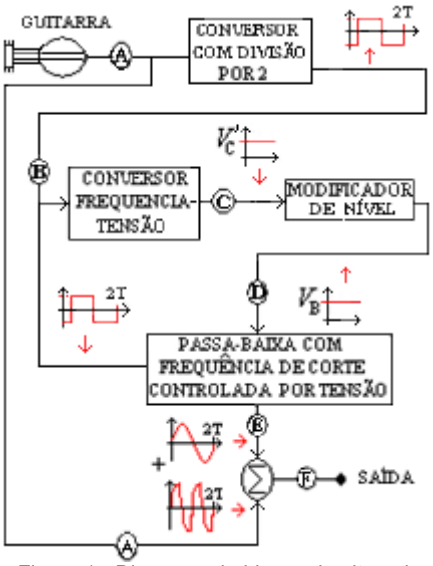


Figura 1 - Diagrama de blocos do oitavador.

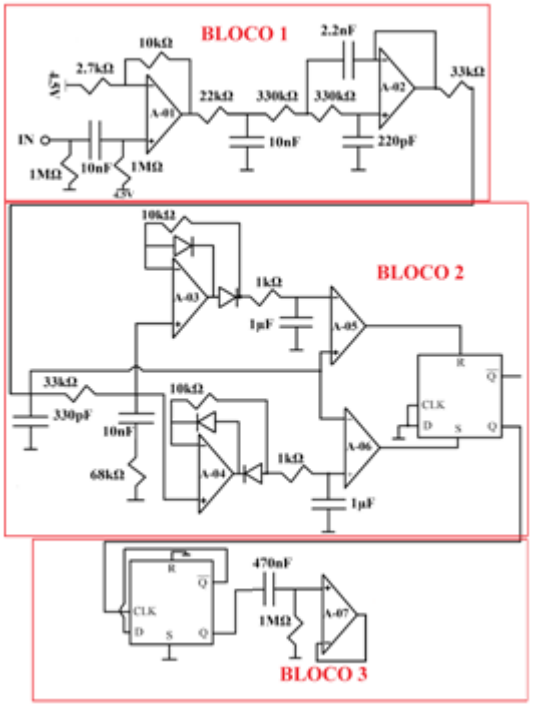


Figura 2 – Bloco conversor do sinal da guitarra em um sinal de onda quadrada.

O bloco 1 é um filtro passa-baixa que reduz as harmônicas dos sinais da guitarra. O ganho do amplificador A-01 garante que os sinais das frequências altas tenham amplitudes suficientemente grandes para que o circuito opere como o desejado.

O bloco 2 é responsável pela conversão do sinal filtrado em uma onda quadrada. Os amplificadores A-03 e A-04 estão na configuração de detectores de envelope, sendo um responsável pela detecção de valores positivos, e outro dos negativos. Ao passar pelos comparadores compostos por A-05 e A-06, os sinais se tornam pulsos defasados de 90º. As saídas dos comparadores estão devidamente conectadas aos terminais *SET* e *RESET* do primeiro *flip-flop*. Devido a esta diferença de fase entre os pulsos, os dois *flip-flops* juntos fornecem uma onda quadrada, com um *duty cycle* de 50%, com a mesma frequência

fundamental produzida pela corda percutida. Esse sinal aplicado no bloco 3 (contador [6]) divide a frequência do sinal por dois, proporcionando a onda quadrada sub-oitavada.

3.2 Bloco Conversor Frequência-Tensão

No bloco conversor frequência-tensão, é usado o circuito integrado LM331 [5], o qual analisa a onda quadrada de entrada e fornece um sinal contínuo V_C' que, segundo o *data-sheet* do fabricante, é dada por:

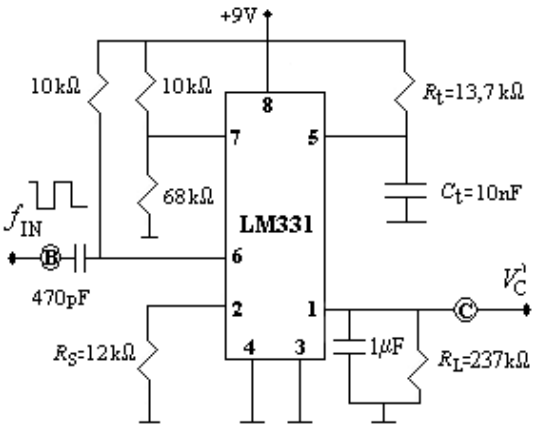


Figura 3 – Bloco conversor frequência-tensão (FVC).

$$V_C' = f_{\mathrm{IN}} \times 2,09 \times \frac{R_L}{R_S} \times R_T C_T \qquad (1)$$

onde f_{IN} é a frequência fundamental da onda quadrada. Conforme é calculado na subseção 3.3, os valores máximos e mínimos de V_C' devem ser $V'_{C\max} = 3,719\,\mathrm{V}$ para $f_{\max} = 660\,\mathrm{Hz}$ e $V'_{C\min} = 0,2254\,\mathrm{V}$ para $f_{\min} = 40\,\mathrm{Hz}$. Então, arbitrando $C_t = 10\,\mathrm{nF}$, $R_S = 12\,\mathrm{k\Omega}$ e $R_L = 237\,\mathrm{k\Omega}$, utilizando (1) determina-se R_t.

$$R_t = \frac{3,719 \times 12}{660 \times 2,09 \times 237 \times 10^{-8}} = 13,65 \cong 13,7\,\mathrm{k\Omega} \qquad (2)$$

Os valores dos demais componentes mostrados na Figura 3 são sugeridos no *data-sheet* do fabricante [5].

3.3 Bloco Modificador de Nível

O modificador de nível (somador inversor), obtido com o circuito do amp. op. da Figura 4, tem como objetivo produzir a tensão de polarização $(V_B = V_{\mathrm{C}} - 7,94\,\mathrm{V})$ do filtro passa-baixa obtido com o OTA LM13700 [7]. Este é descrito na subseção 3.4, cuja frequência de corte aumenta quando a tensão V_{C} aumenta. O objetivo é transformar a dependência com V_{C}, que é uma função incrementalmente linear, em uma função linear. A razão da tensão negativa de -7,94V é explicada a seguir.

A corrente de polarização do OTA, segundo o *data-sheet* do fabricante [7], é dada por (3), onde $2V_D$ é a tensão de polarização direta de duas junções PN em série ($2V_D \cong 1,06\mathrm{V}$ para a corrente de polarização usada) e $-V_{CC} = -9\,\mathrm{V}$ é a tensão de alimentação negativa.

$$I_B = \frac{V_B - 2V_D - (-V_{CC})}{R_B} = \frac{V_B + 7,94}{R_B} \qquad (3)$$

Fazendo com que $V_B = V_{\mathrm{C}} - 7,94\,\mathrm{V}$, a corrente de polarização passa a ser:

$$I_B = \frac{V_{\mathrm{C}} - 7,94 + 7,94}{R_B} = \frac{V_{\mathrm{C}}}{R_B}. \qquad (4)$$

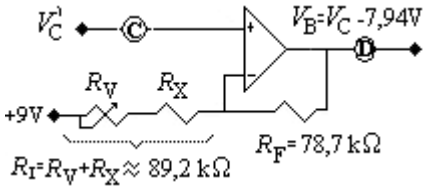


Figura 4 – Circuito do modificador de nível.

No circuito da Figura 4, para $V_{\mathrm{CC}} = +9\,\mathrm{V}$ e $R_F = 78,7\,\mathrm{k\Omega}$, deduz-se que $R_I = (9/7,94)R_F = 89,2\,\mathrm{k\Omega}$. Para permitir um ajuste fino da tensão V_{C}, R_I é obtido pela associação em série de um resistor fixo $R_X = 80,6\,\mathrm{k\Omega}$ e um trimpot multivoltas de valor $R_V = 20\,\mathrm{k\Omega}$.

A razão da colocação do resistor $R_V = 20\,\mathrm{k\Omega}$ é possibilitar o ajuste da corrente de polarização dos OTA'S e, consequentemente, sintonizar a frequência f_{P} do filtro da Figura 5. Isso porque as equações relacionadas aos OTAs são aproximadas e todos os componentes utilizados (capacitores e resistores) possuem tolerâncias de fabricação. O resistor R_V deve ser ajustado para a frequência mínima $f_{\min} = 40\,\mathrm{Hz}$ porque a sensibilidade em relação à tensão $2V_D$ é maior em razão da tensão produzida pelo FVC ser a menor dentre todas as tensões produzidas.

O valor de tensão contínua V_C' produzido pelo FVC mostrado na Figura 3 deve ter seu valor calculado de tal forma que o valor de V_{C} resultante seja aproximadamente 7V na mais alta frequência (660Hz) e (7/16,5) = 0,4242V na mais baixa frequência (40Hz). O valor de 16,5 corresponde à relação entre as frequências máxima e mínima processadas pelo FVC ($f_{\max} / f_{\min} = 16,5$).

$$V_{\mathrm{C}} = V_C' \frac{R_{\mathrm{F}} + R_{\mathrm{I}}}{R_{\mathrm{I}}} \qquad (5)$$

Considerando (5), os valores máximos e mínimos de V_C' serão:

$$V'_{C\max} = V_{\mathrm{C}} \frac{R_{\mathrm{I}}}{R_{\mathrm{F}} + R_{\mathrm{I}}} = 7 \frac{89,2065}{89,2065 + 78,7} = 3,719\,\mathrm{V} \qquad (6)$$

$$V'_{C\min} = 7/16,5 = 0,2254\,\mathrm{V} \qquad (7)$$

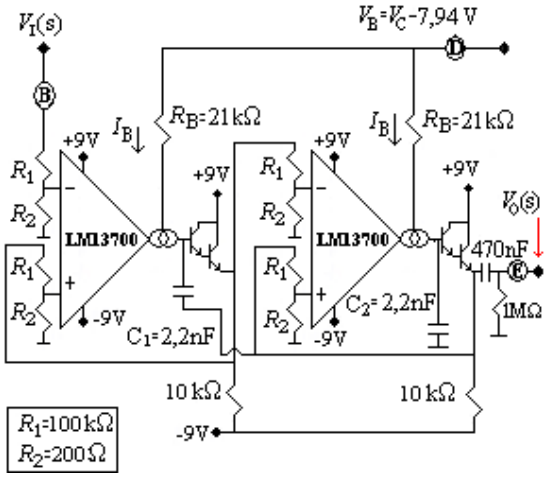


Figura 5 – Circuito do filtro passa-baixa.

3.4 Bloco do Filtro Passa-Baixa Controlado por Tensão

Para obter uma distorção relativamente baixa, optou-se por usar um filtro passa-baixa (ver Figura 5) de ordem 2 com fator de qualidade Q > 0,707 por este apresentar um sobrepassamento em frequência. Portanto, sintonizando o filtro no valor de pico da magnitude da resposta em frequência, a fundamental da onda é amplificada e os harmônicos ímpares da onda quadrada são filtrados com uma taxa de -12dB/o. A razão de não ter sido usado um filtro passa-faixa de segunda ordem é porque com este tipo de filtro os sinais harmônicos seriam filtrados com uma taxa de -6dB/o.

Deduzindo a função de transferência (FT) $T(s)$ do filtro passa-baixa da Figura 5, encontra-se:

$$T(s)=\frac{V_O(s)}{V_I(s)}=\frac{\dfrac{\alpha^2 g_{m1}g_{m2}}{C_1C_2}}{s^2+\dfrac{\alpha g_{m1}}{C_1}s+\dfrac{\alpha^2 g_{m1}g_{m2}}{C_1C_2}} \qquad (8)$$

onde $\alpha=R_2/(R_1+R_2)$. O divisor obtido com $R_2=200\Omega$ e $R_1=100\,\text{k}\Omega$ resulta em $\alpha\cong 0{,}002$. Esse divisor é usado para garantir níveis de tensão baixos na entrada e, consequentemente, baixa distorção na saída. Para $g_{m1}=g_{m2}=g_m$ e $C_1=C_2=C$, a FT se resume a:

$$\frac{V_O(s)}{V_I(s)}=\frac{\alpha^2 g_m^2/C^2}{s^2+(\alpha g_m/C)s+\alpha^2 g_m^2/C^2} \qquad (9)$$

$$\frac{V_O(s)}{V_I(s)}=\frac{K\omega_0^2}{s^2+(\omega_0/Q)s+\omega_0^2} \qquad (10)$$

Comparando com a função genérica de um filtro passa-baixa de segunda ordem dada por (10) com (9), obtém-se $K=1$, $Q=1$ e $\omega_0=\alpha g_m/C$. Para $Q>0{,}707$, como é o caso, o máximo valor da magnitude ocorre na frequência ω_P dada por [8]:

$$\omega_P=2\pi f_P=\omega_0(1-(1/2Q^2))^{1/2} \qquad (11)$$

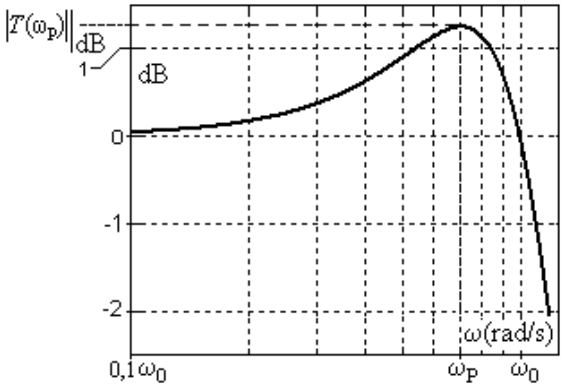


Figura 6 – Magnitude da resposta em frequência do filtro projetado.

A magnitude de maior valor é dada por [8]:

$$|T(\omega_P)|=|K|Q/(1-(1/4Q^2))^{1/2} \qquad (12)$$

Para $Q=1$, $K=1$ e $\omega=2\pi f$, $f_P=0{,}707 f_0$ obtém-se $|T(\omega_P)|=1{,}1547$. A Figura 6 ilustra as grandezas ω_P e $|T(\omega_P)|$.

A transcondutância g_m é dada por $g_m\cong 19{,}2\times I_B$ [10] e $\alpha\cong 0{,}002$. A corrente de polarização I_B é dada por (4), ou seja, $I_B=V_C/R_B$. Na frequência $f_{Pmax}=660\,\text{Hz}$, $V_C=7\,\text{V}$. Arbitrando $C=2{,}2\,\text{nF}$, e que $f_P=0{,}707 f_0$, tem-se:

$$f_0=\frac{\omega_0}{2\pi}=\frac{\alpha g_m}{2\pi C}. \qquad (13)$$

Então, a relação entre f_P e R_B é:

$$f_P=0{,}707\frac{\alpha g_m}{2\pi C}=0{,}707\frac{0{,}002\times 19{,}2\times 7}{2\pi\times 2{,}2\times 10^{-9}\times R_B} \qquad (14)$$

ou:

$$R_B=\frac{13748209{,}106}{f_P}\cong 21\,\text{k}\Omega. \qquad (15)$$

3.5 Bloco do Misturador de Saída

O estágio de saída mostrado na Figura 7 é um simples somador que mistura o sinal da guitarra com o sinal oitavado. Os valores de R_O, que controla o nível do sinal oitavado, e de R_G, que controla o nível do sinal da guitarra, devem ser escolhidos ao gosto do projetista.

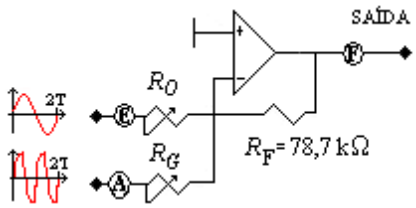


Figura 7 – Misturador de saída.

3.6 Cálculo da Distorção Harmônica

Segundo a norma IEC 268-3, de 1988, a distorção harmônica total (THD) de um sinal é obtida pela razão entre o valor RMS (ou de pico) de todos os harmônicos e o valor RMS (ou de pico) da frequência fundamental somada aos valores RMS (ou de pico) dos sinais harmônicos, conforme em (16).

$$\mathrm{THD} = \frac{\sqrt{V_{2f}^2 + V_{3f}^2 + \cdots + V_{nf}^2}}{\sqrt{V_f^2 + V_{2f}^2 + V_{3f}^2 + \cdots + V_{nf}^2}} \qquad (16)$$

Para uma onda quadrada com valor médio nulo, a amplitude $V(n\omega)$ da fundamental e das harmônicas, para todo n ímpar, é dada por (17):

$$V(n\omega) = (2/n\pi)\mathrm{sen}(n\omega) \qquad (17)$$

A magnitude da resposta em frequência do filtro é dada por (18), onde $\omega_0 = 1{,}414\omega_P = (1{,}414 \times 2\pi \times 660)\,\mathrm{rad/s}$:

$$\left|\frac{V_O(\omega)}{V_I(\omega)}\right| = \frac{\omega_0^2}{\sqrt{(\omega_0^2 - \omega^2)^2 + \omega_0^2\omega^2}} \qquad (18)$$

Com o auxílio de (18) e calculando, por simplicidade, até o sétimo harmônico, as amplitudes são $|T(\omega_P)| = 1{,}1547$, $|T(3\omega_P)| = 0{,}24434$, $|T(5\omega_P)| = 0{,}083117$ e $|T(7\omega_P)| = 0{,}041639$.
Calculando as amplitudes da fundamental e dos harmônicos, usando (17) e (18), obtém-se:

$$V_f^2 = [1{,}1547 \times (2/\pi)]^2 = 0{,}540379 \qquad (19)$$

$$V_{3f}^2 = [0{,}24434 \times (2/3\pi)]^2 = 0{,}002688 \qquad (20)$$

$$V_{5f}^2 = [0{,}083117 \times (2/5\pi)]^2 = 1{,}119953\text{e-}4 \qquad (21)$$

$$V_{7f}^2 = [0{,}041639 \times (2/7\pi)]^2 = 1{,}43405\text{e-}5 \qquad (22)$$

Calculando a THD dada por (16), encontra-se:

$$\mathrm{THD} = \frac{\sqrt{V_{3f}^2 + V_{5f}^2 + V_{7f}^2}}{\sqrt{V_f^2 + V_{3f}^2 + V_{5f}^2 + V_{7f}^2}} \qquad (23)$$

$$\mathrm{THD} = \frac{0{,}053050314}{0{,}7370165} = 0{,}071979 \Rightarrow 7{,}1979\% \qquad (24)$$

A distorção obtida é perfeitamente aceitável para os músicos que não gostam dos efeitos com nível de distorção relativamente mais alto (*fuzz*, por exemplo).

A distorção de 7,1979% pode se diminuída se for colocado mais um filtro passa-baixa em cascata.

4 CONCLUSÕES

Neste trabalho, é apresentado o projeto de um oitavador analógico para sinais de uma guitarra, que apresenta níveis de distorção relativamente baixos em relação aos oitavadores analógicos disponíveis no mercado. O princípio do oitavador se baseia na obtenção de uma onda quadrada cuja frequência fundamental é a metade da fundamental do sinal de uma guitarra. Essa função é realizada das mais variadas formas com circuitos relativamente complexos, devido aos vários cruzamentos por zero dentro de um período do sinal produzido. A solução encontrada faz uso de um circuito encontrado na literatura pertinente. Após, a onda quadrada é aplicada em um conversor frequência-tensão que produz um sinal DC cuja amplitude aumenta quando a frequência aumenta.

Foi encontrada uma solução simples que transforma uma função incrementalmente linear em uma função linear no que diz respeito ao controle do filtro passa-baixa controlado por tensão. Essa simples solução pode ser aplicada em muitos circuitos que fazem uso de um OTA.

O sinal DC controla a frequência de corte de um filtro passa-baixa de forma que os harmônicos da onda quadrada são atenuados, diminuindo a distorção. A distorção obtida é relativamente menor do que aquelas obtidas com os quatro oitavadores conhecidos, que, por sua vez, apresentam uma menor distorção em relação a todos os outros oitavadores comerciais pesquisados.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] ROLAND BOSS. Detalhes sobre o pedal BOSS OC-3. Disponível em: http://bossinfo.roland.com.br/gear/542/608 . Acesso em: 28/01/2015.
[2] ELECTRO HARMONIX. Detalhes sobre o pedal EHX Micro POG. Disponível em: http://www.ehx.com/products/micro-pog. Acesso em: 28/01/2015.
[3] EBS. *Análise do pedal EBS Octabass*. Disponível em: http://www.ebssweden.com/content/reviews/blp/blp_BassGuitar2004.pdf . Acesso em: 28/01/2015.
[4] DUNLOP MXR . *Análise do pedal MXR Octave Deluxe*. Disponível em: http://www.jimdunlop.com/product/m288-bass-octave-deluxe . Acesso em: 28 jan 2015.
[5] Boss OC-2 Octave pedal schematic diagram http://www.hobby-hour.com/electronics/s/oc2-octave.php Acesso em: 08/04/2015 2015.
[6] CD4024 datasheet.pdf. Disponível em http://media.digikey.com/pdf/Data%20Sheets/Fairchild%20PDFs/CD4024BC.pdf Acesso em: 28 jan 2015
[7] LM13700 datasheet.pdf
[8] S. Noceti Filho, Filtros Seletores de Sinais, 3 ed., Florianópolis, SC, Editora UFSC, 2010

Sociedade de Engenharia de Áudio

Artigo de Congresso

Apresentado no 13º Congresso de Engenharia de Áudio
19ª Convenção Nacional da AES Brasil
25 a 28 de Maio de 2015, 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 e Reprojeto de Mesa de Alta Qualidade para Gravação Acústica

Marcelo Johann[1], Andrei Yefinczuk[1], Rodolfo Pirotti[1] e Marcio Chiaramonte[2]
[1]UFRGS - Instituto de Informática - PGMICRO/PPGC
Porto Alegre, RS, 95501-970, Brasil
[2]Metalúrgica Meber
Bento Gonçalves, RS, Brasil
johann@inf.ufrgs.br, andrei.yefinczuk@gmail.com, rodolfo.p.pirotti@gmail.com, marcio@meber.com.br

RESUMO

Neste trabalho apresentamos uma breve análise de uma mesa de som de alta qualidade, dedicada a captação sonora acústica, especialmente de concertos de música erudita, que foi construída artesanalmente e usada durante mais de 20 anos para o registro da grande maioria dos concertos realizados na cidade de Porto Alegre no período. Apresentamos em seguida um relato do processo de reprojeto, pelo qual construímos uma nova e atualizada versão dessa ferramenta, e os resultados obtidos. O artigo apresenta como principal contribuição o registro histórico e técnico desse equipamento, através de seu esquemático, enfatizando a importância do projeto dedicado de circuitos de qualidade excepcional, que normalmente não é possível atingir satisfazendo as restrições usuais da indústria e do mercado.

0. INTRODUÇÃO E CONTEXTO HISTÓRICO

Na área de áudio, existem vários exemplos emblemáticos de pessoas que não tiveram graduação em engenharia eletrônica, mas que contribuíram de forma muito significativa com o desenvolvimento de equipamentos eletrônicos, sendo reconhecidos por sociedades profissionais com prêmios e distinções. O lendário inglês Gilbert Briggs, fundador da companhia Wharfedale™ é considerado como o pai do áudio de alta-fidelidade (*Hi-Fi*) [1], por seu trabalho pioneiro e levado à perfeição na construção de alto-falantes, e sua famosa turnê na década de 1950, em parceria com Peter Walker, dos amplificadores Quad™, para demonstrar as características excelentes de gravação e reprodução de alta fidelidade, a platéias ao vivo em diversas grandes cidades do mundo.

Com semelhante paixão pela música, especialmente música erudita, e por equipamentos de reprodução e gravação da mais alta qualidade possível, o odontólogo brasileiro Marcello Sfoggia (1931-2015) seguiu desde cedo os passos dessa corrente de áudio *Hi-Fi*, especialmente Gilbert Briggs e outros tantos entusiastas. Durante sua vida, desenvolveu suas habilidades como técnico em eletrônica, sempre com o foco em alta-qualidade, construindo incontáveis equipamentos valvulados e superdimensionados para audição, reprodução de

gravações em Vinyl e em fitas [3]. Seu interesse ilimitado o levou diretamente ao processo de gravação, e passou a usar, modificar e construir equipamentos para captação acústica profissional, isto é, para gravação de concertos, especialmente de música erudita, com a acústica do ambiente [2].

Nesse caminho evolutivo, passou de gravações em rolos de fita analógica, para gravações digitais em fitas de video, em DAT, ADAT™, e recentemente, em computador. Na parte de microfones e pré-amplificadores, permanece, na sua opinião, como na de muitos outros, a preferência sólida por circuitos puramente valvulados, os quais são em geral grandes e caros. Entretanto, como Sfoggia se tornou o responsável pela gravação rotineira, semanal, de concertos das orquestras na cidade de Porto Alegre, ele acabou desenvolvendo uma pequena mesa de som compacta, com amplificadores operacionais (*op-amps*) integrados [4]. Apesar do aparente compromisso de qualidade na construção de um equipamento mais leve e compacto, essa pequena mesa de gravação exibe as mesmas características sonoras principais de outros equipamentos seus: uma qualidade irrepreensível, um som sólido e detalhado, resultado do uso de componentes selecionados e técnicas de projeto e montagem cuidadosas.

Além de registrar as apresentações semanais da Orquestra Sinfônica de Porto Alegre (OSPA) e de várias outras orquestras locais e visitantes, essa pequena mesa tem sido empregada por ele nesses últimos 20 anos com sucesso para diversas gravações especiais, eventos, álbuns, discos e CDs, missão para a qual equipamentos de estúdio muito caros seriam demandados, e freqüentemente com um resultado inferior, principalmente quando combinada com seu processo também próprio de mixagem valvulada subsequente.

Neste trabalho, efetuamos uma breve análise desta mesa compacta construída por Sfoggia e apresentamos o projeto de uma mesa de som com características semelhantes, conforme descrito a seguir. Como motivação e objetivos, procuramos:

- iniciar o registro e documentação de sua produção tecnológica;
- ter condições de oferecer manutenção e dar continuidade ao uso de seus equipamentos;
- dar continuidade ao processo iniciado por ele de gravação da produção musical local com uma qualidade diferenciada;
- desenvolver e divulgar o conhecimento sobre construção de equipamentos com qualidade para gravação (de estúdio) e de construção dedicada, especificamente carente no Brasil;
- dispor de uma plataforma para teste de outros circuitos e componentes, de forma a continuar aprimorando o trabalho a partir do ponto até onde Sfoggia chegou;

O restante desse artigo está assim organizado. Na seção 1 apresentamos definições e características principais deste equipamento, as linhas gerais nas quais se fundamentam o trabalho de montagem dedicada nele empregado e o circuito levantado resultado da análise realizada. Na seção 2 apresentamos o processo de reprojeto dessa ferramenta e os resultados obtidos até então com algumas medidas e testes comparativos. Finalmente, apresentamos nossas conclusões.

1. A MESA CONSTRUÍDA POR SFOGGIA

Uma das estratégias mais tradicionais de microfonação indica o uso do menor número possível de microfones para captação de uma orquestra, de forma a preservar a imagem, a dinâmica e a acústica particulares do local. Em técnicas como *estéreo A-B* ou *Decca Tree*, emprega-se tipicamente um par espaçado de microfones omnidirecionais para captar a maior parte da orquestra, e mais um microfone central para melhor capturar as madeiras, que estão muito distantes dos microfones laterais. Os metais e percussão já tem maior energia, e microfones adicionais são necessários apenas para instrumentos solistas, como piano, harpa, cantores, etc. Neste contexto, tanto os microfones como os pré-amplificadores devem ser muito sensíveis, hábeis para captar detalhada e uniformemente os sons relativamente distantes de toda a orquestra.

1.1 Definições e Características Gerais

Para atender a essa finalidade, a mesa de gravação em questão consiste fundamentalmente em um pequeno conjunto de pré-amplificadores neutros e precisos. Ao contrário da microfonação de instrumentos individuais e de gravações de estúdio, não se costuma empregar equalização nos canais, pois estes (e os microfones neles ligados) devem capturar o som da forma mais fiel ao som ambiente possível, sem alterações ou correções.

Além dos pré-amplificadores, que têm saídas individuais para gravação multi-canal, a mesa em questão apresenta os estágios somadores (mix) e de saída com a finalidade principal de gravar uma cópia de segurança. As funções de monitoria e de difusão do som ao vivo, por exemplo, para Rádio e TV, são mais frequentemente feitas na interface de gravação e com outra mesa de linha, já que o volume dos pré-amplificadores é dedicado a ajustar adequadamente os níveis de gravação.

1.2 Princípios Norteadores

É possível identificar alguns pontos que formam uma filosofia de trabalho específica nos circuitos projetados e montados por Sfoggia. Esses princípios

norteadores podem ser resumidos nos seguintes aspectos:

- Construção de circuitos dedicados, com técnicas específicas, algumas artesanais, livres de compromissos de tamanho, custo e praticidade típicos dos projetos comerciais;
- Simplicidade: quanto mais direto, melhor; em alguns casos, componentes e equipamentos simples oferecem um desempenho comparável ou superior ao de equipamentos caros e sofisticados;
- Projeto sobre-dimensionado, aliviando a carga sobre os componentes, especialmente nas fontes de alimentação, com transformadores que não devem esquentar, bons dissipadores, filtros RCL, capacitores grandes;
- Componentes de qualidade e selecionados, muitos deles testados individualmente;

1.3 Análise da mesa

No dia 19/05/2014 os autores deste artigo reuniram-se com autorização do Dr. Sfoggia para analisar e documentar os circuitos implementados em sua mesa de som. A Figura 1 apresenta uma das fotografias do circuito, feitas na ocasião.

Figura 1 - Foto do circuito completo da mesa original

Algumas características gerais que podem ser imediatamente notadas são as seguintes:

- o circuito é montado ponto-a-ponto, praticamente sem uso de placas, minimizando todos os caminhos e interações entre os sinais;
- a distribuição de terra tem uma estrutura bastante grande, com fios de até 2.5 mm individuais dos transformadores de entrada, e ligando cada estágio de amplificador operacional em filas e entre si, sendo estes terras todos unidos em um único e massivo ponto junto à entrada de alimentação.
- todos os *op-amps* tem filtros individuais, compostos por capacitores eletrolíticos de 220 μF e separados da linha de alimentação por resistores de 10 Ω, o mesmo ocorrendo na alimentação de 48 V para os microfones (*phantom power*).

1.4 Esquema dos Circuitos

As Figuras 4 e 5 apresentam respectivamente os esquemáticos do caminho de gravação em cada canal, que representam os pré-amplificadores, e do caminho de mixagem, que sai logo após o primeiro operacional, repete o segundo estágio de ganho, o qual endereça o circuito panorâmico (pan), o somador e mais um estágio final.

Os transformadores Jensen™ JT-115K-E já provêem um ganho de 10x no sinal. O primeiro operacional possui uma chave seletora, apresentando ganho variável em três posições, de aproximadamente 1x, 3x e 10x. Esse estágio é praticamente idêntico ao esquemático apresentado na folha de dados do amplificador operacional integrado de precisão OP27 [5]. Nessa nota de aplicação, os engenheiros não só apresentaram o esquema como um amplificador de microfone de altíssimo desempenho para áudio, como indicam especificamente o transformador de entrada JT-115K-E, e fazem uma comparação mostrando que o ruído resultante desse circuito está apenas 3.6 dB acima do ruído gerado por um resistor de carga de 150 Ω e resistores indicados R1 e R2.

Finalmente, o segundo operacional oferece um ganho de mais 10x, totalizando +60 dB nesse caminho de gravação. A pré-amplificação em dois estágios oferece menor distorção e maior banda, como empregada no pré-amplificador Jensen Twin-Servo 990 [6]. A colocação do atenuador de volume do canal entre esses dois estágios é uma característica interessante que permite ao segundo estágio ser o *driver* do cabo de gravação, sem necessidade de um estágio de potência adicional na saída. Após o segundo operacional, apenas um conjunto de capacitores de alta qualidade está presente.

Como se pode observar, o circuito construído é tecnicamente muito simples, minimalista, mas a montagem e a sofisticação de usar componentes de qualidade e filtros individuais é um dos fatores que diferenciam esse projeto da maioria dos circuitos existentes em equipamentos profissionais. Na sua simplicidade está justamente um dos seus maiores valores, já que praticamente não existem circuitos assim disponíveis para o mercado.

2. PROJETO E CONSTRUÇÃO DE NOVA MESA DE GRAVAÇÃO

Considerando a motivação e os objetivos já citados, foi projetada e construída uma nova mesa de som com as mesmas características. Os pré-amplificadores (caminho de gravação) foram

preservados o mais fielmente possível. As maiores alterações foram na estrutura física, com uso de pequenas placas, agilizando e sistematizando a montagem, mas mantendo trilhas de comprimento mínimo. Para desacoplamento dos *op-amps*, foram usados capacitores eletrolíticos da série KZ Muse Nichicon™ e aumentado o valor para 470 μF em cada linha de alimentação, em cada *op-amp*.

Os estágios de soma e saída da mesa foram re-projetados, apesar de manterem esquema semelhante ao apresentado na figura 5. Foram incluídos envio e retorno para unidade de efeito, um par de saídas com volumes independentes e um par de amplificadores de fone também independentes. Por não serem considerados partes críticas do projeto, entretanto, os estágios de soma e saída estão em uma pequena placa separada, assim como os amplificadores de fone, ambos com *op-amps* quádruplos, e não serão avaliados neste artigo.

2.1 Testes e Resultados com Protótipo

Com a finalidade de testar o desempenho dos novos pré-amplificadores, verificar e corrigir quaisquer problemas, foi primeiro montado um protótipo completo de dois canais. Os níveis de ruído medidos foram inicialmente comparados com um pré-amplificador de dois canais comercial na faixa de preço de US$ 600. Na Figura 2 pode-se verificar que o ruído do protótipo é menor do que o comercial quando operando com o ganho máximo (60 dB), e um pouco maior quando operando com o ganho mínimo (40 dB), com os ganhos calibrados para os mesmos níveis de sinal. As medidas foram feitas com uma carga de entrada de 200 Ω.

O desempenho acústico foi experimentado com uma série de microfones, incluindo condensador de cápsula pequena, condensador de cápsula grande, valvulado, microfones de fita, e ouvindo com fones Senheiser™ HD650. Em todos os experimentos, o resultado sonoro sempre esteve de acordo com as expectativas: extremamente detalhado, bem definido, sem qualquer ruído ou artifício perceptível.

2.2 Comparação com a Mesa Original

A Figura 3 apresenta as curvas de resposta de frequência da mesa original e da nova mesa, para as três configurações de ganho. As curvas foram obtidas através do *software* Electroacoustics Toolbox, corrigidas pela resposta do laço DA/AD da interface MetricHalo ULN2. As curvas foram geradas com o potenciômetro atenuador em uma posição intermediária, cobrindo de +30 dB a +50 dB de ganho total. A resposta dos pré-amplificadores está dentro de 1 dB na faixa de 4 Hz a cerca de 40 KHz. As medidas de ruído foram semelhantes às da figura 2 tanto para a mesa original quanto para a nova, com níveis variando entre -120 dB e -100 dB para ganhos típicos de +30 a +50dB. A tentativa de medição de THD feita com sinal de 1 KHz a -8 dBFS reporta respectivamente valores de -94.36 dB -91.26 dB para a mesa nova e original. Por essa medida e pela inspeção visual, verifica-se que não há harmônicas distinguíveis dos níveis de ruído de fundo. Assim, não foram identificados problemas significativos, e o desempenho dos circuitos reprojetados foi idêntico ao da mesa original.

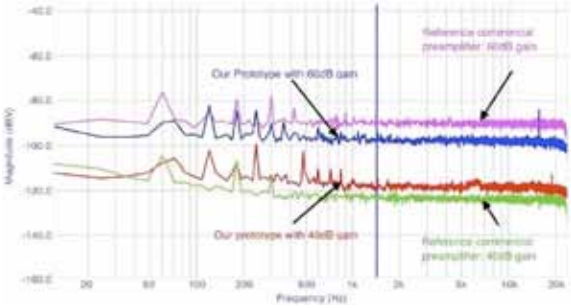


Figura 2 Ruído com ganho de +40 dB e +60 dB comparado com um pré-amplificador comercial de qualidade.

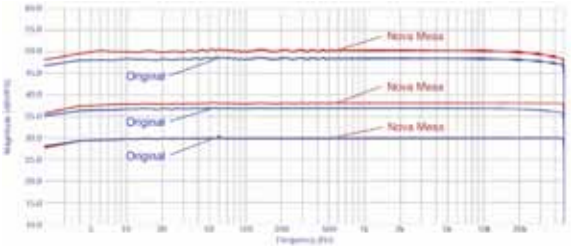


Figura 3 Resposta de frequência da mesa nova e da original, para três posições de ganho, de +30 dB a +50 dB.

3. CONCLUSÕES

Neste artigo apresentamos o registro histórico da mesa de som compacta montada artesanalmente pelo dedicado especialista em gravações Marcello Sfoggia. O circuito baseado em entradas com transformadores Jensen e operacionais de precisão OP27, aliado às técnicas cuidadosas de montagem ponto a ponto resultam em um equipamento com som sólido e de alta definição, sendo um dos responsáveis pela notável qualidade das gravações feitas por este técnico, amplamente reconhecidas.

O processo de análise e reprojeto nos permitiu verificar diversos aspectos de projeto e montagem de circuitos de áudio de alto desempenho. Destacam-se a simplicidade do caminho de sinal, a importância da qualidade de alguns componentes, a qualidade e independência da fonte e das linhas de alimentação. Na construção, verificamos que os transformadores, os amplificadores operacionais individuais, os capacitores de desacoplamento, todos têm custo tipicamente uma ordem de grandeza maior do que componentes normalmente empregados na maioria dos equipamentos comerciais, mesmo profissionais. Somente essa característica já indica parcialmente a superioridade técnica desse projeto.

Concluímos que o projeto de equipamentos dedicados permite mais facilmente se dispor de circuitos de altíssimo desempenho, vencendo restrições que a indústria normalmente precisa respeitar, quanto à praticidade, generalidade de aplicação, custos de fabricação e aceitação no mercado. Esperamos que a publicação desse trabalho e difusão dessa filosofia de projeto possam contribuir para que mais engenheiros e técnicos os empreguem, produzindo equipamentos melhores em geral, e equipamentos dedicados à aplicações especiais.

AGRADECIMENTOS

Agradecemos a Hique Gomes, atual responsável pelas atividades de gravação com esse sistema e ao próprio Marcello Sfoggia, falecido no dia 30 de Março de 2015, por nos autorizar a operar, estudar seus equipamentos, publicar este artigo e continuar seu trabalho, bem como aos membros de sua família. A Yuri Yefinczuk e às empresas MontaBrasil e TecnoScreen pela ajuda na montagem dos novos circuitos. Este trabalho também foi parcialmente apoiado com recursos do processo 484423/2010-0, edital MCT/CNPq No 014/2010.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Briggs, David. (2012), “A Pair of Wharfedales - The Story of Gilbert Briggs and his Loudspeakers”, IM Publications, 359p.

[2] Kafruni, Simone. (1996) “O Apreciado Som do Dentista Sfoggia - O odontólogo Marcello Sfoggia é o mais requisitado técnico para as gravações de música clássica”, Correio do Povo, Domingo, 15 de dezembro de 1996.

[3] Sfoggia M. (1990), “Marcello Sfoggia talking about his audio equipment in 1990”, http://www.youtube.com/watch?v=yOVVPFRx3bQ, http://www.youtube.com/watch?v=27795_PAkP0.

[4] Johann, Marcelo. (2013), “Mesa de Gravação do Marcello Sfoggia”, http://www.youtube.com/watch?v=uQQ0-7N5fb0 , Acessado em Set/2013.

[5] Analog Devices. OP27 Datasheet - Low noise precision operational amplifier. Disponível em: http://www.analog.com/media/en/technical-documentation/data-sheets/OP27.pdf.

[6] Jensen Transformers, Jensen Twin-Servo 990 Mic Preamp, AS083, 1997. Disponível em: "http://www.jensen-transformers.com/ schematics/". Acesso em 9 de mar. 2015

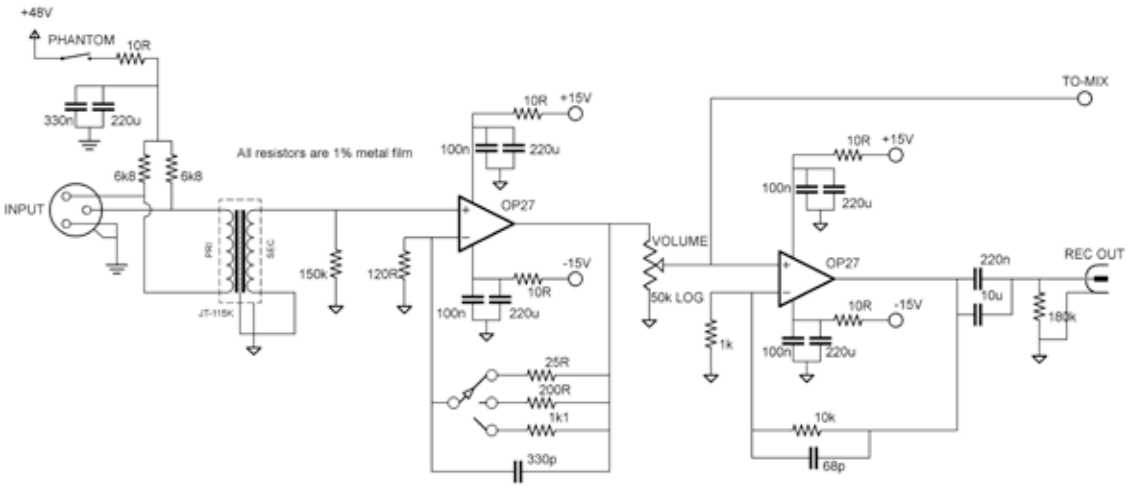


Figura 4 Esquemático do pré-amplificador, caminho de gravação.

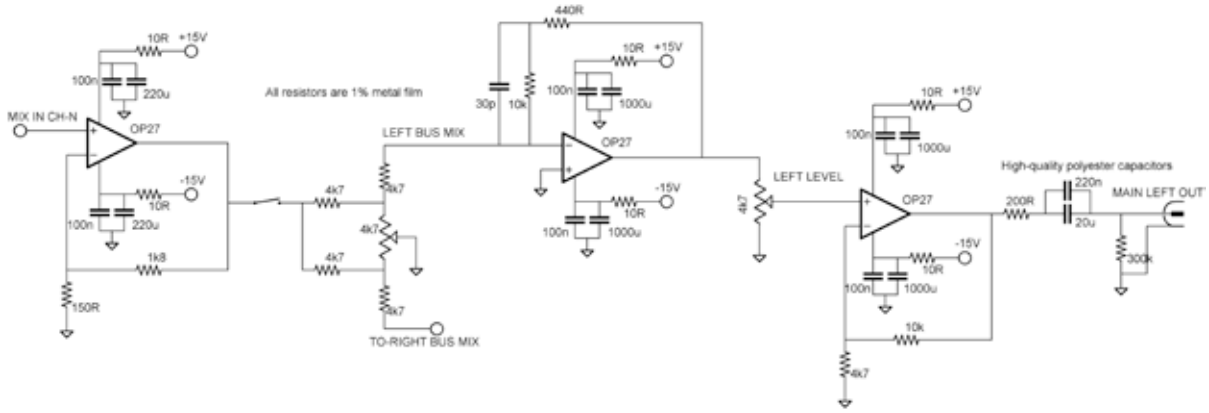


Figura 5 Esquemático do caminho de mixagem, estágios de soma e saída.

Sociedade de Engenharia de Áudio
Artigo de Congresso

Apresentado no 13º Congresso de Engenharia de Áudio
19ª Convenção Nacional da AES Brasil
25 a 28 de Maio de 2015, 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.

Estudo do desenvolvimento de um sistema computacional para taquigrafia musical

Antonio Fernando Penteado[1], José Fornari[1]
[1]Núcleo Interdisciplinar de Comunicação Sonora (NICS)
Universidade Estadual de Campinas (UNICAMP)
Campinas, São Paulo, CEP 13083, Brasil

nando@nandopenteado.com, tutifornari@gmail.com

RESUMO

A notação musical realizada através de recursos computacionais costuma demandar maior tempo de elaboração quando comparada ao modo manuscrito tradicional. Parte disso deve-se ao fato de que as interfaces comuns de entrada (como o mouse e o teclado) costumam limitar a velocidade de inserção de dados musicais, também pela ausência de um método unificado de inserir e editar tais dados entre diferentes ferramentas (softwares e suas versões). Este trabalho visa a apresentar um estudo inicial para a implementação de: 1-) Uma linguagem para representar uma partitura de forma textual, facilitando assim a sua escrita em meio digital, e que utilize o princípio da estenografia para registrar padrões musicais recorrentes de maneira mais rápida e 2-) Um sistema computacional capaz de converter esta representação em formato texto para outros formatos, como por exemplo o MusicXML, propiciando a exportação e compartilhamento de arquivos com outros softwares.

0. INTRODUÇÃO

Pode-se dizer que a notação musical realizada através do computador tornou-se atualmente uma necessidade para o músico, quer seja em termos de sua atuação como compositor, arranjador ou professor de música. A elaboração de partituras e materiais didáticos, através do computador, apresenta diversos benefícios, tais como: correções sem rasuras, impressões de excelente qualidade, separação automática das partes de cada instrumento (no caso de uma grade musical), transposiçãoi mediata, reprodução e distribuição infinita e sem perdas. Tais benefícios são evidentes no que se refere ao processamento e a síntese de dados simbólicos, conforme visto na figura 1. Dados simbólicos referem-se aos dados controle (tal como o MIDI) ou à notação musical [1]. No entanto, a análise de dados simbólicos, referente à sua entrada através de interfaces computacionais ainda é deficitária e normalmente apresenta-se mais lenta do que a inserção manuscrita de notação musical em partitura, de modo tradicional

Figura 1 - Os 3 estágios de tratamento de dados simbólicos.

O foco deste trabalho teórico é esboçar um método que permita minimizar o tempo de inserção de dados simbólicos. Esta pesquisa parte da hipótese de que tal objetivo pode ser alcançado através do estudo e catalogação de padrões simbólicos repetitivos, para que estes sejam organizados numa tabela de fácil acesso pelo usuário, semelhante à um sistema estenográfico [2]. O posterior desenvolvimento de

um sistema computacional que lance mão de recursos de estenografia simbólica notacional, poderá vir a contribuir para acelerar o processo de análise simbólica.

1. ESTENOGRAFIA E MÚSICA

Por estenografia (do grego: *stenos*, apertado ou estreito; e *graphein*, escrever) entende-se aqui com sendo a técnica de escrita abreviada por meio de sinais convencionais, por vezes também chamada de taquigrafia (do grego *thakys*, rápido; e *graphein*, escrever) [3]. O primeiro sistema organizado de estenografia remonta o século I A.C., conhecido como "Notas Tironianas" [4]. Atualmente no Brasil a estenografia é utilizada para documentar sessões legislativas através de transcrições em tempo real, sendo um sistema de registro muito rápido da fala humana.

Segundo [5] música e estenografia envolvem a escrita de sons e o sistema estenográfico moderno coloca fonemas da fala em escalas, que vagamente se assemelham à escalas musicais. No entanto, até onde sabemos, não existe na literatura um método estenográfico para a notação musical. Do mesmo modo que ocorre na catalogação da sonoridade de fonemas, se for possível identificar padrões musicais recorrentes e cataloga-los textualmente, então é possível que estes poderão ser registrados por meio de um método semelhantes àqueles utilizados pelos atuais recursos estenográficos.

2. PADRÕES MUSICAIS

O uso de uma "linguagem de padrões" foi proposto pelo arquiteto Christopher Alexander em sua obra "*A Pattern Language*" como ferramenta para solucionar problemas de design. Segundo [6], cada padrão descreve um problema sempre recorrente em nosso ambiente e portanto descreve a essência da solução para aquele problema. Seus estudos deixaram o escopo da arquitetura para serem adaptados a diversas áreas, incluindo a ciência da computação, como evidencia [10] no artigo em que conclui que padrões não se limitam a um domínio específico.

A identificação e o estudo da ocorrência de padrões musicais dentro de diversos gêneros e estilos de música é um dos objetivos deste trabalho. Exemplificando, a figura 2 apresenta os primeiros compassos do choro "Ande Ligeiro" de Pixinguinha [7]

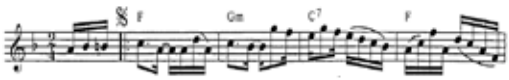


Figura 2 - Exemplo de padrões musicais, extraídos do choro "Ande Ligeiro" de Pixinguinha.

Pode-se dividir a busca de padrões do trecho da figura 2 em três categorias: harmônicos, melódicos e rítmicos.

Padrões harmônicos: o tom do trecho é Fá Maior (F) onde tem-se uma cadência do tipo I IIm V7 I. Desse modo, uma possível catalogação estenográfica para representar este padrão harmônico seria: F1251.

Padrões melódicos: a partir da segunda nota do terceiro compasso inicia-se uma escala de Fá Maior (F) descendente, a partir da segunda maior, caminhando em graus conjuntos até a terça maior (A). Uma possível catalogação estenográfica para este trecho poderia ser: F7-3

Padrões rítmicos: no gênero musical do exemplo (choro) é comum a ocorrência de células rítmicas com tamanho equivalente a uma semínima, se algumas delas forem elencadas e referidas por uma numeração específica, utilizando o sinal menos (-) pra referir que a célula inicia com uma pausa, ou o sinal de mais (+) pra referir que a mesma liga-se à próxima célula, obtém-se um mapeamento como aquele indicado na figura 3. A frase se resumirá à seguinte sequência numérica: -48+48+66444

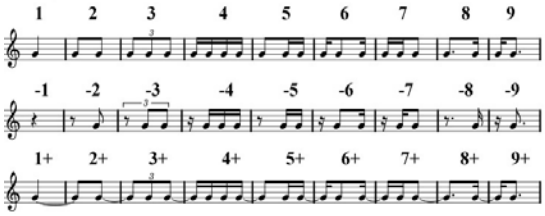


Figura 3 - Possíveis padrões rítmicos de catalogação estenográfica.

Fica assim evidente a grande diminuição de símbolos usados desse modo, para representar o padrão rítmico do trecho musical exemplificado.

3. FERAMENTAS E FORMATOS

Existe atualmente no mercado uma ampla gama de softwares de edição musical. Em janeiro de 2004 a empresa Recordare LLC (que detém os direitos do Finale) criou um formato de arquivos para auxiliar no intercâmbio e compartilhamento de dados entre diferentes softwares, aceitos atualmente pela grande maioria deles como o próprio Finale, Sibelius, Encore, Musescore e mais uma extensa lista de 180 nomes. Este formato é denominado "MusicXML". Este é baseado na linguagem textual XML [8]. Trata-se de um formato muito bem documentado e definido por uma série de DTDs (*Document Type Definitions*) e XSDs (*XML Schema Definitions*) [11].

No caso do método esboçado neste trabalho, é suficiente a elaboração de um sistema em software

que seja capaz de utilizar as ferramentas e formatos já desenvolvidos, recebendo as informações num formato estenográfico como aquele exemplificado acima e exportando os dados simbólicos para o formato MusicXML, de modo a permitir, caso necessário, o processamento e a síntese de tais dados simbólicos em outras ferramentas de software já estabelecidas.

Num primeiro estágio de desenvolvimento a maneira mais conveniente de entrada de dados simbólicos será através de texto sem formatação (ASCII), o que facilitará a sua portabilidade e dispensará o uso de uma interface gráfica, dispensando também o uso obrigatório do mouse que a maior parte dos softwares exige e que eventualmente contribui para dificultar a acessibilidade, como no caso de músicos que sejam deficientes físicos ou visuais.

Num segundo estágio, o sistema poderá ser implementado como um "plugin" a ser acoplado a outras ferramentas de edição musical que aceitem o padrão MusicXML. Isto facilitará a curva inicial de aprendizagem desta ferramenta, uma vez que o usuário estará utilizando uma ferramenta de sua preferência e proficiência.

4. INTERATIVIDADE E ACESSIBILIDADE

A importância deste estudo e de seus desdobramentos vem da necessidade de um processo mais rápido e mais intuitivo de notação musical através do computador.

Para situações como a descrita em [9] de notação musical interativa, em que se combinam a composição generativa através de dados acústicos coletados da execução musical em tempo real intermediadas por um computador, são ainda mais evidentes os benefícios de um método mais ágil de transcrição das ideias musicas em notação formal.

O formato de entrada de dados escolhido para o sistema aqui apresentado (texto sem formatação) dispensa o uso de uma interface gráfica. Isto reforça a importância desta pesquisa nos quesitos de acessibilidade e inclusão digital. Portadores de deficiência visual e/ou deficiência motora, com impossibilidade ou dificuldade no uso do mouse, ficam automaticamente excluídos de produzirem uma partitura eletrônica sem o auxílio de alguém, e mesmo a realização de parte do trabalho é extremamente difícil, como pudemos constatar em visita ao LAB (Laboratório de Acessibilidade da UNICAMP), onde o Prof. Dr. Vilson Zattera, que coordena o LAB e é músico e deficiente visual, nos mostrou passo a passo as fases desde a edição até a impressão de uma partitura em braille (em alto-relevo), operação que pode necessitar de até 5 softwares para ser completada. Sabe-se também que crianças pequenas e idosos tem dificuldade em elaborar uma partitura utilizando o mouse [12]. A entrada textual para dados simbólicos pode vir a constituir-se numa melhoria substancial do desempenho de ambos os grupos etários.

5. DESENVOLVIMENTO METODOLÓGICO

Este trabalho pretende ser desenvolvido baseado nos seguintes critérios: 1) Elaboração de uma linguagem musical textual em forma de estenografia para agilizar o processo de escrita de partituras com foco em músicos brasileiros, sua cultura, sua língua e sua música, uma linguagem que seja rápida para a produção de uma partitura, que seja intuitiva e simples de aprender e que garanta acessibilidade a portadores de deficiência visual e/ou motora. 2) Criação de um software livre capaz fazer a intermediação entre esta nova linguagem e outros softwares existentes no mercado, de modo a garantir a compatibilidade que permita gerar imagens impressas ou eletrônicas, sons, impressão em braille, etc. Esta intermediação ocorrerá pelo intercâmbio de dados através de um arquivo de exportação no formato MusicXML.

As ações para a condução desta pequisa irão se dividir em 5 etapas, a saber: 1) Coleta de dados iniciais: entrevistar um número estatisticamente significativo de músicos profissionais portadores ou não de alguma deficiência visual e/ou motora com o objetivo de: conhecer o(s) software(s) que mais utilizam, identificar suas demandas e necessidades para a edição de partituras e também medir a velocidade com que elaboram uma partitura específica. 2) Com base na análise dos dados coletados, elaborar uma linguagem textual estenográfica. 3) Criação do software livre capaz de converter dados escritos nesta linguagem para o formato music XML. 4) Coleta de dados: testar o uso da linguagem e do software elaborados com músicos iniciantes, profissionais e com portadores de deficiência visual ou motora. 5) Análise de dados, avaliação da performance da linguagem estenográfica e do software, no que diz respeito à velocidade de criação de uma partitura e à sua eficiência nos quesitos acessibilidade e facilidade de uso.

6. CONCLUSÃO

Com base no que foi descrito acima, conclui-se que a notação musical em computadores apresenta algumas dificuldades no que diz respeito à velocidade de elaboração de uma partitura, dificuldades estas inerentes à interface e aos sistemas disponíveis atualmente para a inserção de dados (ideias musicais) e também à curva de aprendizagem dos softwares ou de suas novas versões. Este trabalho apresenta uma pesquisa inicial que pretende diminuir ou resolver tal problema, acelerando o processo de criação com o uso de uma linguagem abreviada (estenográfica) e bastante intuitiva, que também seja capaz de representar textualmente alguns padrões musicais repetitivos, reduzindo o volume de informações a serem inseridas.

Foram aqui apresentadas as bases para o desenvolvimento futuro de um sistema computacional capaz de receber dados em formato de texto e convertê-los para o formato MusicXML, a fim de permitir sua compatibilidade com um amplo conjunto com os softwares da categoria. A opção por um sistema em modo texto terá como benefícios diretos a portabilidade, a acessibilidade e a inclusão digital, favorecendo também o seu uso para pessoas que por ventura tenham dificuldades ou impossibilidades de manusear um mouse (tais como deficientes visuais, deficiência motriz nas mãos, crianças muito pequenas e idosos).

A metodologia empregada será baseada em entrevistas com usuários de software de notação musical para pesquisar suas demandas e registrar a velocidade de inserção de dados em seu software preferido; elaboração de uma linguagem textual estenográfica inclusive apoiada na demanda dos usuários; criação de um modelo computacional capaz de converter dados nesta linguagem para o formato MusicXML; novas entrevistas para apresentar aos músicos a linguagem e o software, e assim medir suas performances no software desenvolvido; analisar o desemprenho da linguagem e do sistema computacional, no que diz respeito à sua velocidade e eficiência.

O resultado pretendido é uma propiciar uma nova ferramenta de software livre para edição rápida de partituras e a criação de uma maneira inovadora de se representar música de maneira textual e abreviada.

Como futuro desenvolvimento deste trabalho tem-se, dentre outras, a possibilidade da inclusão do sistema aqui descrito na forma de módulo de extensão (*plugin*) para o editor de texto LibreOffice, (www.libreoffice.org/), onde será possível propiciar a escrita de partituras de maneira nativa dentro do próprio ambiente deste editor de textos, desse modo facilitando e acelerando o trabalho de elaboração de materiais musicais didáticos e acadêmicos.

7. REFERÊNCIAS BIBLIOGRÁFICAS

1. RAFAILIDIS D., Nanopoulos A., Manolopoulos Y., and Cambouropoulos E. (2008) Detection of stream segments in symbolic musical data. In Int. Conf. on Music Information Retrieval (ISMIR), pages 83–88.
2. FRIDRICH, Jessica; M. Goljan and D. Soukal (2004). Searching for the Stego Key. Proc. SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents VI 5306: 70–82.
3. LAROUSSE. (1999) Grande Dicionário da Língua Portuguesa. Editora Nova Cultural.
4. CURY, Waldir. (1994) Palestra apresentada em Aracaju, no II Encontro Nacional de Taquígrafos, em novembro de 1994. http://www.taquigrafiaemfoco.com.br
5. CRAMER, Alfred William (2006) Material melody, immaterial writing: music and stenography in mid-19th century Germany. Claremont Discourse Lectures – http://ccdl.libraries.claremont.edu/col/cdl/
6. ALEXANDER, C., et al. (1977) A Pattern Language. Oxford University Press.
7. CARRASQUEIRA, Maria José. (1997) O melhor de Pixinguinha. São Paulo: Editora Irmãos Vitale.
8. BRAY, Tim., Jean Paoli, and Charles M. Sperberg-McQueen (eds.) (2000). Extensible Markup Language (XML) 1.0 (Second Edition). World Wide Web Consortium (W3C) Recommendation, October 6, 2000. www.w3.org/TR/2000/REC-xml-20001006
9. FORNARI, José, Igor Maia. (2011) Interactive Notation: Towards a Computer Model for Interactive Musical Composition. Revista EIMAS. http://www.ufjf.br/anais_eimas/ ISSN: 2236-8353. Encontro Internacional de Música e Artes Sonoras (EIMAS) 2011. Universidade Federal de Juiz de Fora (UFJF) - Juiz de Fora - MG. 13 - 16 de Setembro 2011.
10. DUELL, Michael. (1997) Non-Software Examples of Software Design Patterns, Object Magazine, Vol. 7, No. 5, Julho 1997, pp. 52-57.
11. GOOD, M. (2006) Lessons from the Adoption of MusicXML as an Interchange Standard. In: Proceedings of XML
12. MAARSE F.J., L.J.M. Mulder, A.N. Brand, A.E. Akkerman. (2006) Clinical Assessment, Computerized Methods, and Instrumentation (Computers in Psychology). Psychology Press

ARTIGOS
COMPLETOS
FULL
PAPERS

Sociedade de Engenharia de Áudio
Artigo de Congresso

Apresentado no 13º Congresso de Engenharia de Áudio
19ª Convenção Nacional da AES Brasil
25 a 28 de Maio de 2015, 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.

Fan Chirp Transform with nonlinear time warping

Isabela F. Apolinário,[1] Luiz W. P. Biscainho,[1] Martín Rocamora,[2] and Pablo Cancela[2]

[1] Federal University of Rio de Janeiro, COPPE, Electrical Engineering Program
Rio de Janeiro, Brazil

[2] Universidad de la República, Instituto de Ingeniería Eléctrica, Montevideo, Uruguay

isabela.apolinario@smt.ufrj.br, wagner@smt.ufrj.br, rocamora@fing.edu.uy, pcancela@fing.edu.uy

ABSTRACT

This paper proposes an extension of a method for time-frequency analysis of nonstationary harmonic signals: the Fan Chirp Transform (FChT). In its original form, the FChT considers that each fundamental frequency (along with the higher harmonics) of the signal may vary linearly in a period of time. This model, however, may be considered poor for some types of signals, especially those whose fundamental frequencies vary rapidly with time. By allowing quadratic frequency variation, this article presents a solution to this problem, which may be considered the next step of the FChT. The proposed technique is assessed in the context of music signals.

0 INTRODUCTION

Analyzing the frequency content of a signal is one of the most essential operations in Signal Processing. To do that, one usually considers the signal under analysis to be time-invariant, and computes its Fourier Transform. However, this is not appropriate when the signal has rapid fluctuations in frequency, such as the ones produced by pitch variations in speech signals.

Many methods have been proposed to analyse those type of signals, such as the Short-Time Fourier Transform (STFT), the Wigner Distribution, the Wavelet Transform, and so on [1], [2], [3]. Among them, the Fan Chirp Transform (FChT) was introduced in [4], originally devised for speech processing. The goal of this transform is to provide a representation as concentrated as possible of the energy of a harmonic linear chirp in the time-frequency plane.

When it comes to music, it makes no sense to analyze the signal as a whole. Its frequency content is changing with each musical note, which in turn depends on how it is played by the musician. It is important to consider such frequency fluctuations in time in order to have a more precise description of the time-frequency content of a music signal. Because of that, music-

oriented time-frequency transformations, such as sinusoidal modeling [5], [6] and the Constant-Q Transform [7] have become powerful tools. In [8], the FChT was applied to the analysis of music signals by means of the Short-Time Fan Chirp Transform (STFChT).

For some applications, there is a certain interest in exploring the sparsity provided by the FChT, especially in higher harmonics. Works as [9] and [10] take this advantage into account when computing the sinusoidal modeling and extracting the main melody of a music signal, respectively.

The present work still focuses on audio analysis, but allows that the fundamental frequency of the signal varies quadratically in time instead of only linearly as in previous works. This new approach is expected to yield better resolution for signals with rapid frequency fluctuations such as vibratos.

The next section details the FChT and discusses its FFT-based implementation. In Section 2, the FChT with nonlinear warping is introduced. Some experimental results are shown in Section 3 while conclusions and future work are presented in Section 4.

1 THE FAN CHIRP TRANSFORM

This section defines the FChT and briefly discusses its implementation and use in real signals.

1.1 Definition

As said before, the FChT provides an acute representation of harmonically related linear chirp signals. It is described in [8] as

$$X(f,\alpha) \triangleq \int_{-\infty}^{\infty} x(t)\phi'_\alpha(t)\mathrm{e}^{-\mathrm{j}2\pi f\phi_\alpha(t)}dt, \qquad (1)$$

where $\phi_\alpha(t)$ is a linear time warping function given by

$$\phi_\alpha(t) = \left(1+\frac{1}{2}\alpha t\right)t. \qquad (2)$$

Applying the variable change $\tau = \phi_\alpha(t)$ to Equation (1), one obtains

$$X(f,\alpha) = \int_{-1/\alpha}^{\infty} x(\phi_\alpha^{-1}(\tau))\mathrm{e}^{-\mathrm{j}2\pi f\tau}d\tau, \qquad (3)$$

where α is the chirp rate parameter, $\phi_\alpha^{-1}(t)$ is given by

$$\phi_\alpha^{-1}(t) = -\frac{1}{\alpha} + \frac{\sqrt{1+2\alpha t}}{\alpha}, \qquad (4)$$

and one assumes that $x(t) = 0$ for $t \leq -1/\alpha$ to avoid aliasing [4].

From Equation (3), it is possible to see that the FChT is the Fourier Transform of a time-warped version of signal $x(t)$. Therefore, the FChT can be calculated by taking advantage of the fast implementation of the Fourier transform, the FFT algorithm [8].

1.2 Implementation

This work considers the analysis of audio signals as the main application. The fan geometry of the FChT seems appropriate to represent these types of signal, as long as they are essentially composed by tones, each of them consisting of a fundamental frequency and higher harmonics. Nevertheless, this fundamental frequency can be well approximated by a linear chirp only for a short period of time, which forces the FChT to be calculated in consecutive short-time signal frames. This is called the STFChT and can be seen as a generalization of the spectrogram [8].

The first step of an implementation of the FChT is the time warping of each frame of the discrete signal $x[n]$. This step is performed via a nonuniform resampling. Since one only has access to its samples at time instants nT_s, where T_s is the sampling period, an interpolation is carried out [8]. Next, the FFT of the time-warped signal is calculated.

The main difficulty here is to find the appropriate value of α. For doing so, an exhaustive search is performed, where all the admissible values of fundamental frequencies f_0 and chirp rates α are considered. An auxiliary function $\rho(f_0,\alpha)$, called salience plane, is created to help with this task. This function consists on an harmonic accumulation performed for every (f_0,α) pair. If f_0 actually represents an existing fundamental, then the energy at its partials is significant and, therefore, a higher value of $\rho(f_0,\alpha)$ is expected. Likewise, if the correct α value is applied, the sparsity of $\rho(f_0,\alpha)$ is maximum, i.e., the peak value obtained is the higher possible one. For any other α value, the energy of the peak (corresponding to the existing fundamental) would be spread among adjacent bins.

The procedure [8] is briefly described below:

- Many instances of the FChT are calculated for different pre-determined values of α.
- A fundamental frequency grid is defined and, for each f_0 and FChT instance, the sum of the harmonics' log-magnitudes is calculated as [11]

$$\rho(f_0,\alpha) = \frac{1}{n_\mathrm{H}}\sum_{i=1}^{n_\mathrm{H}} \log|S(if_0,\alpha)|, \qquad (5)$$

  where $S(f,\alpha)$ is the FChT and n_H is a predetermined number of harmonics.
- Now, one has a dense plane $\rho(f_0,\alpha)$ that concentrates energy in some points (f_0,α), each of which represents an audio source with fundamental frequency f_0 increasing (or decreasing) at the rate of $|\alpha|$ Hz.
- The highest value of the salience plane $\rho(f_0,\alpha)$ for each f_0 is chosen, yielding a salience function $\overline{\rho}(f_0)$. The peaks of $\overline{\rho}(f_0)$ represent, as said before, audio sources; from them, the chirp rate parameter α for each source is obtained.

In practice, there is no such correct α value, since the proposed model is a first order approximation of the fundamental frequency. Additionally, because of complexity purposes, only a finite number of parameter values is tested. These assumptions introduce small errors to the estimation of α.

It is also important to emphasize that it is only possible to choose one α value to compute the FChT for each time frame. This way, when the signal involves various simultaneous sound sources, it can properly represent one of them at a time, while giving a poor representation for the remaining ones [8].

The desired time-frequency signal representation is provided by the STFChT, which is built as the concatenation of all previously frame-wise computed FChTs. By concatenating the salience functions $\overline{\rho}(f_0)$, one generates a "summarizing" time-frequency representation known as F0gram, which shows the temporal evolution of pitch for all harmonic tones in a music signal [8], and can give some insights about the estimated α values, as will be done in the following sections.

2 THE FCHT WITH NONLINEAR TIME WARPING

The fundamental frequency of a music signal can sometimes present rapid fluctuations in a short period of time. In this case, its approximation by a linear function would be poor, whereas choosing a set capable of representing higher variations in frequency could result in a sparser transformation. In this work, a second-order polynomial is chosen to approximate the fundamental.

Figure 1 (up) shows an example of the ground-truth melody (fundamental frequencies) of an opera excerpt along time, where zero values represent note absences in the foreground melody. It can be noticed that the analyzed signal presents rapid fluctuations in frequency and, as mentioned before, the fundamental frequency may not be well approximated by a linear function when considering short periods of time. By adding a third term to Equation (2) one expects to improve the representation of the fundamental frequency. One has

$$\phi_{\alpha,\beta}(t) = \left(1 + \frac{1}{2}\alpha t + \frac{1}{3}\beta t^2\right) t, \qquad (6)$$

where β is called the curvature parameter.

Note that, by this definition, the instantaneous frequency is given by

$$\nu(t) = f\frac{\mathrm{d}}{\mathrm{d}t}\,\phi_{\alpha,\beta}(t) = (1 + \alpha t + \beta t^2)f. \qquad (7)$$

This shows that the fundamental frequency will now be approximated by a second-order polynomial instead of first-order as before.

This section aims at verifying whether this representation is viable and, then, searching for the best choice of values for parameters α and β. The next step is to add the nonlinear warping to the current FChT implementation and see whether or when it brings an improvement to the time-frequency representation.

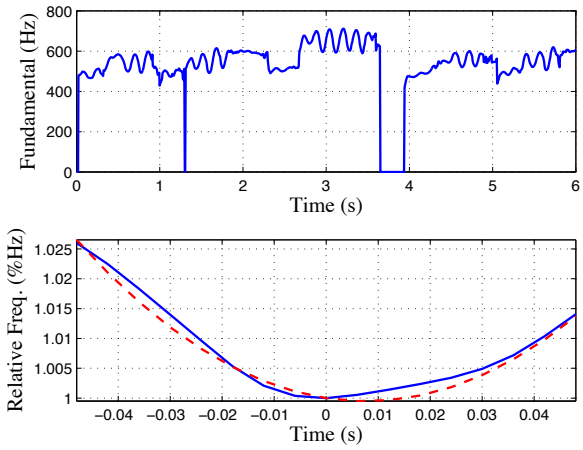


Figure 1: Fundamental frequency of an opera signal (upper graph) and quadratic approximation in dashed red for 100 ms of an audio signal in blue (lower graph).

2.1 Parameter Sampling

A 100-ms frame extracted from the frequency track shown in Figure 1 (up) is zoomed in blue in the bottom plot, and compared with its second-order polynomial approximation drawn in dashed red. Both curves are normalized w.r.t. their frequencies in the middle of the frame ($t = 0$ s).

As mentioned before, for this example a second order approximation seems clearly more suitable. For most signals, though, the fundamental frequency variation may be subtle enough to be approximated by a linear function. The main goal here is, however, to analyze such music signals that in fact exhibit this rapid fluctuations in pitch, like the opera excerpt.

The first step toward the nonlinear warping is to determine which are the possible values for parameters α and β. A database from MIREX [12] containing excerpts of polyphonic audio for which the main melodies' fundamental frequency had been manually labeled (one of which depicted in Figure 1) was employed to aid in this task. Each signal was divided into 100-ms frames[1], and the values of α and β that yielded the best fitting were calculated for each frame. It is important to point out that the fundamental frequency frame should be normalized as shown in Figure 1 (bottom) prior to parameter computation.

The obtained set of pairs (α, β) was then used to construct a histogram. Parameter ranges were set to $[-4, 4]$ for α and $[-50, 50]$ for β, partitioned into 22 bins each. The result can be seen in Figure 2. We see

[1] Since we are interested in following typical pitch variations of audio signals, a time frame way larger than the 20-ms standard was chosen in order to bring forth significant values of β.

that the majority of α and β values concentrate in point (0,0), but there is still a considerable amount of energy around it. From the preferential values for pairs (α, β) depicted in the histogram, different samplings can be proposed, of which three examples are shown in Figure 2. In the first case, the sampling consists of 23 points representing the two main directions of the (α, β) plane: $\alpha = 0$ and $\beta = 0$. In the second case, the sampling consists of the first one added to 12 additional points around the origin $(0, 0)$. In the third case, the sampling consists of the 175-point ellipse around the origin with 90% of the values of α and β. Since an exhaustive search is performed during the computation of the FChT, the number of sampling points is directly related to the computational cost. This fact determined the adoption of restricted ranges for both parameters.

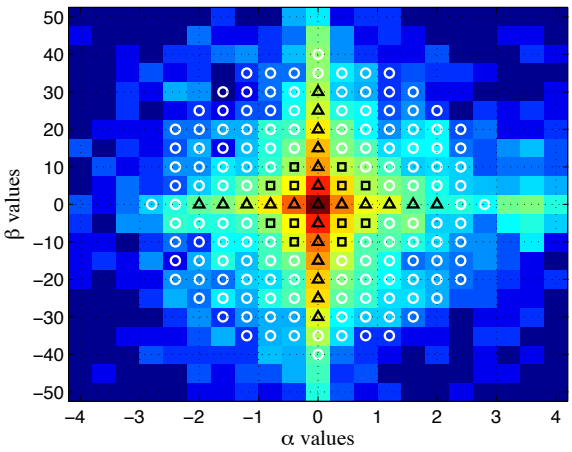


Figure 2: Examples of sampling strategies: 1) the cross denoted by black triangles; 2) the points of the cross plus the ones denoted by black squares; 3) the first two samples plus the points denoted by white circles. The image shows the logarithm of the results.

After choosing an adequate sampling, it is now possible to proceed to the implementation.

2.2 Implementation

The implementation of the FChT with nonlinear time warping also requires an exhaustive search. This time, however, one performs the search over a set of pairs (α, β) instead of just α values. Such pairs can be arranged as an N-element vector, where N is the number of sampling points chosen. If each position γ in this vector represents a pair (α, β), the same method explained in Section 1 can be employed: compute a dense plane $\rho(f_0, \gamma)$ to find the peaks corresponding to audio sources.

3 EXPERIMENTS AND RESULTS

In order to illustrate the effect of the proposed change in the representation, two signals were considered. The first one is a synthetic harmonic signal frequency-modulated by a sinusoid. Its fundamental frequency f_0 is given by the following expression:

$$f_0 = f_1(1 - 2^{1/12}) \sin(2\pi f_2 t) + f_1,$$

where f_1 is the central frequency and f_2, the modulation frequency. The values were chosen to mimic a typical vibrato, as found in singing voice performances, namely, 500 Hz and 6 Hz, respectively. The second signal is the opera excerpt introduced in Section 2, which presents rapid pitch fluctuations with time. For both signals, the sampling frequency is 44100 Hz.

Figure 3 shows the magnitude of the STFChT of the synthetic signal for both linear (second column) and nonlinear (third column) warpings. The STFT for the same signal is also shown (first column). Three different window sizes were chosen for comparison: 1024 samples (first row), 2048 samples (second row), and 4096 samples (third row). This is done in order to enable a fair comparison of the methods.

It is possible to notice the effect of the chosen window size in the resolution of high and low frequencies in the STFT. Increasing the window length yields higher frequency resolution, which provides a better representation for slow varying harmonics, as found in low frequencies. On the other hand, a shorter window is needed to improve the temporal resolution of rapid varying harmonics, which can be seen in the high frequency range. This is a classical issue when dealing with the STFT whose effects are mitigated, as shown, by the STFChT. As a matter of fact, it is desirable to have the largest possible window when using the STFChT, since this means a smaller spread in frequency due to the windowing process. This upper bound is restricted by the chosen warping type. The window length can be increased as long as the frequency variations within the frame can be properly approximated by the warping. For instance, when the window length is increased to 4096 samples, the linear warping is no longer capable of modeling the evolution of the harmonics, and a resolution decrease arises especially in the regions with high curvature. When using the nonlinear warping, one has one more degree of freedom to model the fundamental frequency variations and can, therefore, allow the analysis window to have a larger number of samples. It is important to notice, however, that the computation of the STFChT is dependent on the parameter α (and β, for the nonlinear warping case). They have to be correctly estimated in order to allow a sparse representation.

The analysis obtained with the nonlinear warping, shown in Figure 3, exhibits a sparser representation compared to the linear warping, especially around peaks and valleys of the partial contours. Not surprisingly, the higher order model can approximate with more detail the actual frequency variations. Some artifacts can be observed in this representation due to the estimation of α and β values from a discrete set.

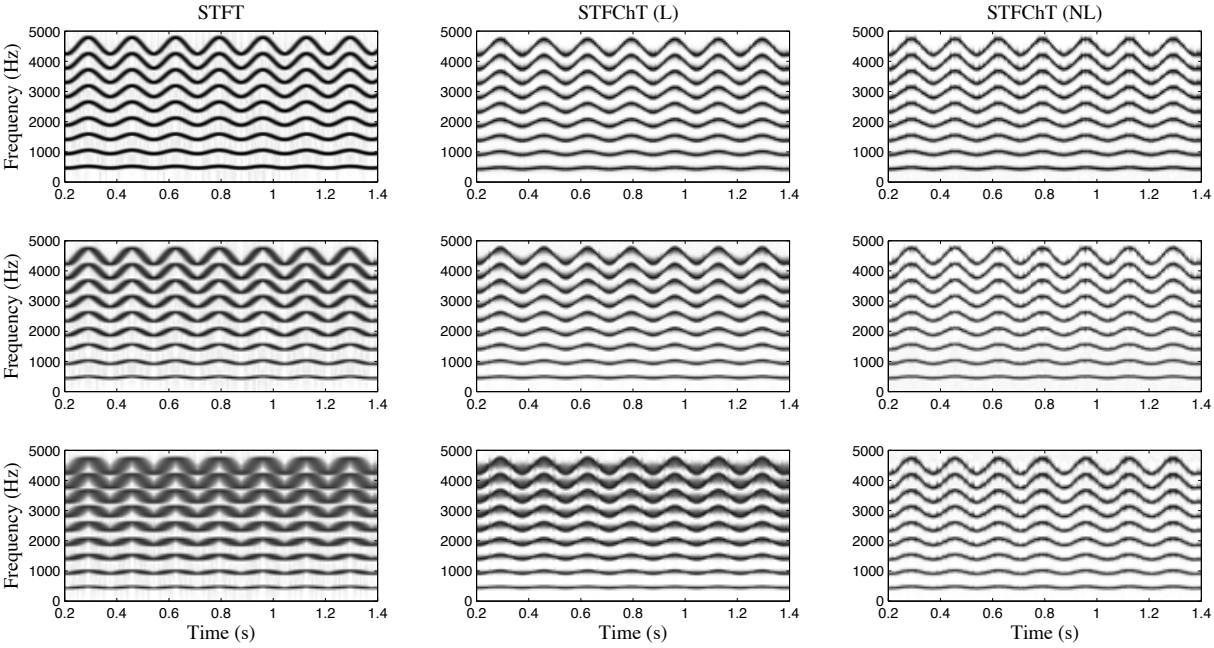


Figure 3: STFT (first column), STFChTs with linear (second column, L) and nonlinear (third column, NL) warpings of a synthetic signal. The following window sizes were used: 1024 (first row), 2048 (second row), and 4096 (third row) samples.

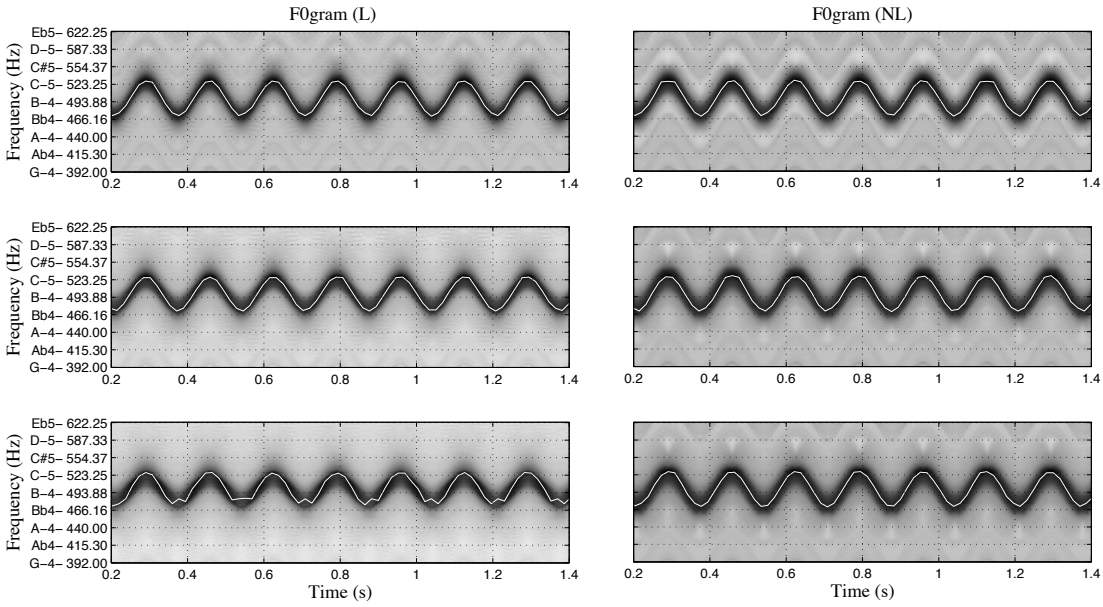


Figure 4: F0gram graphics for the STFChTs with linear (first column, L) and nonlinear (second column, NL) warpings of a synthetic signal. The following window sizes were used: 1024 (first row), 2048 (second row), and 4096 (third row) samples.

Such behavior can also be outlined through the corresponding F0gram graphics, shown in Figure 4. As before, the linear (left column) and nonlinear (right column) warping cases are presented with window sizes of 1024 samples (first row), 2048 samples (second row), and 4096 samples (third row). The fundamental frequency, estimated as the maximum value for each time instant, is shown in white on top of the F0gram. The time-frequency resolution of the representation has a noticeable impact on the accuracy of the fundamental frequency estimation. In particular, the linear warping for a 4096 samples window fails to properly follow the actual fundamental frequency contour.

Figure 5 shows the magnitude of the STFChT of the chosen real music signal for both linear (middle) and nonlinear warpings (bottom). Again, the magnitude

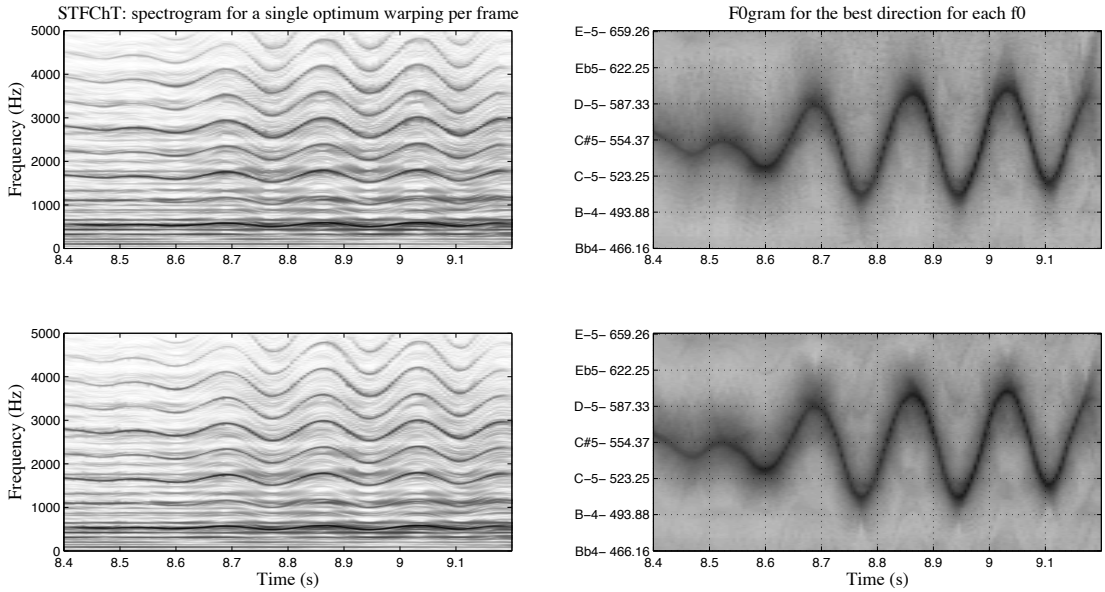


Figure 6: STFChT with linear (higher row) and nonlinear (lower row) warpings of an opera signal.

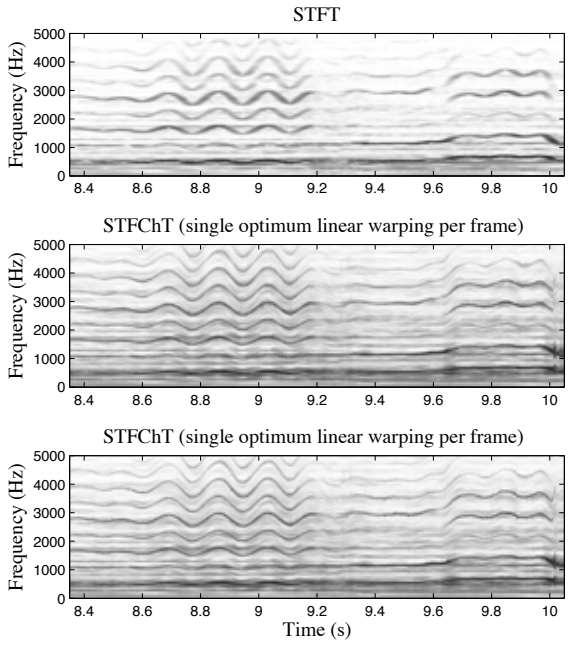


Figure 5: STFT (upper graph) and STFChT with linear (middle graph) and nonlinear (lower graph) warpings of an opera signal.

of the STFT (up) is shown for comparison. The window size used for the these images was 2048 samples, which, considering a sampling frequency of 44100 Hz, correspond to approximately 46.4 ms. Since this signal presents rapid enough pitch fluctuations, a time frame of around half the duration proposed in Section 2.1 was applicable to highlight the main differences between the three representations. As previously stated, it is possible to see that the representation shown in the lower image (STFChT with nonlinear warping) exhibits higher resolution.

Figure 6 shows the F0grams for both the STFChT with linear (up) and nonlinear warpings (bottom). Both methods seem to correctly estimate the fundamental frequency of the signal (not represented here), but it is possible to notice that, for the nonlinear case, the F0gram graphic is slightly sparser.

Another important aspect to consider is the processing time of the nonlinear method. For the nonuniform case and the third type of sampling, for example, since the search is done over a larger amount of samples, the time needed to calculate the FChT is approximately 5 times higher. A further analysis of the complexity is still in progress. This time difference should be taken into account in order to balance this issue and the benefits the nonlinear warping can bring to the representations of signals with fast pitch variations.

4 CONCLUSION

The presented work is an expansion of the formerly proposed FChT that improves the time-frequency resolution of signals that present rapid pitch fluctuations with time.

Among other applications, the FChT can be used in systems for melody detection, denoising, and sound source separation [8]. The proposed modification can improve the performance of the FChT as a result of a finer time-frequency resolution.

It is important to point out that, for complexity reasons, a relatively sparse grid should be employed with the modified STFChT, which can eventually lead to slightly worse resolution in linear segments than the conventional transform with finer α grid would yield. Moreover, the parameter ranges must be blindly chosen, when there is no *a priori* information on the signals. A possible solution to this issue could be guiding

the obtained parameters in an adaptive way, in order to take into consideration only currently plausible values of α and β in the exhaustive search.

It should also be mentioned that the required processing time is higher for the nonlinear warping and thus its application should be restricted to those situations in which fast frequency variations call for better tracking. Future work must include a strong effort to alleviate the computational requirements inherent to the proposed method.

REFERENCES

[1] L. Cohen, *Time-frequency Analysis*, Prentice Hall, Englewood Cliffs, USA, 1995.

[2] P. Flandrin, *Time-frequency/Time-Scale Analysis*, Academic Press, France, 1999.

[3] C. S. Burrus, R. A. Gopinath, , and H. Guo, *Introduction to Wavelets and Wavelet Transforms*, Prentice Hall, Upper Saddle River, USA, 1998.

[4] L. Weruaga and M. Képesi, "The fan-chirp transform for non-stationary harmonic signals," *Signal Processing*, vol. 87, no. 6, pp. 1504–1522, June 2007.

[5] P. R. Cook, *Real Sound Synthesis for Interactive Applications*, A K Peters, Natick, USA, 2002.

[6] 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.

[7] J. C. Brown, "Calculation of a constant q spectral transform," *Journal of the Acoustical Society of America*, vol. 80, no. 1, pp. 425–434, January 1991.

[8] P. Cancela, E. López, and M. Rocamora, "Fan-chirp transform for music representation," in *11th Conference on Digital Audio Effects (DAFx-10)*, Graz, Austria, September 2010.

[9] M. Bartkowiak, "Application of the fan-chirp transform to hybrid sinusoidal+noise modeling of polyphonic audio," in *16th European Signal Processing Conference (EUSIPCO 2008)*, Lausanne, Switzerland, August 2008, pp. 161–166.

[10] Z. Tang and D. Black, "Melody extraction from polyphonic audio of western opera: A method based on detection of the singer's formant," in *15th International Society for Music Information Retrieval Conference (ISMIR 2014)*, Taipei, Taiwan, October 2014.

[11] M. Képesi and L. Weruaga, "Adaptive chirp-based time-frequency analysis of speech signals," *Speech Communication*, vol. 48, no. 5, pp. 474–492, May 2006.

[12] J. Downie, "The music information retrieval evaluation exchange (2005-2007): A window into music information retrieval research," *Acoustical Science and Technology*, vol. 28, no. 4, pp. 247–255, September 2008.

Sociedade de Engenharia de Áudio
Artigo de Congresso

Apresentado no 13º Congresso de Engenharia de Áudio
19ª Convenção Nacional da AES Brasil
25 a 28 de Maio de 2015, 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.

Tempos de reverberação em salas comuns de residências típicas no sul do Brasil: Resultados de medições in-situ

Michael Édison Klein[1] e Stephan Paul[2]

[1] Universidade Federal de Santa Maria, Centro de Tecnologia, Santa Maria, RS, 97.105-900, Brazil

[2] Universidade Federal de Santa Catarina, Centro das Engenharias da Mobilidade
Joinville, SC, 89.218-035, Brazil

michaelklein92@hotmail.com, stephan.paul.acoustics@gmail.com

RESUMO

Os tempos de reverberação de uma sala são os parâmetros mais importantes para avaliação da acústica de ambientes destinados à transmissão da fala ou para apreciar música. Para elaborar projetos de áudio para este tipo de salas o projetista deve conhecer os tempos de reverberação, para adequar o projeto eletroacústico e a acústica da sala caso houver necessidade. Porém, nem sempre é viável realizar medições in-situ na sala objeto do projeto, pois o equipamento para medição é caro e a medição demorada e sensível a ruído de fundo. A fim de conhecer os tempos de reverberação de salas de estar e dormitórios típicos no Brasil foram realizadas medições em 30 salas de estar e 30 dormitórios em residências na cidade de Santa Maria, no sul do Brasil. Verificou-se que nas edificações com idade média de 34 anos, 40% estavam com piso de madeira, não tão utilizado na construção moderna. A mediana do T_{30} em $1\mathrm{kHz}$ para salas de estar com um volume médio de $63,60\,\mathrm{m}^3$ foi de 0,68 segundos. A mediana do T_{30} em $1\mathrm{kHz}$ para quartos com volume médio de $33,76\,\mathrm{m}^3$ foi de 0,49 segundos.

0 INTRODUÇÃO

Em acústica, a caracterização de salas é de grande importância para várias finalidades e pode ser realizada por meio de vários parâmetros, tanto objetivos quanto subjetivos. Uma das características mais importantes em salas é a reverberação, que pode ser quantificada através do tempo de reverberação ou do tempo de reverberação inicial. A importância deste parâmetro reside na possibilidade de poder prever o comportamento do som na sala, e avaliar se as condições para a compreensão da fala ou a apreciação de música são boas. Desta forma, este parâmetro é interessante em

salas destinadas à comunicação através da fala e em ambientes destinados para a escuta de música, compreendendo desta forma desde salas de aula, salas de concerto, salas de teatro e muitos outros, também salas de estar ou outros ambientes residenciais. Ao fazer projetos eletroacústicos como *home-theaters*, *music rooms* etc , destinados para serem executados nestas salas, o projetista também deveria conhecer os parâmetros da sala pré-projeto para decidir se é preciso alguma intervenção na acústica da sala ou algum ajuste no projeto eletro-acústico. Entretanto, a medição dos parâmetros de qualidade acústica, o que inclui os tempos de reverberação, é um processo demorado e no caso do tempo de reverberação muito suscetível a ruído de fundo, principalmente em baixas frequências pois diferença entre NPS produzido pela fonte e o NPS do ruído de fundo deve ser de $60\,\mathrm{dB}$ para determinar T_{60}, $30\,\mathrm{dB}$ para determinar T_{30} e $20\,\mathrm{dB}$ para determinar T_{20}. Enquanto uma diferença de $60\,\mathrm{dB}$ dificilmente pode ser atingida, diferenças de $30\,\mathrm{dB}$ e $20\,\mathrm{dB}$ são mais viáveis a serem atingidas em medições em campo. Ainda assim, altos níveis de pressão sonora em baixas frequências podem resultar em dificuldades na determinação de T_{20} e T_{30}, fazendo com que várias medições precisam ser realizadas para obter dados satisfatórios.

Como os tempos de reverberação em salas dependem da absorção característica das superfícies da sala e dos objetos dentro delas, a possibilidade de estimar o tempo de reverberação de cada faixa de frequência como uma função do volume do ambiente representaria uma otimização de tempo ao realizar em medições acústicas.

Ainda, para simplificar, a norma ISO 10140:2010 [1] assume um tempo de reverberação padrão de $0,5\,\mathrm{s}$ para salas. Este valor corresponde a valores de tempo de reverberação encontrados em estudos realizados na década de 1950 no Reino Unido de salas de estar mobiliados [2] para a maioria das frequências. Tendo em vista que o interesse na determinação do tempo de reverberação em ambientes construídos no Brasil cresceu imensamente, devido à nova norma NBR 15575:2013 [3, 4], que faz considerações sobre as exigências de isolamento sonoro em edificações habitacionais, uma avaliação da situação no Brasil no que diz respeito ao tempo de reverberação é de grande interesse. Considerando que os sistemas construtivos e ambientes internos (pisos, tratamento das paredes) na Espanha e no Brasil são similares, a comparação dos resultados encontrados por Diaz e cols. [5] é indicada.

Assim, o interesse primário do estudo relatado neste artigo é determinar valores típicos para o tempo de reverberação, em termos de T_{20}, T_{30} e EDT a partir de medições em ambientes habitacionais no sul do Brasil.

1 MEDIÇÕES

As medições de tempo de reverberação foram realizadas no sul do Brasil, na cidade de Santa Maria-RS. A determinação do número de residências nas quais os tempos de reverberação e demais parâmetros devem ser medidos deu-se de forma estatística. Considerando-se uma população infinita, o valor mínimo de amostras para medição dos parâmetros de interesse foi de 30 residências (30 salas de estar e 30 dormitórios).

Para as medições do tempo de reverberação em ambientes comuns utilizou-se as recomendações da norma ISO 3382-2:2008 [6] e determinou-se o *Early Decay Time*, o T_{20} e o T_{30} a partir do método do ruído interrompido, utilizando-se o ruído rosa. Os pontos de microfone foram escolhidos conforme determina o método de engenharia, que exige três posições de microfone para cada uma das duas posições de fonte.

Para as medições acústicas, utilizou-se:

- Medidor de Nível de Pressão Sonora, Brüel & Kjaer Tipo 2270,
- Fonte omnidirecional, Brüel & Kjaer Tipo 4292,
- Amplificador, Brüel & Kjaer Tipo 2716,
- Microfone pré polarizado, Brüel & Kjaer Tipo 4189, e
- software Minitab para análise estatística de dados.

A medição é esquematizada de acordo com a figura 1. As medições abrangeram uma faixa de frequência de $80\,\mathrm{Hz}$ à $5\,\mathrm{kHz}$.

Figura 1: Esquema da medição.

2 RESULTADOS

2.1 Características gerais das residências

A amostra total foi constituída por trinta ambientes conforme com características arquitetônicas conforme indicam os exemplos da Figura 2.

Os anos de construção das residências avaliadas na cidade de Santa Maria (RS) variam entre 1970 e 2010. Ao utilizar o método de caixas (Figura 3) para avaliação de valores discrepantes constatou-se que a idade de uma das residências era um valor discrepante. Apesar disso,

Figura 2: Salas de estar e dormitórios.

verificou-se, que seus valores de volume e tempo de reverberação permaneceram próximos aos valores de medidas de tendência central destes parâmetros, o que justifica manter a observação na amostra analisada. Nas residências avaliadas, 40% tiveram pisos de parquet, o que é uma particularidade, já que é mais típica para a década de 1980, enquanto nas últimas duas décadas a preferência tem sido por piso cerâmico. Além destes dois tipos de piso, as residências tiveram pisos de vinílico, carpete e laminado, conforme indica o histograma na Figura 3.

2.2 Tempo de reverberação

2.2.1 Tempo de reverberação em salas de estar

O tempo de reverberação típico em salas de estar no sul do Brasil, varia de acordo com os móveis, elementos decorativos, e superfícies da envoltória que proporcionam absorção sonora e dispersão sonora. As formas e quantidade de elementos de absorção e dispersão variam de uma sala de estar para outra.

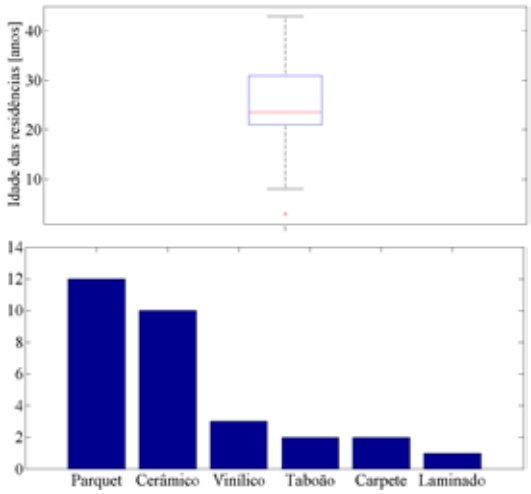


Figura 3: Diagrama de caixas da idade das residências e histograma dos tipos de pisos das residências.

Para todos os parâmetros acústicos, bem como para os parâmetros volume, temperatura, umidade, foram feitas análises estatísticas, utilizando ferramentas da estatística descritiva (histograma, diagramas de caixa, parâmetros de tendência central).

A Figura 4 apresenta os dados de volume, temperatura e umidade das salas de estar em termos de histogramas e diagramas de caixa. A partir da análise dos volumes dos ambientes observou-se, que há três valores discrepantes de volume na amostra (acima do valor extremo superior). Esses valores não se aproximam da mediana de volumes de 63,6 m^3.

As temperaturas encontradas durante as medições estavam dentro da faixa permitida segundo a norma ISO 3382-2:2008 [6] para medições em campo. Apenas em um dos casos o valor 15 °C não foi atingido, sendo que nesta medição a temperatura ambiente era de 12 °C.

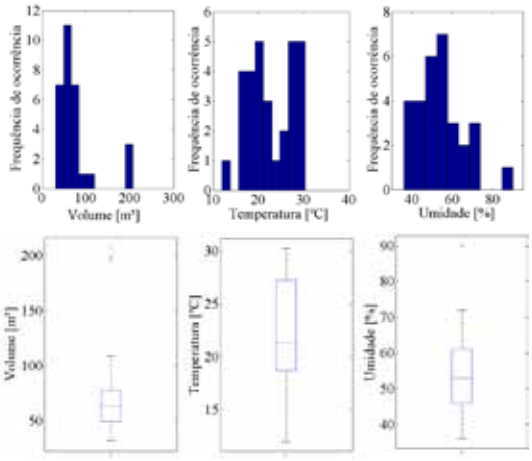


Figura 4: Histograma e diagramas de caixa para as variáveis volume, temperatura e umidade em salas de estar.

A Figura 5 apresenta os dados para T_{20}, T_{30} e EDT encontrados nas salas de estar da amostra. Analisando-se a relação entre amostras que tiveram volume ou umidade muito diferente dos demais (valores discrepantes) observou-se que estas não apresentaram valores discrepantes em termos dos tempos de reverberação, justificando a retenção destas medições na amostra, dando origem às estatísticas descritivas da Tabela 1.

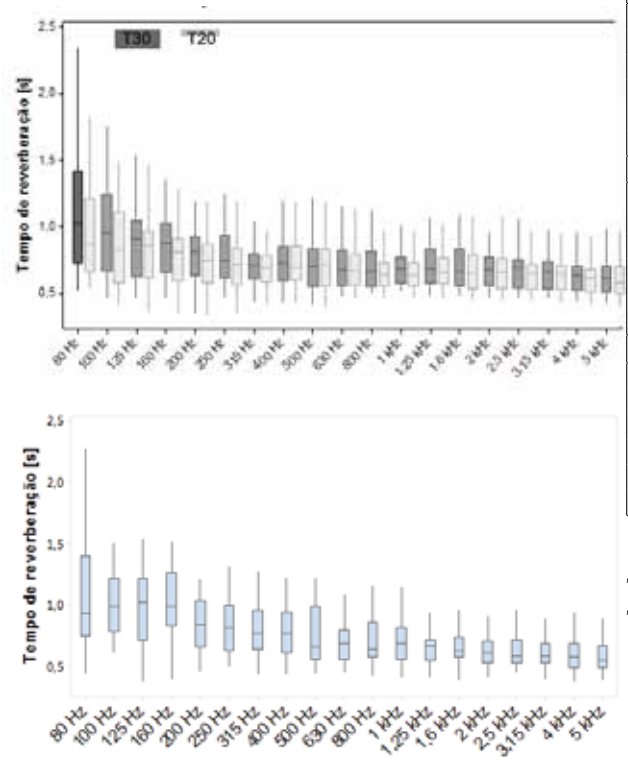


Figura 5: Diagramas de caixa dos dados do T_{20} e T_{30} [diagrama superior] e EDT [diagrama inferior] em salas de estar para cada frequência.

Para determinar se os valores dos parâmetros estavam distribuídos de forma normal, utilizou-se o método de Kolmogorov-Smirnov. Devido a distribuição não-normal dos dados coletados para cada variável, foi utilizado a mediana como medida de tendência central para todos os parâmetros.

As estatísticas dos parâmetros auxiliares Volume, Temperatura e Umidade Relativa encontrados nas medições nas salas de estar encontram-se reunidas na Tabela 2.

Como para todas as demais variáveis, para a análise do tempo de reverberação nas salas de estar, utilizou-se a mediana como medida de tendência central. A Figura 6 apresenta as medianas de T_{20}, T_{30} e EDT para as frequências centrais obtidas nas salas de estar. Para a freuquência central de 1kHz determinou-se $T_{20}(1\,\mathrm{kHz}) = 0,65\,\mathrm{s}$, $T_{30}(1\,\mathrm{kHz}) = 0,68\,\mathrm{s}$ e o

Tabela 1: Média e mediana do T_{20}, T_{30} e EDT das salas de estar.

f central	80	100	125	160	200
T20 média	0,97	0,86	0,82	0,78	0,75
T20 mediana	0,87	0,82	0,86	0,81	0,75
T30 médio	1,14	0,97	0,87	0,86	0,79
T30 mediana	1,07	0,95	0,91	0,87	0,81
EDT médio	1,06	1,01	1,09	1,04	0,87
EDT mediana	0,93	0,99	1,03	0,98	0,84
f central	250	315	400	500	630
T20 média	0,72	0,7	0,72	0,71	0,7
T20 mediana	0,72	0,69	0,7	0,72	0,67
T30 médio	0,78	0,71	0,73	0,72	0,71
T30 mediana	0,74	0,71	0,72	0,7	0,68
EDT médio	0,83	0,84	0,78	0,74	0,71
EDT mediana	0,81	0,76	0,76	0,65	0,69
f central	800	1000	1250	1600	2000
T20 média	0,67	0,67	0,67	0,67	0,67
T20 mediana	0,65	0,65	0,66	0,65	0,66
T30 médio	0,7	0,69	0,71	0,69	0,67
T30 mediana	0,66	0,68	0,68	0,67	0,68
EDT médio	0,7	0,7	0,66	0,66	0,64
EDT mediana	0,64	0,68	0,67	0,63	0,61
f central	2500	3150	4000	5000	$\overline{80\mathrm{Hz} - 5\,\mathrm{kHz}}$
T20 média	0,65	0,64	0,62	0,61	0,75
T20 mediana	0,65	0,65	0,62	0,58	0,75
T30 médio	0,68	0,65	0,63	0,62	0,79
T30 mediana	0,69	0,67	0,63	0,61	0,81
EDT médio	0,63	0,63	0,61	0,59	0,87
EDT mediana	0,59	0,58	0,58	0,54	0,84

Tabela 2: Estatística das variáveis auxiliares Volume, Temperatura e Umidade Relativa em salas de estar.

	Mediana Mediana	Valor mínimo	Valor máximo
Volume [m^3]	63,6	31,7	207,3
Temperatura [°C]	21,3	12	30,3
Umidade relativa [%]	53	36	90
Pé direito [m]	2,62	2,5	2,8

$\mathrm{EDT}(1\,\mathrm{kHz}) = 0,68\,\mathrm{s}$.

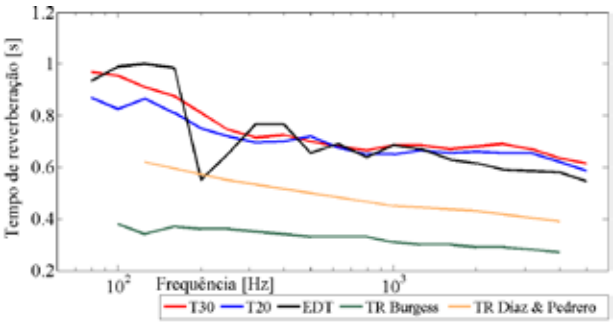


Figura 6: Medianas de T_{20}, T_{30} e EDT em função da frequência em salas de estar.

2.2.2 Tempo de reverberação em dormitórios

Da mesma forma que nas medições em salas de estar mediu-se nos dormitórios o volume, a temperatura,

a umidade e os tempos de reverberação e foram realizadas as análises estatísticas (Figura 7 e 8, Tabela 3).

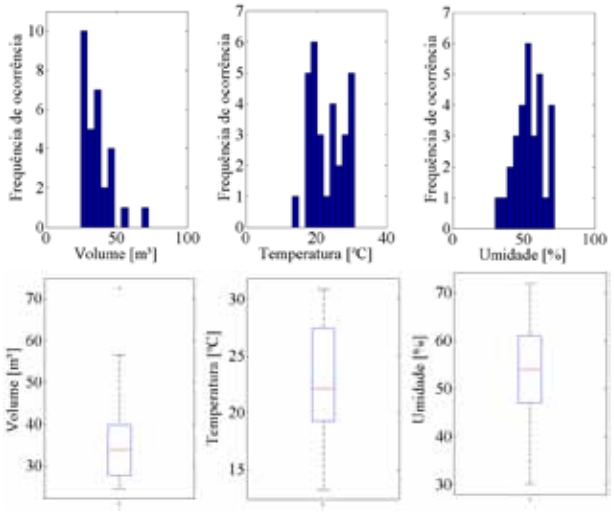


Figura 7: Histograma e diagrama de caixa para os parâmetros auxiliares nos dormitórios.

Tabela 3: Estatística das variáveis auxiliares Volume, Temperatura e Umidade Relativa nos dormitórios.

	Mediana Mediana	Valor mínimo	Valor máximo
Volume[m^3]	33,76	24,5	72,44
Temperatura[°C]	22,1	13,2	30,9
Umidade [%]	44	30	72
Pé direito [m]	2,63	2,49	2,8

Mesmo possuindo um valor discrepante para o volume dos dormitórios e uma medição fora da faixa de temperatura (13,2 °C), não houve valores discrepantes nos tempos de reverberação (Figura 8).

Da mesma forma que para as salas de estar, utilizou-se o teste de normalidade de Kolmogorov-Smirnov para verificar a normalidade da distribuição dos dados de T_{20}, T_{30} e EDT. Observou-se que as distribuições dos dados era não-normais optou-se novamente pela mediana como medida de tendência central. A Figura 9 apresenta as medianas dos parâmetros T_{20}, T_{30} e EDT medidos nos dormitórios em função da frequência. Em 1kHz, constatou-se o $T_{20} = 0,44\,\mathrm{s}$, $T_{30} = 0,49\,\mathrm{s}$ e para o EDT registrou-se $0,42\,\mathrm{s}$.

3 CONCLUSÃO

Para salas de estar concluiu-se que os tempos de reverberação, em 1 kHz, foram $T_{20}(f_c = 1\mathrm{kHz}) = 0,65\,\mathrm{s}$, $T_{30}(f_c = 1\mathrm{kHz}) = 0,68\,\mathrm{s}$ e EDT$(f_c = 1\mathrm{kHz})$ = $0,68\,\mathrm{s}$. O valor $T_{30}(f_c = 1\mathrm{kHz})$ determinado neste estudo é superior ao $T_{30}(f_c = 1\mathrm{kHz}) = 0,45\,\mathrm{s}$ determinado por Diaz e cols [5] para salas de estar em construções típicas na Espanha. Da mesma forma, é

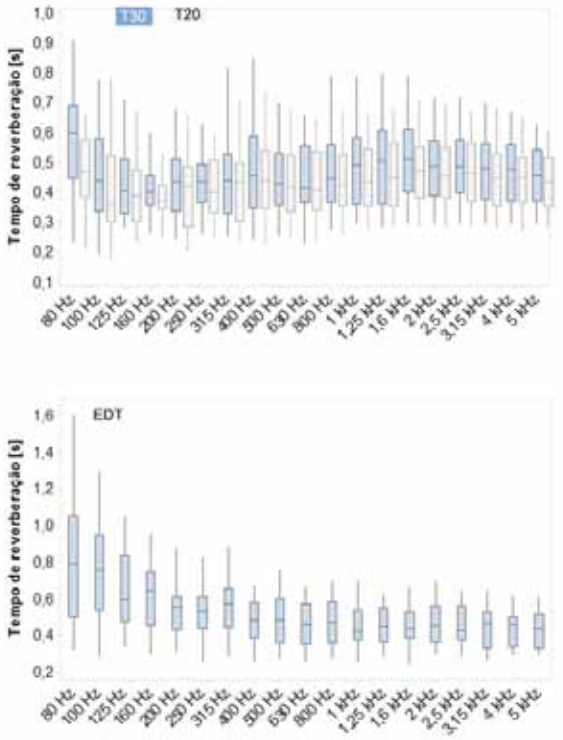


Figura 8: Diagramas de caixa dos dados do T_{20} e T_{30} [diagrama superior] e EDT [diagrama inferior] em dormitórios para cada frequência.

Tabela 4: Média e mediana do T_{20}, T_{30} e EDT dos dormitórios.

f central	80	100	125	160	200
T20 média	0,48	0,4	0,41	0,38	0,4
T20 mediana	0,47	0,36	0,39	0,38	0,42
T30 médio	0,59	0,46	0,43	0,41	0,42
T30 mediana	0,59	0,44	0,4	0,4	0,43
EDT médio	0,81	0,75	0,74	0,64	0,56
EDT mediana	0,79	0,75	0,59	0,64	0,55
f central	250	315	400	500	630
T20 média	0,42	0,43	0,45	0,44	0,44
T20 mediana	0,4	0,44	0,44	0,42	0,41
T30 médio	0,43	0,45	0,47	0,44	0,45
T30 mediana	0,43	0,44	0,45	0,43	0,41
EDT médio	0,54	0,56	0,47	0,48	0,46
EDT mediana	0,53	0,57	0,48	0,48	0,46
f central	800	1000	1250	1600	2000
T20 média	0,44	0,46	0,47	0,47	0,47
T20 mediana	0,43	0,44	0,45	0,47	0,46
T30 médio	0,47	0,49	0,5	0,5	0,49
T30 mediana	0,44	0,49	0,5	0,51	0,48
EDT médio	0,47	0,45	0,45	0,45	0,46
EDT mediana	0,47	0,42	0,45	0,43	0,45
f central	2500	3150	4000	5000	$\overline{80\mathrm{Hz} - 5\,\mathrm{kHz}}$
T20 média	0,47	0,45	0,45	0,44	0,44
T20 mediana	0,47	0,45	0,45	0,44	0,42
T30 médio	0,48	0,47	0,46	0,45	0,45
T30 mediana	0,48	0,48	0,47	0,45	0,43
EDT médio	0,45	0,46	0,44	0,43	0,56
EDT mediana	0,42	0,46	0,46	0,43	0,55

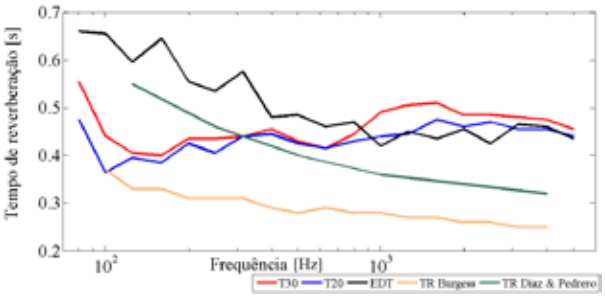


Figura 9: Medianas de T_{20}, T_{30} e EDT em função da frequência em dormitórios.

superior ao $T_{30}(f_c = 1\text{kHz})$ determinado por Burgess [2] no Reino Unido nos anos 1980, que foi de apenas $0,33\,\text{s}$, para uma faixa de volumes similares deste tipo de ambientes. Fazendo a média sobre todas as centrais obtêm-se valores de entre 0,75 e $0,87\,\text{s}$ para os diferentes parâmetros T_{20}, T_{30} e EDT.

Da mesma forma, para os dormitórios, os tempos de reverberação em $1\,\text{kHz}$ foram $T_{20}(f_c = 1\text{kHz}) = 0,44\,\text{s}$, $T_{30}(f_c = 1\text{kHz}) = 0,49\,\text{s}$ e $\text{EDT}(f_c = 1\text{kHz}) = 0,42\,\text{s}$, sendo igualmente superiores aos valores encontrados por Diaz e cols [5] e Burgess [2] que determinaram 0,36s e $0,28\,\text{s}$ respectivamente para uma faixa de volumes similares dos dormitórios. Fazendo a média sobre todas as centrais obtêm-se valores de entre 0,44 e $0,86\,\text{s}$ para os diferentes parâmetros T_{20}, T_{30} e EDT. O valor global de tempo de reverberação (média aritmética dos nas bandas de oitava de 500 Hz, 1000 Hz e 2000 Hz) determinado neste estudo está próximo ao valor indicados na ISO 10140:2010 como tempo de reverberação padrão.

Comparando-se salas de estar e dormitórios observa-se que nas salas de estar o tempo de reverberação diminui com a frequência, conforme esperado, mas aumenta levemente com a frequência nos dormitórios. Já os EDT diminuem com a frequência nas salas de estar e nos dormitórios.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] "ISO 10140-2: Laboratory measurement of sound insulation of building elements - measurement of airborne sound insulation," .

[2] M. A. Burgess, "Reverberation times in British living rooms," *Applied Acoustics*, , no. 18, pp. 369–380, 1985.

[3] Erasmo Felipe Vergara and Stephan Paul, "Introducing the Brazilian standard NBR 15575 regarding the acoustical requirements, criteria and assessment methods for residential dwellings," in *2011*, 2011.

[4] "NBR 155575 edificações habitacionais - desempenho," .

[5] César Díaz and Antônio Pedrero, "The reverberation time of furnished rooms in dwellings," *Applied Acoustics*, , no. 66, pp. 945–956, 2005.

[6] "ISO 3382-2: Mensurement of room acoustic parameters - reverberation time in ordinary rooms," 2008.

Sociedade de Engenharia de Áudio
Artigo de Congresso

Apresentado no 13º Congresso de Engenharia de Áudio
19ª Convenção Nacional da AES Brasil
25 a 28 de Maio de 2015, 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.

Projeto e Construção de uma Interface Gestual para Improvisação Livre

Eduardo Aparecido Lopes Meneses,[1] José Eduardo Fornari Novo Jr.[2] e Marcelo Mortensen Wanderley[3]

[1] UNICAMP - Universidade Estadual de Campinas, Instituto de Artes, Departamento de música
Campinas, SP, 13.083-970, Brasil

[2] NICS - Núcleo Interdisciplinar de Comunicação Sonora (UNICAMP), Campinas, SP, 13.083-970, Brasil

[3] McGill University, Music Research
Montreal, Quebec, H3A 0G4, Canada

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

RESUMO

A utilização de interfaces gestuais na performance de música eletroacústica mista é uma importante frente de desenvolvimento da música contemporânea. A interação entre algoritmos e intérpretes tem auxiliado uma significativa quantidade de artistas a explorarem e a transcenderem algumas fronteiras da performance musical; bem como tem promovido a interação distribuída e a cooperação remota de intérpretes em redes sociais de performances multimodais. Este trabalho apresenta um estudo neste tópico onde sessões de improvisação livre foram realizadas utilizando interfaces gestuais controlando algoritmos criados em software de código livre. A construção do dispositivo, assim como as implicações práticas da utilização de um instrumento aumentado (IA) específico em sessões de improvisação livre são aqui apresentadas, descritas e discutidas.

0 INTRODUÇÃO

Os avanços tecnológicos ocorridos na metade do século XX possibilitaram a utilização de dispositivos eletrônicos na composição e performance musical. Célebres compositores contemporâneos criaram peças seminais do gênero musical atualmente conhecido por música eletroacústica, onde utilizaram artisticamente tal aparato tecnológico. Para citar algumas importantes e pioneiras obras da música eletroacústica, tem-se: Musica su Due Dimensioni (de

Bruno Maderna), Mikrophonie I e II (de Karlheinz Stockhausen), Philomel (Milton Babbitt), Thema: Omaggio a Joyce (de Luciano Berio), Répons (de Pierre Boulez). As primeiras experiências de John Cage e Pierre Schaeffer durante a década de 1950, lançaram as bases da música eletroacústica, abrindo possibilidades de registro, manipulação e criação sonoras que podem tanto ser previamente criadas em estúdio, comumente chamado de tape[1], quanto em tempo real, ou *live electronics*[2]. Tais procedimentos também puderam ser utilizados durante a performance, junto com instrumentos musicais tradicionais, expandindo e transcendendo suas possibilidades acústicas.

No final do século XX e início do século XXI computadores, sensores, transmissores e receptores eletrônicos tornaram-se cada vez mais baratos e ubíquos, passando a serem utilizados amplamente nas mais diversas áreas de atuação humana, o que possivelmente atribuiu a este período contemporâneo o título de "era da tecnologia da informação'. Os computadores são atualmente utilizados para recuperar, transmitir, receber e processar dados em tempo real, formando uma rede informacional que promove a interação cibernética entre humanos e máquinas. Na música, isto passou a significar que o acesso à manipulação da informação acústica de uma performance, anteriormente intangível, agora pudesse ser registrada, copiada, distribuída, editada e processada em tempo real. "Essas possibilidades vão além do repertório de técnicas estendidas, de certa forma já incorporadas na música contemporânea, mas pressupõem o uso inusitado de sons provenientes do instrumento, como no caso do piano preparado, apenas para citar um exemplo já clássico. Com a incorporação de dispositivos eletrônicos surgem os instrumentos expandidos, ou seja, instrumentos musicais cujos mecanismos de produção sonora são ampliados pelo acoplamento de aparelhos eletrônicos"[1].

Este trabalho se concentra no estudo de performances que utilizam simultaneamente instrumentos musicais acústicos e recursos eletrônicos de manipulação sonora em tempo real, como forma de manipular o seu som natural em tempo real, bem como formas de captura de dados gestuais para a parametrização do controle de processos de síntese sonora. O estudo aqui realizado foca na improvisação livre. Neste trabalho, Improvisação é entendida como a elaboração e desenvolvimento de idéias musicais em tempo real, sem necessariamente gerar uma notação musical. Composição é entendida como a criação musical necessariamente baseada no desenvolvimento *off-line* de conceitos musicais formalizados em notação musical, de modo que outros músicos possam estudar esta notação e posteriormente reproduzir uma performance musical similar [2].

A improvisação quase sempre esteve presente em diversas épocas e gêneros musicais, muitas vezes como uma complementação natural da composição, durante a fase de interpretação [3]. No entanto, só mais recentemente, a improvisação tem sido relevada como um significativo campo acadêmico de investigação. Ao longo da história da música, a improvisação tendeu a ocorrer em alguns gêneros musicais específicos, e com o apoio de uma estrutura musical pré-definida, como uma partitura guia. Este é o caso do jazz, onde temos a estrutura formal (*lead sheet*) contendo compassos e progressões de acordes, que são usados pelos músicos jazzistas para improvisar. No caso da música eletroacústica, uma forma de improvisação vem sendo cada vez mais explorada. Trata-se da improvisação livre. Este transcende as fronteiras musicais impostas pelas estruturas composicionais, gêneros ou formas musicais, permitindo assim que um grupo de músicos interaja no sentido de explorar novas possibilidades musicais ou mesmo criando paisagens sonoras. De acordo com Costa [4], alguns conceitos-chave da improvisação livre são encontrados na obra filosófica de Gilles Deleuze, tais como: estratificação, território, plano de consistência, molaridade e molecularidade, corpo sem órgãos, ritmo, meio e ritornelo. Em vez de depender de uma estrutura organizacional pré-definida, a improvisação livre possibilita o surgimento de uma estrutura original, que pode ser temporária (que dure apenas enquanto a improvisação ocorre) ou registada acusticamente (por meio da gravação de seu áudio) ou simbolicamente (por meio de geração automática de notação musical).

1 A COOPERAÇÃO MUSICAL ENTRE INTÉRPRETE E COMPUTADOR

A realização de uma performance de música eletroacústica mista[3] requer a comunicação eficaz entre o intérprete e o dispositivo responsável pela manipulação sonora. Na música eletroacústica mista esta comunicação é normalmente realizada com a ajuda de outro intérprete responsável pelo gerenciamento do computador e que lida com as manipulações de áudio ou o disparo de eventos sonoros. Uma alternativa para esta prática é o uso de dispositivos eletrônicos que coletam e transmitem dados de controle e de áudio entre a interface utilizada pelo intérprete na performance e o computador responsável pelo processamento sonoro.

Estes dispositivos coletam dados gestuais do interprete e parâmetros de controle, realizados durante a performance, em tempo real, possibilitando o controle de diversos processos de manipulação sonora executados por um computador. Tais equipamentos, quando idealizados e construídos para sua utilização com instrumentos musicais acústicos, permitem ao intérprete

[1] Gravações de áudio previamente manipuladas que podem ser tanto disparadas como eventos em uma performance de música eletroacústica mista como constituir a base de uma performance acusmática.

[2] Música para meios ou dispositivos eletrônicos com manipulação sonora em tempo real.

[3] Música que se utiliza tanto de instrumentos acústicos quanto de meios eletrônicos.

a possibilidade de controlar parâmetros do processamento sonoro que serão realizados pelo computador em tempo real. Esta união entre um instrumento musical e dispositivos eletrônicos, ainda que de forma não invasiva, possui resultados semelhantes aos instrumentos musicais aumentados [5]. Para que este resultado seja alcançado, é necessário considerar o que (COOK, 2001) chama de *spare bandwidth*, ou seja, as possibilidades motoras de cada instrumento musical disponíveis durante a performance.

Outro aspecto importante na utilização destes equipamentos é a relação entre a interface física e as possibilidades gestuais do executante. Esta é a ligação musical primordial entre o intérprete e a máquina. Marshall [6] estudou vários modelos para a construção de interfaces de comunicação e DMIs (instrumentos musicais digitais) propostos por Bert Bongers, Marcelo Wanderley, Perry Cook e David Birnbaum. Esse estudo mostrou que tais modelos normalmente consistem de três componentes principais: 1) a interface física (manipulada pelo intérprete), com sensores (que coletam os dados gestuais) e atuadores (que geram um retorno sensorial, ou *feedback*, para o intérprete); 2) o algoritmo de síntese de dados, responsável pela geração em tempo real de áudio, visual e táctil (tal como um *feedback* táctil); 3) o algoritmo de mapeamento que comunica os parâmetros da interface física com a síntese.

Mapear os dados gestuais, transformando-os em parâmetros que controlam o algoritmo de síntese, é fundamental para proporcionar uma interação satisfatória entre o intérprete e a máquina. Existem inúmeras possibilidades de mapeamento de dados gestuais em parâmetros de controle. Cada método se encaixa melhor para uma determinada intenção estética, que é particular a cada compositor e composição. Wanderlei [7] aponta que o gesto musical vai além do movimento das mãos. Um flautista ou clarinetista, por exemplo, pode gerar dados gestuais com o movimento da cabeça e ombros. Estes gestos musicais podem ser vistos como sendo movimentos auxiliares, de acompanhamento [8] ou mesmo movimentos expressivos [9]. Desse modo, um gesto musical não necessariamente tem uma relação direta com a geração sonora [10].

Toda expressão corporal é regulada pelo córtex motor. Esta parte do cérebro é capaz de simultaneamente controlar a movimentação das articulações do corpo, de onde resultam os gestos. Se a capacidade inata do córtex motor puder ser utilizada para o controle em tempo real de modelos de síntese sonora, um intérprete poderia muito bem ser capaz de explorar intuitivamente técnicas estendidas na música contemporânea por meio de instrumentos musicais virtuais cuja possibilidade de geração de novas sonoridades vai muito além das possibilidades dos instrumentos musicais acústicos [11].

O controle de gestos realizado facilmente pelo córtex motor é automaticamente regulado pela percepção, cognição e afeto. Portanto, gestos musicais requerem uma complexa integração de tais habilidades mentais. Observações em estudos de psicologia e neurociência têm sugerido que uma modulação inibitória recíproca, da cognição pela emoção e vice-versa. No entanto, ainda não está claro como os estados cognitivos de fato influenciam as execução de performances musicais. [12].

Uma vez que as performances musicais vão além dos movimentos do instrumentista, para gerar a sonoridade através do seu instrumento, estas são também reguladas por processos mentais cognitivos e emocionais, respectivamente responsáveis pela realização de performances consideradas cognitivas (propensas a serem mais conscientes e atentamente controladas) ou afetivas (propensas a serem mais inconscientes e emocionais), o surgimento de um padrão gestual pode ser observado entre os músicos, não apenas no nível individual, mas também entre músicos de um mesmo instrumento. Alguns destes gestos parecem não estar relacionados com qualquer intenção ou necessidade de gerar uma certa sonoridade, mas são constatados como estando presentes em muitos músicos, o que pode se transformar em uma assinatura gestual que identifica sua fonte; o intérprete [13].

2 A CONSTRUÇÃO DO DISPOSITIVO

A parte prática (o corpo experimental) desta pesquisa foi realizada em duas etapas. A primeira delas foi a criação de um dispositivo para uso com um violão em performances de música eletroacústica mista. Um protótipo deste instrumento aumentado (IA) utilizando interface gestual foi construído e teve sua primeira utilização durante o 1° Colóquio Franco-Brasileiro de Análise e Criação Musicais com Suporte Computacional, realizado em 2014 na sede do NICS[4]. Este foi um evento que reuniu pesquisadores e compositores no campo da música contemporânea que utilizam meios eletrônicos na sua produção, e que apresentaram neste evento, além de suas pesquisas teóricas, também os resultados artísticos dos trabalhos, sob a forma de equipamentos e performances, em sessões de improvisação livre. A Figura 1 mostra o dispositivo montado no violão e utilizado neste caso.

O controlador do IA aqui apresentado foi construído utilizando a plataforma Arduino[5] para conectar os sensores escolhidos e transmitir os dados coletados ao computador. Este microcontrolador programável pode ser visto como um conjunto de componentes intercambiáveis que podem ser utilizados para construir dispositivos para diversas finalidades, que podem executar tarefas diversas como, neste caso, controlar um processo de manipulação sonora em tempo real.

[4] Núcleo Interdisciplinar de Comunicação Sonora da Universidade Estadual de Campinas (UNICAMP).

[5] Arduino é uma plataforma livre de computação física baseado em uma placa de microcontrolador simples, trabalhando em conjunto com um aplicativo utilizado como ambiente de desenvolvimento para escrever algoritmos e enviá-los ao *hardware*. Maiores informações: http://arduino.cc/en/Guide/Introduction.

Figura 1: Violão Aumentado: instrumento, interface gestual e computador.

O software escolhido para realizar a síntese e manipulação sonora foi o Pure Data (Pd)[6].

Um *patch* foi criado em Pd para realizar um processo de síntese aditiva, modulações em amplitude (AM) e freqüência (FM), utilizando sinais senoidais e amostras de áudio capturadas em tempo real. O *patch* também utiliza um compressor de áudio e FFT para a análise espectral em tempo real dos dados coletados de áudio.

Para a construção da interface gestual foi necessária a análise da *spare bandwidth* relacionada à performance violonística. É possível identificar que, apesar da utilização das duas mãos para tocar o instrumento, existe a possibilidade de aproveitamento de movimentos e gestos tanto voluntários quanto involuntários, de ambas as mãos. A opção selecionada foi pela utilização de gestos voluntário da mão direita. Para isto foi utilizado um sensor ultrassônico HC-SR04 montado no corpo do violão (figura 2).

Figura 2: Box contendo sensores acoplados ao instrumento.

[6]Pd é uma linguagem de programação de código aberto desenvolvido originalmente no IRCAM por Miller Puckette e projetado para criar algoritmos (patches) de geração, controle e manipulação de dados de controle, audio e vídeo, em tempo real.

Na prática violonística é usual que o instrumentista execute o instrumento sentado, tornando possível a utilização dos pés na função de disparo de eventos sonoros e processos de manipulação sonora. O coração deste dispositivo (Figura 3) consiste no microcontrolador Frearduino Leonardo R3[7], 5 interruptores momentâneos modelo SPST (interruptores de uma via de acionamento ativo), além de um *display* LCD modelo JHD162a de 2 linhas por 16 colunas, que possui a função de indicar o acionamento dos interruptores e sensores do sistema.

Figura 3: Interruptores, display LCD e microcontrolador do dispoditivo.

Por fim, um sensor inercial (acelerômetro) foi também instalado junto ao violão, possibilitando a captura de deslocamentos do instrumento que foram utilizados no controle da manipulação sonora com o deslocamento do corpo do violão, resultando em uma captação sensível das variações de posicionamento deste instrumento.

A programação do *patch* em Pure Data procurou inserir os procedimentos de manipulação sonora selecionados de forma a permitir a comunicação entre o programa e o dispositivo construído. Houve uma preocupação em proporcionar *feedback* visual, já que o dispositivo não dispunha, no momento, da possibilidade de enviar informação de feedback (do computador para o intérprete).

Há ainda uma rotina desse*patch* que é responsável pela comunicação entre o dispositivo e o computador. Esta é responsável pelo mapeamento dos dados coletados pelo dispositivo e a sua conexão com os parâmetros de cada processamento sonoro.

3 UTILIZANDO O DISPOSITIVO NA IMPROVISAÇÃO LIVRE

A segunda etapa consistiu na execução de *workshops* com sessões de improvisação livre com estudantes de música do Programa Guri[8]. A Figura 5 mostra uma imagem de uma dessas sessões, onde

[7]O Frearduino Leonardo R3 é um dos muitos clones do Arduino Leonardo R3 que apresentam as mesmas características do Arduino original.

[8]O Programa Guri é um programa de educação musical e inclusão social iniciado em 2008 através de uma parceria entre a Secretaria de Cultura do Estado de São Paulo e a Santa

Figura 4: Patch programado em PD.

este grupo se reuniu para improvisar utilizando instrumentos acústicos e o dispositivo desenvolvido, responsável pelo envio de informações do violão ao computador, controlando assim os parâmetros de geração e manipulação sonora em tempo real.

Esta atividade teve duração de 3 horas e foi realizada em 22 de outubro de 2014, reunindo 16 estudantes e 2 professores de diferentes instrumentos musicais, todos do Programa Guri. A maioria dos participantes realizou aqui seu primeiro contato com a improvisação livre, bem como o uso de técnicas de processamento sonoro em tempo real junto com instrumentos musicais acústicos.

Em duas sessões de improvisação os alunos intuitivamente se organizaram em torno de uma frase, ou gesto musical, que surgiu naturalmente durante a atividade, proporcionando unidade à resultante sonora. Algumas interações musicais foram naturalmente criadas entre os participantes, enquanto outras foram previamente combinadas.

Mesmo sem uma estrutura musical formal e pré-definida, foi possível verificar a criação de uma organização interna entre os participantes durante a sessão de improvisação, o que culminou num discurso musical coerente.

Figura 5: Sessão de improvisação livre do *workshop* intitulado Música, Tecnologia e Criatividade.

Estes participantes foram convidados a interagir uns com os outros, criando sonoridades que estes considerassem interessantes ou relevantes. Durante a sessão de improvisação um microfone captava o áudio dos instrumentos musicais utilizados e enviava esses dados para o computador, onde um dos professores poderia manipulá-los através do*patch* programado em Pd, controlado assim este processamento. O som resultante foi reinserido na sessão com a utilização de alto-falantes, compondo a paisagem sonora criada pela sonoridade dos instrumentos acústicos e do instrumento aumentado.

Ao final da atividade os alunos foram requisitados a preencher um questionário sobre esta experiência, e possíveis outras anteriores, com tecnologia aplicada à prática musical e atividades de improvisação, além de registrar a opinião de cada aluno sobre a utilização de computadores no fazer musical, antes e depois do que vivenciaram neste *workshop*. Por fim, os voluntários responderam questões sobre o quanto apreciaram esta atividade. Em cada uma das questões existiam cinco níveis de gradação qualitativa da resposta. Adorei, gostei, não gostei, odiei até "nunca pensei nisto" eram as opções disponíveis nas questões sobre experiência pessoal. Impossível, difícil, fácil, "muito fácil" até "nunca pensei nisto" eram as opções disponíveis nas questões acerca da utilização de tecnologia em música ou sobre improvisação livre. Além disto, em todas as perguntas havia um campo opcional que permitia a cada voluntário complementar a sua resposta de modo descritivo.

Do total de 16 alunos voluntários, 11 deles (68,75%) relataram que adoraram a experiência, enquanto 5 deles (31,25%) disseram que gostaram (nenhum deles selecionou as opções não gostei ou odiei). Também interessante foi verificar como esta atividade mudou o conceito dos alunos sobre o uso de recursos eletrônicos em tempo real na improvisação musical livre. 6 alunos (37,5%) relataram que antes do *workshop* acreditavam que o uso de tais recursos em apresentações musicais era impossível ou muito difícil. No final da atividade este número diminuiu para 2 alunos (12,5%).

Vários estudantes inicialmente classificaram a improvisação como algo inatingível, reservada apenas aos grandes músicos. O *feedback* após a atividade foi bastante positivo e os participantes passaram a ver a improvisação livre com recursos eletrônicos como uma atividade viável, gratificante e agradável.

4 DISCUSSÃO

Levando em consideração o vínculo normalmente formado entre um intérprete e seu instrumento, é interessante conceber a criação de um dispositivo que possibilite a sua utilização com diversos instrumentos, para diversos intérpretes. Uma das premissas assumidas na construção do dispositivo aqui apresentado é que este se tornasse o menos invasivo possível, ou seja, que o instrumento não necessitasse de modificações permanentes em sua estrutura física para utilizar esta interface gestual, podendo, desta forma, ser utilizado em qualquer violão, transformado-o assim num instrumento au-

Marcelina – Organização Social de Cultura (http://www.gurisantamarcelina.org.br/).

mentado. A solução encontrada foi a montagem de uma pequena caixa plástica contendo o acelerômetro e o sensor ultrassônico, que são afixadas ao instrumento musical através da utilização de ventosas.

A interface construída não prevê neste momento a utilização de componentes capazes de fornecer *feedback* ao intérprete. Porém durante a execução de uma performance utilizando este dispositivo é possível acompanhar os diferentes procedimentos sonoros pela tela do computador, além de manipular o *patch* através dos controles do dispositivo ou do computador.

É comum encontrar estudantes de música com bom aproveitamento em seus instrumentos ou nas aulas de canto, mas com pouca ou nenhuma experiência em improvisação livre, ou outras atividades criativas, do mesmo modo que tem pouco ou nenhum contato com a música contemporânea. Muito se fala hoje sobre o desenvolvimento criativo e o papel da música neste desenvolvimento humano. Para vários pensadores do campo da educação musical (como Swanwick, Dalcroze, Gainza, Willems e Kodály) a improvisação é parte fundamental da educação musical e deve ser incentivada nas fases iniciais do ensino musical, o que pode proporcionar aos estudantes iniciantes desenvolvimento de criatividade, maior conforto com o instrumento musical (aproximando-o de uma extensão do corpo do instrumentista), estímulo da audição crítica, proatividade e, no caso da improvisação coletiva, senso de equipe. Em um ambiente de aprendizagem coletiva da música, a improvisação é muitas vezes relegada a um segundo plano, atrasando o aprendizado musical ou mesmo prejudicando a relação do instrumentista com o seu instrumento.

5 CONCLUSÃO

Após a utilização do IA, composto pelo dispositivo gestual e um violão, foram realizadas algumas constatações interessantes. Os movimentos e gestos capturados pela mão e consequentemente braço direito são melhor aproveitados na função de controle de parâmetros e manipulação sonora, enquanto que disparos de eventos sonoros ou mesmo a ativação de determinados processamentos são controlados de forma mais eficiente através de interruptores.

A possibilidade de controle diretamente através do computador mostrou-se interessante quando há mais de um instrumento musical cujo som é manipulado pelo*patch*. Neste caso, a manipulação dos sons do violão podem ser controladas com o dispositivo aqui apresentado, enquanto que os outros intérpretes podem manipular ou disparar outros sons através do*patch*.

Ainda em relação ao *patch* utilizado, a rotina mais importante trata do mapeamento dos sensores e interruptores. Este mapeamento deve ser pensado não somente sob o ponto de vista pragmático, mas também estético e criativo, onde os melhores resultados musicais foram obtidos quando associamos um gesto mapeado a um comportamento esperado de instrumentos acústicos. Um exemplo disso vem da utilização do deslocamento do violão no eixo transverso para controlar a frequência da onda moduladora em uma modulação por amplitude e frequência. Esta associação entre parâmetro e gesto é facilmente assimilada pelo intérprete, principalmente quando se utiliza uma baixa frequência (aproximadamente 20Hz) e obtemos sonoridades típicas respectivamente de vibrato (FM) e *trêmolo* (AM).

Como foi observado, este *workshop* de improvisação livre funcionou como um laboratório onde os alunos participantes experimentaram possibilidades de uso de dispositivos eletrônicos aplicada à música e à improvisação, como ferramentas tangíveis e descomplicadas de expressão. Pesquisas futuras sobre este estudo podem incluir a reestruturação do dispositivo gestual criado a fim de torná-lo mais eficiente em sua utilização conjunta ao violão, além de possibilitar a sua interação com outros instrumentos musicais. Embora o dispositivo atual já possa ser usado com outros instrumentos musicais, este está no momento otimizado para a sua utilização com um violão. Além disso, novas oficinas estão previstas para ocorrer em 2015 e os participantes terão a oportunidade de interagir mais profundamente com a criação de *software* e *hardware*, explorando assim as técnicas e procedimentos de manipulação sonora, onde esses alunos serão capazes de experimentar o papel de protagonistas na produção da música do século XXI.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Fernando Iazzetta, "Performance na música experimental," in *Performa '11 – Encontros de Investigação em Performance*, Aveiro, 2011, pp. 1–10.

[2] José Fornari, "Um processo de criação computacional interativa de notação musical," *V!rus - Revista do Nomads.usp*, pp. 6–8, Dec. 2011.

[3] A. Berkowitz, *The Improvising Mind: Cognition and Creativity in the Musical Moment*, OUP Oxford, 2010.

[4] Rogério Luiz Moraes Costa, "A livre improvisação musical e a filosofia de gilles deleuze," *Per Musi*, , no. 26, pp. 60–66, 2012.

[5] Marcelo M. Wanderley and Philippe Depalle, "Gestural control of sound synthesis," in *Proceedings of The IEEE*. IEEE, Abril 2004, vol. 92.

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

[7] Marcelo M. Wanderley, *Performer-Instrument Interaction: Applications to Gestural Control of Music*, Ph.D. thesis, University Pierre et Marie Curie – Paris VI, 2001.

[8] F. Delalande, *La Gestique de Gould. In Glenn Gould Pluriel.*, Louise Courteau, 1998.

[9] Jane W. Davidson, "Visual perception of performance manner in the movements of solo musicians," *Psychology of Music*, vol. 21, pp. 103–113, 1993.

[10] Marcelo M Wanderley, "Instrumentos Musicais Digitais - Gestos , Sensores e Interfaces," in *Em Busca da Mente Musical*, vol. 60, p. 21 páginas. Editora da Universidade Federal do Paraná, 2006.

[11] José Fornari, "Gestural controlled virtual percussion instruments," in *Proceedings do IX SBGames*, Florianópolis - SC, 8 a 10 de Novembro 2010.

[12] Kodama Higuchi, Marcia K, Cristina M. Del Ben, Frederico G. Graeff, Joao Pereira Leite, and José Fornari, "Reciprocal modulation of cognitive and emotional aspects in pianistic performance," *PLoS ONE*, vol. 6, 2011.

[13] Marcelo M. Wanderley, "Quantitative analysis of non-obvious performer gestures.," *In I. Wachsmuth and T. Sowa, eds. Gesture and Sign Language in Human-Computer Interaction*, pp. 241–253, 2002.

Sociedade de Engenharia de Áudio
Artigo de Congresso

Apresentado no 13º Congresso de Engenharia de Áudio
19ª Convenção Nacional da AES Brasil
25 a 28 de Maio de 2015, 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.

AQUISIÇÃO SIMULTÂNEA MULTICANAL PARA APLICAÇÕES EM ARRANJOS DE MICROFONES

Marcelo Luiz da Costa Ferreira[1] e Ivandro Sanches[1]

[1] Centro Universitário da FEI, Departamento de Pós-Graduação em Engenharia Elétrica
São Bernardo do Campo, SP, 0950-000, Brasil

marcelocosta@fei.edu.br, isanches@fei.edu.br

RESUMO

Arranjos de microfones, ou *microphone arrays*, têm a forte tendência para estar cada vez mais presentes em dispositivos da vida cotidiana, ou seja nos aparelhos de telefonia celular, consoles de jogos, equipamentos militares e aqueles pertencentes ao campo da medicina, como os de implantes auditivos. Este artigo descreve o trabalho de desenvolvimento de uma arquitetura de captura e conversão analógico/digital (A/D) simultânea de 8 canais de microfones para se conectar a um microcontrolador ou microprocessador e que possui um custo abaixo de US$1,000, que é um valor razoável para este tipo de dispositivo comparado com um produto comercial com custo por volta de US $5,000. A escolha dos principais componentes como o TS472 e o ADS8568 permitiu cumprir os requisitos. O sistema implementado possibilita o desenvolvimento de futuras pesquisas em filtragem espacial para localização de fontes sonoras e melhoria da relação sinal-ruído.

INTRODUÇÃO

Este artigo apresenta uma arquitetura de captura e conversão A/D simultânea para 8 canais de microfones com barramento serial ou paralelo e, taxa de amostragem variável para conexão a microcontroladores ou computadores. Esta proposta propicia a confecção de sistema de baixo custo em comparação aos sistemas comerciais equivalentes disponibilizados pelo mercado. A funcionalidade de acoplamento de sistemas (*daisy-chain*) tenderá a multiplicar significativamente o número de topologias a ser estudadas, por exemplo, espaçamento não uniforme entre microfones, configuração tridimensional, etc.

Arranjos ou matrizes de microfones (*microphone arrays*, MA) estão cada vez mais presentes no cotidiano das pessoas e das organizações, deixando apenas de ser domínio exclusivo de uso militar ou de segurança. Os arranjos podem beneficiar pessoas com necessidades auditivas por meio dos implantes cocleares e aparelhos auditivos [1]. A aplicação dos implantes cocleares

já é uma realidade citada por [2] como, por exemplo, a prótese sensorial mais efetiva na história da medicina. O implante bilateral ou binaural, dentre os modelos existentes, é o mais efetivo pela capacidade de localização e discriminação de sons em ambientes ruidosos, ao facilitar ao seu portador o acompanhamento de uma conversação mais próxima do natural.

Telefones celulares utilizam MAs para o cancelamento do ruído ambiente, como demonstrado em [3], ao permitir que a pessoa na outra ponta da chamada telefônica consiga escutar e entender o usuário do aparelho celular, mesmo que ele se encontre em um ambiente ruidoso. O mesmo se aplica aos computadores portáteis (*laptops*) equipados com MA, permitindo direcionalidade na captura dos sons, inibição das fontes sonoras indesejáveis que não estejam posicionadas diretamente à frente do equipamento ou a um azimute programado pelo usuário. O processo de filtragem espacial é realizado pelo processamento digital do sinal de áudio, via algoritmos específicos que necessitam de posicionamento de microfones adicionais nos dispositivos, bem como uma capacidade computacional adequada para o processamento.

Nos campos militar e de segurança, os MAs estão cada vez mais presentes tanto na área de telecomunicações quanto na área de geolocalização, com a evolução dos microprocessadores e do tamanho dos *arrays* os resultados estão mais eficientes e precisos, ao possibilitar, também, a redução do tamanho dos equipamentos, tornando-os possíveis de incorporação em drones e de serem carregados em mochilas. Em [4] foi proposta a utilização de uma aeronave não tripulada, *Unmanned Aerial Vehicle* (UAV), para coleta e processamento de sinais de uma vasta rede de sensores acústicos.

Geolocalização de fontes sonoras como explosões, disparos de armas de fogo e movimentações com produção acústica indevida são exemplos de usos nos campos militar e de segurança. Os MAs são utilizados como radar para a localização da direção e distância das fontes sonoras, como demonstrado em [5], por meio do projeto de uma "bússola sonora". Como exemplo desta utilização cita-se a localização de disparos efetuados por atiradores de elite ou *snipers* na Segunda Guerra do Golfo (2003), no qual um MA detectou a origem do disparos realizados contra os soldados da força de Coalizão permitindo-lhes maior proteção e a execução de suas missões em campo.

Na produção de um *hardware* que tenha um razoável custo $\times$ benefício, sem prejuízo aos resultados, é necessária escolha de uma arquitetura *Digital-to-Analog Converter* (DAC) e subsistemas para suprirem os pré-requisitos exigidos. Os pré-requisitos dependem da destinação dada ao produto final e de sua resolução, seja ele cancelador de ruído ou geolocalizador.

Existem fabricantes deste tipo de *hardware* no mercado no qual a empresa National Instruments é referência constante. Ela é conhecida por linha de produtos de análise e desenvolvimento, como os dispositivos de captura simultânea de diversos canais de entrada. Entretanto os seus elevados custos restringem uma maior difusão dos produtos o que impacta negativamente na produção de conhecimento científico e acadêmico.

Este artigo traz em sua seção metodológica uma introdução aos conversores analógico/digital e tecnologias de amostragem simultânea de múltiplos canais e uma seção de resultados com detalhamento do desenvolvimento do *hardware*. O artigo conclui com a exposição dos benefícios do sistema de aquisição direcionado para o desenvolvimento de pesquisas na área.

METODOLOGIA

Os requisitos básicos para a elaboração deste estudo estão listados a seguir. Os componentes escolhidos estão descritos com mais detalhes na seção de resultados.

- Unidade de captura de áudio ser remota, ou seja, que fosse capaz de ser instalada em diferentes configurações;
- Amostragem simultânea de diversos canais analógicos;
- Conectividade com plataforma de microcontrolador ou microprocessador;
- Capacidade de multiplicação do número de canais com conexão entre módulos, ou seja, *daisy-chain interconnection*.

Para atender ao primeiro requisito é necessária a utilização de um microfone omnidirecional e prever um modelo de transmissão de sinais por fios que seja imune a ruídos. Isto é possível pela transmissão por par de sinais diferenciais.

Segundo [6], o conversor analógico/digital, frequentemente abreviado por conversor A/D ou ADC, é um dispositivo eletrônico capaz de gerar uma representação digital, na forma de números binários, a partir de uma grandeza analógica contínua no tempo, normalmente um sinal representado por um nível de tensão. O processo de conversão A/D é mais complexo e consome mais tempo do que o processo de conversão reversa, ou seja, digital/analógico (D/A). Os ADCs são muito úteis na interface entre dispositivos digitais, tais como, microprocessadores, microcontroladores e dispositivos analógicos e são utilizados em aplicações como leitura de sensores, digitalização de áudio e vídeo. Isto permite que um sistema digital consiga interagir com o mundo real.

A figura 1 apresenta os principais blocos de um conversor analógico/digital, nominalmente:

- Amostrador - Amostragem do sinal em tempo discreto;

- Quantizador - Aproximação do valor de tensão amostrado a um dos 2^N níveis possíveis, por arredondamento ou truncagem, onde N é o númeo de bits do conversor;
- Codificador - Conversão do valor amostrado em um código específico;
- Interface - Conversão série/paralelo e/ou latching

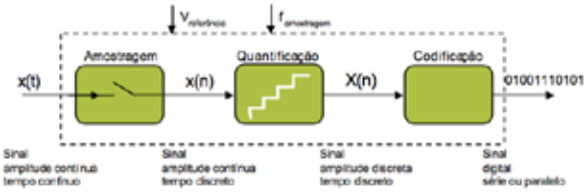


Figura 1: Blocos do conversor analógico/digital ou ADC

A fim de garantir a eficácia da filtragem espacial, é imperativo que a aquisição seja simultânea entre os canais, figura 2. Pela simultaneidade da aquisição, diferenças de fase e correlação entre os canais serão convenientemente utilizadas nos algoritmos, seja para a melhoria da relação sinal-ruído (SNR) do sinal de interesse ou da determinação do azimute da fonte sonora.

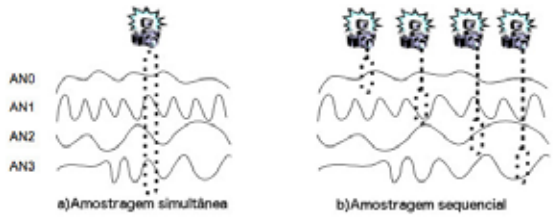


Figura 2: Amostragem simultânea versus amostragem sequencial. Na amostragem simultânea os quatro canais analógicos, AN0, AN1, AN2 e AN3 são capturados ao mesmo tempo. Fonte: National Instruments™

Existem diversas tecnologias que permitem a captura de amostras de maneira simultânea, entre elas, temos:

- Múltiplos canais com Simultaneous Sample & Hold (SSH) multiplexados, figura 3;
- Múltiplos canais com conversores analógico/digital (ADC) dedicados, figura 4;

A principal vantagem da arquitetura de ADCs dedicados é que o processo de conversão de todas as entradas é simultâneo e o tempo consumido neste processo é praticamente o tempo de conversão de um dos canais da arquitetura SSH multiplexados. Este estudo optou

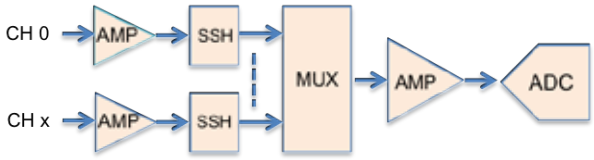


Figura 3: Múltiplos canais com SSH

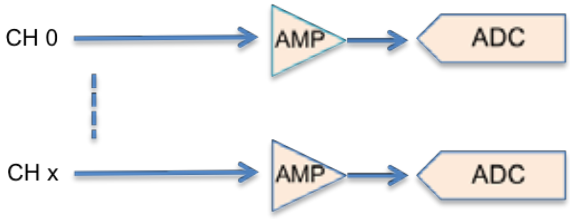


Figura 4: Múltiplos canais com ADCs dedicados

pela escolha da arquitetura de ADCs dedicados principalmente pelo reduzido tempo de conversão proporcionado.

O componente ADC escolhido possui 8 canais analógicos de entrada onde cada canal possui o seu próprio ADC dedicado, barramento de comunicação paralelo e serial e capacidade de conexão *daisy-chain*. Estas características atendem aos requisitos básicos do projeto.

RESULTADOS

A seguir são apresentados os resultados obtidos do estudo para a realização do projeto que atende aos principais requisitos, principalmente a relação custo-benefício, com a seleção de componentes eletrônicos específicos para a realização de cada função e sua integração com uma plataforma de desenvolvimento de baixo custo como o MCF52233DEMO da Freescale. A figura 5 exibe o diagrama geral do sistema desenvolvido e a seguir é descrito com mais detalhes cada bloco ou subsistema deste estudo.

Subsistema de captura e transmissão de áudio

O sub-sistema de captura de áudio é composto de um microfone omnidirecional [7] com sensibilidade de $-44 \pm 2dB$ e uma curva de resposta em frequência conforme o gráfico da figura 6. Outro componente que faz parte deste subsistema é o integrado TS472 [8] da ST-Microelectronics™ que possui um pré-amplificador e um módulo de transmissão através de sinais diferenciais, como visto na figura 7. A utilização de uma configuração diferencial possui um alto *Power Suply Rejection Ratio* (PSRR), figura 8 e, como o módulo de captura fica distante do módulo de processamento dos sinais, a utilização desta configuração de transmissão torna o subsistema de captura imune a ruídos de fontes eletromagnéticas externas que possam ingressar na

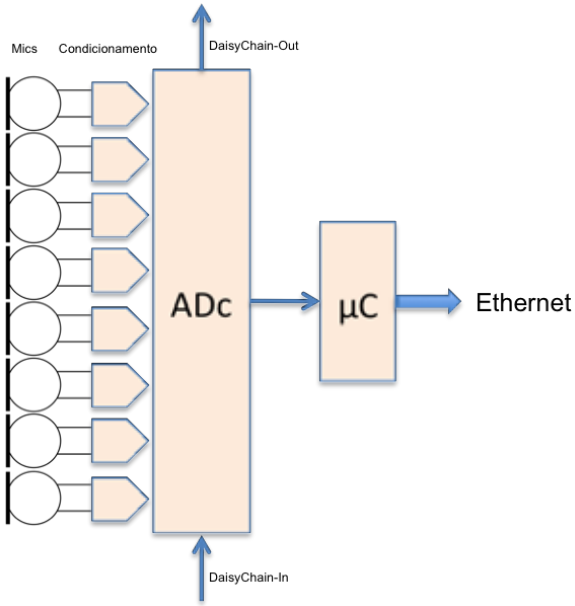


Figura 5: Diagrama em blocos do sistema completo.

linha de transmissão. Este subsistema possui um filtro passa-baixa (LPF) interno de 1ª ordem com o limite de corte superior, F_{CH}, de 40kHz e atenuação de 3dB. É possível reduzir o limite superior de corte com a adição de 2 capacitores externos calculados pela equação 1.

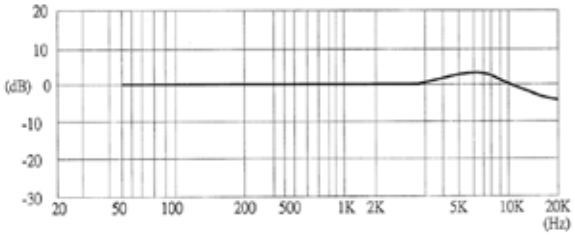


Figura 6: Resposta em frequência do microfone. Fonte: Ficha técnica do componente.

$$F_{CH} = \frac{1}{2\pi * 40 \times 10^3 * (C_{1,2} + 100 \times 10^{-12})} \quad (1)$$

A figura 9 mostra a placa final do subsistema de captura e transmissão de áudio desenvolvida pelo autor.

Subsistema de condicionamento de sinais

O subsistema de condicionamento de sinais, figura 10, é responsável em converter os sinais diferenciais vindo do subsistema de captura e transmissão de áudio em um sinal *single-ended* compatível com a entrada do ADC utilizado. Este subsistema é composto por um amplificador operacional com banda passante compatível com os requisitos do projeto. Um exemplo

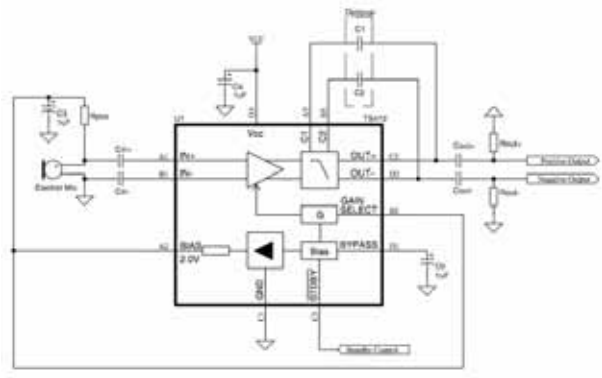


Figura 7: Circuito de captura e transmissão de áudio por par diferencial. Fonte: Ficha técnica do componente.

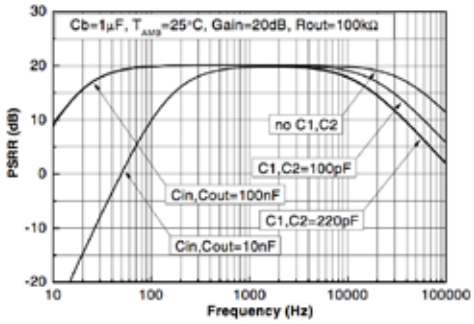


Figura 8: Gráfico de resposta em frequência. Fonte: Ficha técnica do componente.

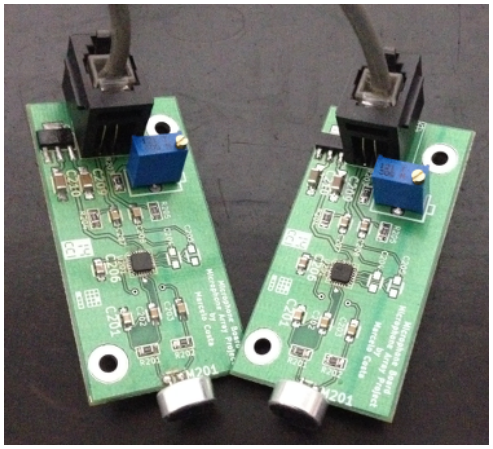


Figura 9: Placa do subsistema de captura produzida pelo autor.

comercial deste componente é o TL072 [9]. Este projeto será utilizado primordialmente para lidar com sinal de voz ou de mais baixas frequências, portanto, um filtro *anti-aliasing* com uma frequência de corte de 8 kHz foi adicionado a este subsistema permitindo frequências de amostragem variáveis a partir de 16 kHz.

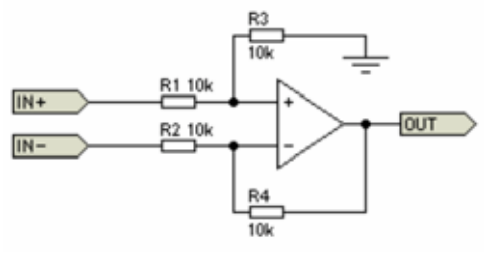


Figura 10: Subsistema de condicionamento de sinais.

A figura 11 mostra a parte na placa final responsável pelo subsistema de condicionamento de sinais desenvolvida especificamente para este projeto.

Figura 11: Vista parcial da placa do subsistema de condicionamento de sinais desenvolvida.

Subsistema de conversão A/D de sinais

O subsistema de conversão A/D é composto pelo circuito integrado ADS8568 da Texas Instruments™ [10], figura 12, e possui a função de amostrar e converter, simultaneamente, os 8 canais de entradas analógicas *single-ended* a uma taxa de amostragem controlada externamente por um sinal digital. Cada grupo de 2 canais possui o seu próprio controle de amostragem o que permite que grupos sejam amostrados a diferentes taxas. A figura 13 ilustra alguns dos sinais envolvidos no processo de amostragem. O tempo de conversão é fixo em 1,7 μs.

Subsistema de *interface* digital

O subsistema de interface digital está construído dentro do próprio ADC utilizado, figura 12, e permite diversos modos de interconexão e sinalização. Este componente possui duas opções de barramentos, paralelo ou serial. O barramento paralelo pode enviar os resultados das conversões a uma alta taxa de transmissão permitindo uma taxa de amostragem na ordem de 510 kSPS (kilo samples per second ou mil amostras por segundo), figura 14. O barramento serial integrado ao ADC possui 4 canais seriais independentes que podem ser utilizados para enviar os resultados dos 4 grupos de entradas simultaneamente para o microcontrolador ou microprocessador ou podem ser utilizados somente 1 ou 2 canais para fazer este envio. A taxa de amostragem que utiliza os 4 canais é de 400 kSPS, este número é reduzido caso se deseje utilizar 1 ou 2 canais somente, figura 15.

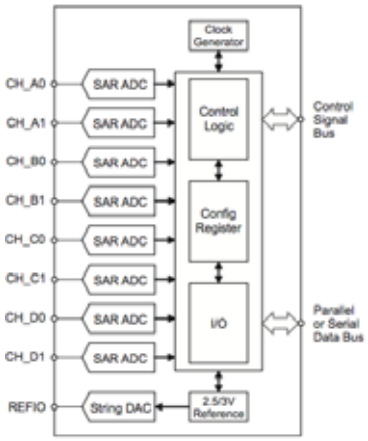


Figura 12: Subsistema de conversão A/D simultânea de 8 canais e interface digital.Fonte: Ficha técnica do componente

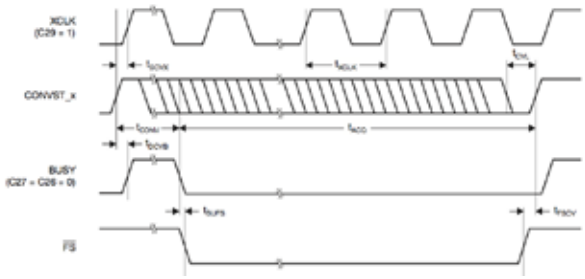


Figura 13: Sinais de controle de amostragem. Fonte: Ficha técnica do componente

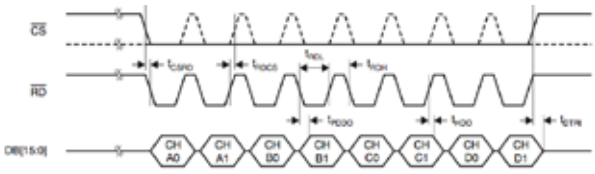


Figura 14: Sinais interface paralela. Fonte: Ficha técnica do componente

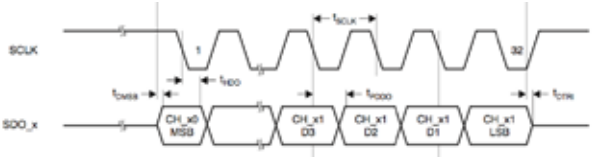


Figura 15: Sinais interface serial. Fonte: Ficha técnica do componente

Ambos modos de interfaces, paralela ou serial, podem ser conectadas a um microcontrolador ou microprocessador. Este estudo utilizou uma placa de desenvolvimento MCF52233DEMO, Freescale, na qual foi

desenvolvido um programa de controle e aquisição das amostras e envio destas por meio de uma interface *FastEthernet* com a utilização do protocolo *User Datagram Protocol* (UDP).

A figura 16 mostra a placa com o subsistema de conversão A/D e o subsistema de condicionamento de sinais.

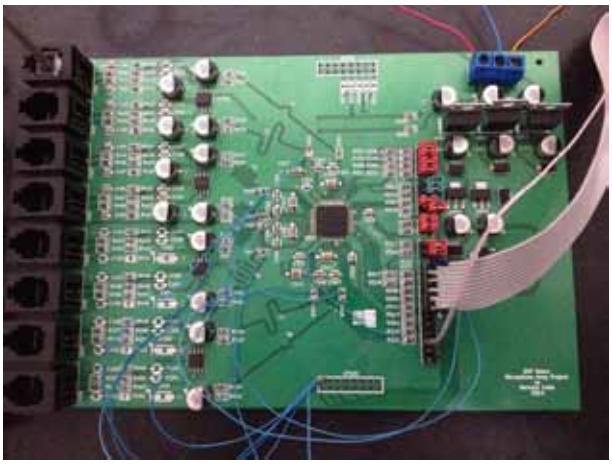


Figura 16: Placa do subsistema de conversão A/D de sinais desenvolvida.

A figura 17 mostra o kit MCF52233DEMO da Freescale utilizado neste projeto.

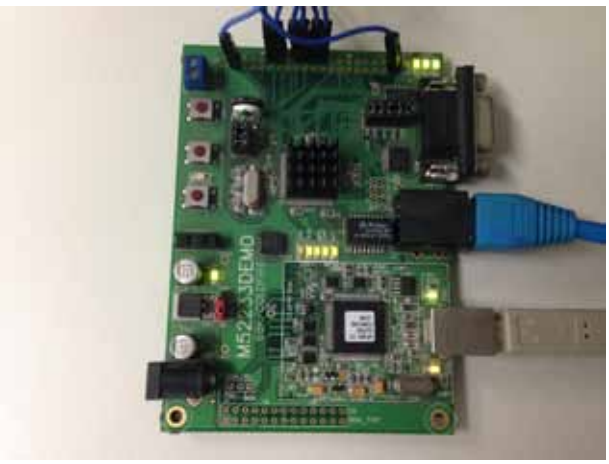


Figura 17: Kit MCF52233DEMO da Freescale utilizado no projeto.

Neste estudo optou-se por utilizar uma frequência de conversão fixa a 16 kHz que é produzida por um microcontrolador PIC16F684 da Microchip por meio de uma saída PWM. A figura 18 mostra o sinal de disparo, de 900ns, em amarelo e o sinal de *busy* (executando a conversão A/D), de 1,6us, em verde. A escolha deste modelo foi para tornar o subsistema de conversão A/D independente do subsistema de aquisição, neste caso o Kit MCF52233DEMO. Desta maneira o autor foi capaz de testar vários módulos de processamento antes da escolha pelo Kit MCF52233DEMO. Na implementação final do projeto pode-se utilizar como sinal de disparo um sinal produzido pelo próprio subsistema de aquisição escolhido. A sincronização do subsistema de conversão A/D e o subsistema de aquisição é feito pelo sinal de *busy* do subsistema de conversão A/D.

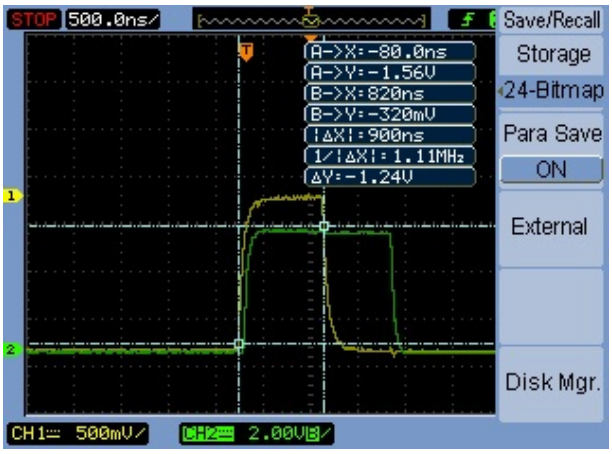


Figura 18: Sinal de disparo em amarelo e *busy* em verde.

Com a escolha da frequência de amostragem de 16 kHz, decidiu-se utilizar a modalidade de transferência entre o ADC e a placa de controle de 4 portas seriais independentes, onde, por cada porta, trafegam os bits de dois canais amostrados. Sendo assim, cada porta transfere 32 bits por ciclo de leitura. Um ciclo de leitura completo dos sinais de disparo, em amarelo, e *clock* de transferência das amostras, em verde, pode ser visto na figura 19.

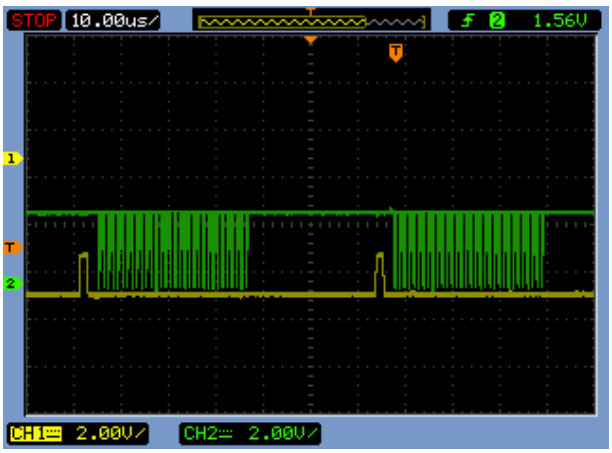


Figura 19: Ciclo de leitura completo.

Em um primeiro instante, para confirmar o funcionamento do sistema como um todo, foram adquiridas amostras de uma fonte sonora a 630 Hz sem uma taxa

de amostragem definida, ou seja, foi utilizado um *loop* para controlar a placa de aquisição à máxima capacidade do subsistema de controle. Para este teste foi utilizada a configuração espacial de microfones ilustrada na figura 20.

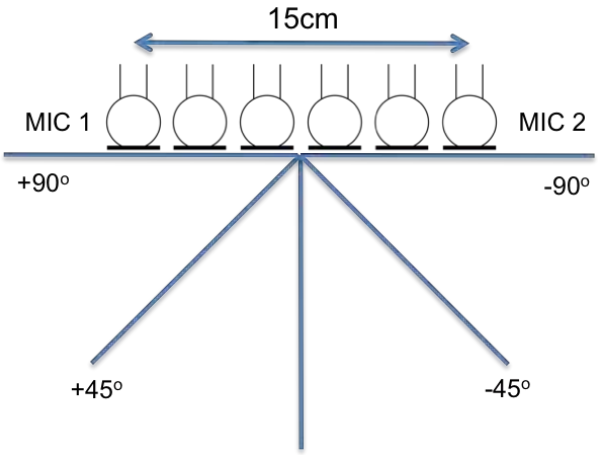


Figura 20: Configuração espacial de microfones utilizada no primeiro teste de funcionamento.

As amostras foram adquiridas com a fonte sonora posicionada a ângulos de $-90°$, $-45°$, $0°$, $45°$ e $90°$. As figuras 21, 22, 23, 24 e 25 exibem as amostras capturadas.

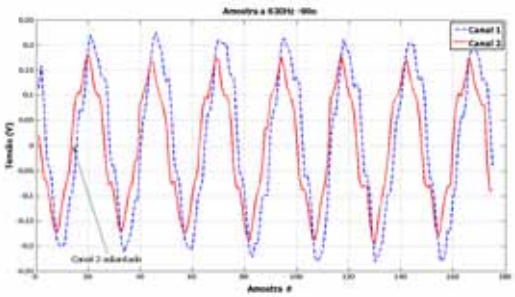


Figura 21: Aquisição de amostras a 630 Hz e $-90°$.

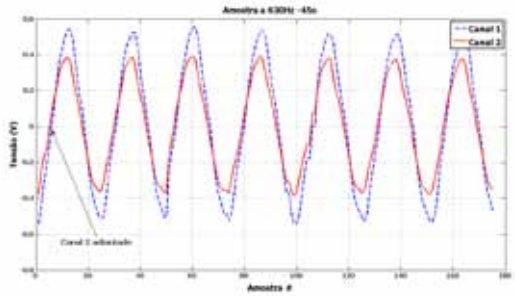


Figura 22: Aquisição de amostras a 630 Hz e $-45°$.

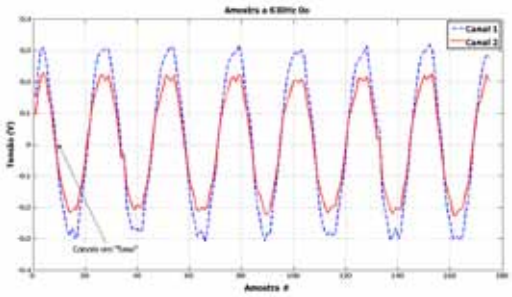


Figura 23: Aquisição de amostras a 630 Hz e $0°$.

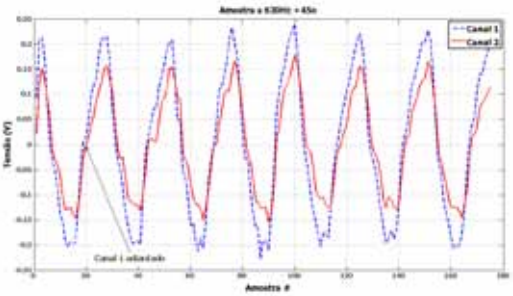


Figura 24: Aquisição de amostras a 630 Hz e $+45°$.

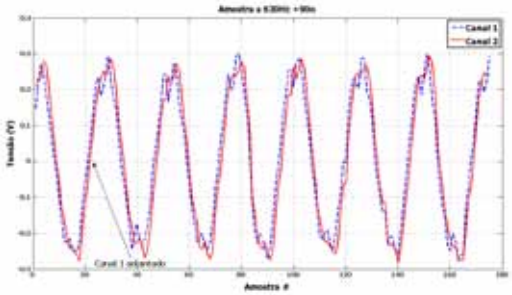


Figura 25: Aquisição de amostras a 630 Hz e $+90°$.

Diversas amostras foram adquiridas e analisadas com o uso do sistema desenvolvido neste projeto e, para isto, foram utilizadas duas configurações espaciais de microfones. As figuras 26 e 27 ilustram essas configurações. A primeira configuração possui uma distância entre microfones de 6 cm e a segunda configuração possui uma distância entre microfones de 50,5 cm. A primeira configuração espacial utilizada não permitiu uma boa visualização da defasagem entre os canais pois, para um sinal de 1 kHz, somente 2 amostras eram capturadas entre a chegada de um sinal sonoro a um microfone e ao seu microfone vizinho. Com a utilização da segunda configuração espacial com 50,5 cm de separação entre os microfones foram adquiridas 36 amostras com a utilização do mesmo sinal.

Na figura 28 é apresentada a resposta ao impulso do ambiente da figura 27 adquirida por um dos canais.

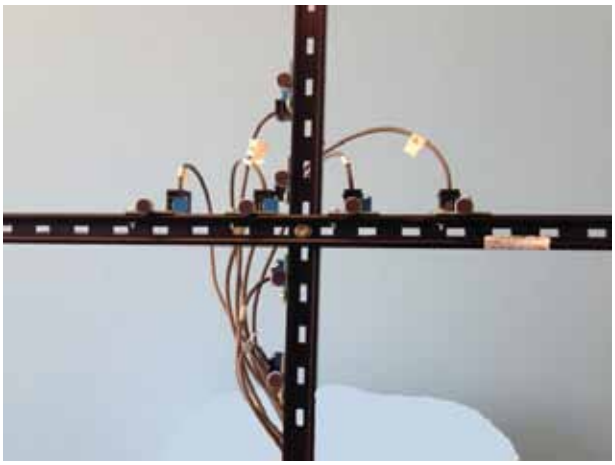


Figura 26: Configuração espacial dos microfones 1.

Figura 27: Configuração espacial dos microfones 2.

Constata-se o alto grau de reverberação desse ambiente.

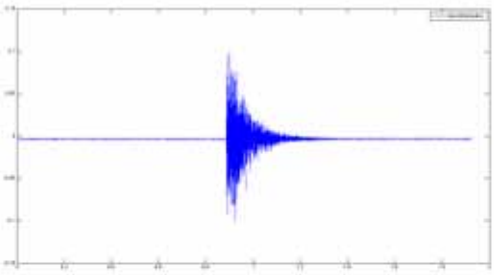


Figura 28: Amostra de um impulso sonoro.

Apesar da alta reverberação, uma primeira análise da defasagem entre canais pode ser feita visualmente conforme a figura 29 onde são exibidas as formas de onda de sinais adquiridos por três microfones sucessivos. As linhas verticais indicam a chegada do mesmo ponto da fonte sonora a cada um dos três microfones utilizados.

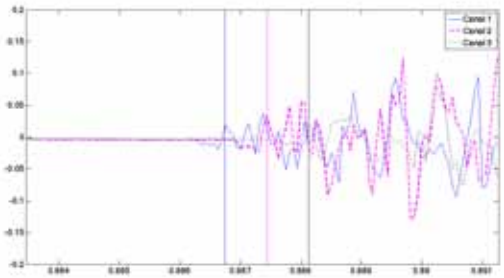


Figura 29: Detalhe da chegada do impulso sonoro.

Outra maneira de analisar a defasagem entre os sinais de dois microfones vizinhos é por meio da convolução de um sinal de referência com a resposta ao impulso de cada canal e verificar as defasagens. Para simplificar esta análise, utilizou-se um sinal senoidal de frequência tal que seu comprimento de onda fosse igual à diferença de caminho percorrido pelo mesmo sinal a um determinado azimute. Para um ângulo de $+45°$ e uma separação entre dois microfones de 50,5 cm, a diferença entre os caminhos é de 35,71 cm. Uma fonte sonora de 1 kHz possui um comprimento de onda de 34,4 cm com a velocidade do som de 34400 cm/s a $20°C$. A diferença resultante de 1,31 cm não pode ser amostrada pois o sistema não possui resolução para tal. Com isso, a convolução de um sinal senoidal com frequência de 1 kHz com os dois sinais amostrados ficam em fase já que a diferença entre eles é de um período completo. A figura 30 exibe o resultado da convolução de cada sinal amostrado.

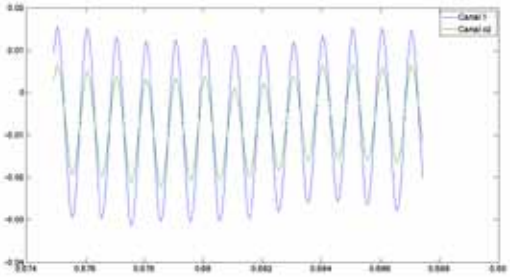


Figura 30: Resultado da convolução de um sinal senoidal de 1 kHz com dois sinais amostrados em microfones vizinhos.

CONCLUSÃO

O sistema desenvolvido é capaz de adquirir, em tempo real, amostras simultâneas de 8 canais analógicos, onde neste estudo foram utilizados microfones de eletreto, mas pode-se utilizar diversos tipos de transdutores analógicos com sinalização diferencial.

A proposta de desenvolvimento de um sistema de baixo custo foi concluída com êxito, pois o sistema possui um custo bem abaixo dos outros sistemas profissionais existentes no mercado, o que irá possibilitar o desenvolvimento da área de processamento de sinais em diversas instituições acadêmicas. Mais detalhes deste projeto podem ser acessados em [11].

Com a utilização de uma interface *ethernet* e protocolo *User Datagram Protocol* (UDP), que é roteável em redes locais ou na Internet, o sistema possibilita a transmissão das amostras em tempo real por meio de redes onde um servidor, regional ou central, concentra os dados obtidos para análise. Esta capacidade também possibilita a descentralização das aquisições, ou seja, é possível ter diversos centros de concentração de dados regionais, por exemplo, com uma posterior agregação em um site central. Um outro uso seria a disponibilização ou troca das informações em tempo real entre centros acadêmicos ou empresas.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Jan Wouters, Simon Doclo, Raphael Koning, and Tom Francart, "Sound processing for better coding of monaural and binaural cues in auditory prostheses," *Proceedings of the IEEE*, vol. 101, no. 9, pp. 1986–1997, 2013.

[2] Miguel Angelo Hyppolito and Ricardo Ferreira Bento, "Directions of the bilateral cochlear implant in Brazil," *Brazilian Journal of Otorhinolaryngology*, vol. 78, no. 1, pp. 2–3, 2012.

[3] Jian Zhang, Risheng Xia, Zhonghua Fu, Junfeng Li, and Yonghong Yan, "A fast two-microphone noise reduction algorithm based on power level ratio for mobile phone," in *Chinese Spoken Language Processing (ISCSLP), 2012 8th International Symposium on*. IEEE, 2012, pp. 206–209.

[4] Daniel J Klein, Sriram Venkateswaran, Jason T Isaacs, Jerry Burman, Tien Pham, João Hespanha, and Upamanyu Madhow, "Localization with sparse acoustic sensor network using uavs as information-seeking data mules," *ACM Transactions on Sensor Networks (TOSN)*, vol. 9, no. 3, pp. 30, 2013.

[5] Jelmer Tiete, Federico Domínguez, Bruno da Silva, Laurent Segers, Kris Steenhaut, and Abdellah Touhafi, "Soundcompass: A distributed MEMS microphone array-based sensor for sound source localization," *Sensors*, vol. 14, no. 2, pp. 1918–1949, 2014.

[6] Ronald J Tocci and Neal S Widmer, *Sistemas digitais: princípios e aplicações*, Pearson Prentice Hall, 8 edition, 2003.

[7] CMA-4544PF-W Datasheet, "CUI inc," *disponível online em http://www.cui.com*, 2008.

[8] TS472 Datasheet, "ST Microelectronics," *disponível online em http://www.st.com*, 2011.

[9] TL072CD Datasheet, "ST Microelectronics," *disponível online em http://www.st.com*, 2001.

[10] ADS8568 Datasheet, "Texas Instruments," *disponível online em http://www.ti.com*, 2011.

[11] Marcelo Costa et al., "Processamento multicanal do sinal de voz," `http://fei.edu.br/~isanches/multichannel.htm`, 2015, [Online; accesso em 06-mar-2015].

Sociedade de Engenharia de Áudio
Artigo de Congresso

Apresentado no 13º Congresso de Engenharia de Áudio
19ª Convenção Nacional da AES Brasil
25 a 28 de Maio de 2015, 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 da Fatoração de Matrizes Não-negativas para a separação de fontes acústicas via Distância Euclidiana Quadrática e Divergência de Kullback-Leibler

Marcelo Borba Santos,[1] Flávia Magalhães Freitas Ferreira[1] e Zélia Myriam Assis Peixoto[1]

[1] Pontifícia Universidade Católica de Minas Gerais, Programa de Pós-Graduação em Engenharia Elétrica, Belo Horizonte, Minas Gerais, Brasil

marcelo.borba.ti@hotmail.com, flaviamagfreitas@pucminas.br, assiszmp@pucminas.br

RESUMO

Este trabalho trata do processamento digital de sinais acústicos. Mais especificamente, visa à separação de sons, gravados ou sintetizados, das peças de uma bateria. Dentre as técnicas disponíveis, utiliza-se o método de Fatoração de Matrizes Não-negativas (NMF - *Non-negative Matrix Factorization*), nas versões supervisionada e não-supervisionada, ambas associadas a diferentes funções de custo, a Distância Euclidiana Quadrática e a Divergência de Kullback-Leibler. São apresentados os fundamentos matemáticos das técnicas envolvidas, resultados de simulações em ambiente computacional e análises comparativas de desempenho por meio de critérios de validação baseados na razão sinal-ruído e na psicoacústica. Por meio dos resultados obtidos, pôde-se concluir pela viabilidade da separação de sons por meio dos métodos selecionados e observar um melhor desempenho para a associação NMF supervisionada e a Distância Euclidiana Quadrática.

0 INTRODUÇÃO

As primeiras investigações sobre a separação de sinais acústicos ocorreram, aproximadamente, em 1954, a partir da proposição de um problema que foi denominado *Cocktail Party*, sobre a capacidade de uma pessoa distinguir as diversas fontes sonoras onde várias pessoas conversam simultaneamente, concentrando-se somente na fonte de interesse [1]. Esse problema foi investigado para o reconhecimento de voz, a partir da leitura de textos com diferentes tipos de estímulos e a utilização de um ou ambos os ouvidos. Buscava-se destacar as características dos sinais que fossem ade-

quadas à implementação do reconhecimento de voz por máquinas, tendo-se concluído pela indicação de técnicas baseadas na teoria da probabilidade e estatística, identificação de fonemas e sintaxe das palavras [1].

Na área musical, a separação de sinais vem sendo utilizada para a extração de informações diversas, como a classificação dos instrumentos contidos em uma música polifônica através de modelos probabilísticos [2]. Mais comumente, os métodos de separação de sinais vêm sendo aplicados à filtragem seletiva de frequências, como em [3], que utilizou o método de Fatoração de Matrizes Não-negativas (NMF) para separar a fala dos ruídos ambiente.

Existem ainda trabalhos que visam, após a extração do som de um determinado instrumento em uma música polifônica, recuperar informações sobre as frequências desse instrumento para a transcrição de sua linha melódica. Em [4], é realizada a separação de sinais pelo método NMF com o objetivo de se identificar as notas musicais contidas no sinal.

As técnicas de separação de sinais podem ser classificadas em relação à disponibilidade de informação prévia ou conhecimento *a priori* de algumas de suas características, como por exemplo, as faixas de frequência utilizadas. Quando não se tem, ou não se usa, informação prévia sobre o sinal, o método é denominada Separação Cega de Fontes (BSS - *Blind Source Separation*. Um exemplo de BSS, por meio do método NMF, é apresentado em [5] para a extração de fontes rítmicas de músicas polifônicas comerciais.

O método de separação não-cega de sinais (ISS - *Informed Source Separation*) é aplicado em [2], com o objetivo de identificar alguns instrumentos, tais como flauta, clarinete, oboé, violino e violoncelo, em músicas polifônicas. Nesse trabalho, as faixas de frequência típicas de cada instrumento são consideradas informações prévias e são utilizados modelos probabilísticos para a identificação de cada instrumento.

Os principais métodos aplicáveis à separação de sinais são indicados a seguir [6][7]:

- Análise de Componente Independente (ICA - *Independent Component Analysis)*;
- Análise de Componentes Morfológicos (MCA - *Morphological Component Analysis*);
- Análise de Componentes por Esparsidade (SCA - *Sparse Component Analysis*);
- Análise de Componente Principal (PCA - *Principal Component Analysis*);
- Fatoração de Matrizes Não-negativas (NMF - *Non-negative Matrix Factorization*);
- Quantização Vetorial (VQ - *Vector quantization*).

Dentre esses, os métodos NMF e ICA vêm se destacando no contexto das aplicações relativas à separação de sinais acústicos.

A qualidade da separação realizada pelo método ICA está condicionada ao número de captadores de áudio (M) e fontes (N), de tal forma que $M \geq N$ [8].

O método NMF, por sua vez, não tem restrições em relação à quantidade de captadores, o que possibilita a aquisição dos sinais de todos os instrumentos de uma bateria apenas com um captador, sem perda de qualidade. Além disso, as regras de atualização do método NMF são mais facilmente parametrizáveis, flexíveis, apresentam maior facilidade para a implementação digital e requerem menor tempo de processamento se comparadas às outras técnicas citadas [9].

Nesse trabalho, os sinais a serem separados foram obtidos do som de uma bateria, composta por um bumbo, caixa e chimbal. As análises utilizam além de sinais sintetizados, como comumente encontrados na literatura atual [4] e [8], também fontes acústicas gravadas. Nesse contexto, tem-se por objetivo principal avaliar os impactos das funções de custo Distância Euclidiana Quadrática e Divergência de Kullback-Leibler sobre o método NMF, nas versões supervisionada e não-supervisionada.

1 FATORAÇÃO DE MATRIZES NÃO-NEGATIVAS

O problema básico a ser solucionado pelo método NMF refere-se ao cálculo de duas matrizes não-negativas, $W \in \mathbb{R}_+^{N\times D}$ e $H \in \mathbb{R}_+^{D\times M}$ que, multiplicadas entre si, resultem em uma matriz $A \in \mathbb{R}_+^{N\times M}$ que seja uma aproximação da matriz original $V \in \mathbb{R}_+^{N\times M}$, ambas não-negativas, conforme a Equação (1).

$$V \approx A = WH \tag{1}$$

Nos casos de separação de sinais acústicos, a matriz V representa a mistura original no domínio tempo-frequência, de tal forma que cada coluna expressa as amplitudes das componentes de frequência do sinal em um determinado intervalo de tempo e cada linha refere-se a uma faixa de frequência. Ou seja, um elemento $v_{nm} \in \mathbb{V}_+^{N\times M}$ representa a amplitude de uma determinada componente de frequência n em um intervalo de tempo m. Aplicada a fatoração, a matriz H irá conter os ganhos relativos às fontes sonoras presentes no sinal original, por sua vez, descritas por meio dos padrões espectrais representados na matriz W [6].

1.1 O Algoritmo NMF

A inicialização do algoritmo NMF requer a definição de algumas ações e parâmetros de entrada, fundamentais ao seu desempenho. São eles:

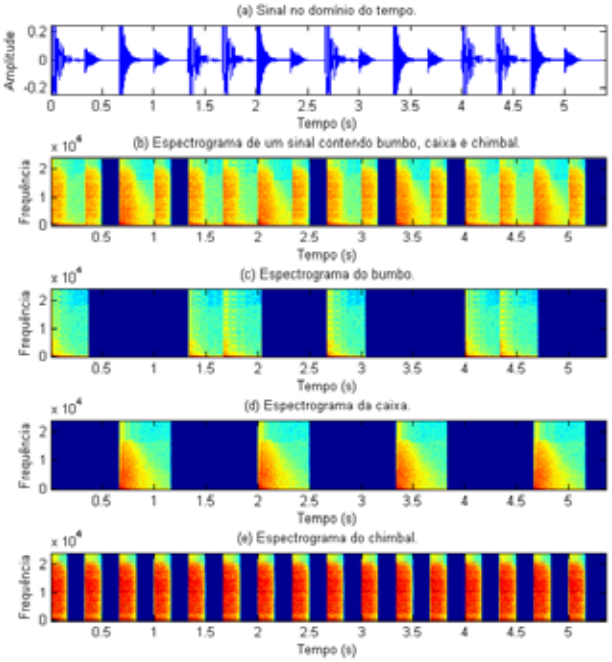
- Construir a matriz não-negativa V;

- Definir o número máximo de iterações;
- Especificar a quantidade de fontes a serem separadas.

A matriz não-negativa V é construída a partir da aplicação da Transformada de Fourier de Curta Duração (STFT - *Short-Time Fourier Transform*) ao sinal de áudio capturado, de forma a se obter o espectrograma do sinal, consideradas apenas as informações sobre as amplitudes das componentes de frequência.

A Figura 1 apresenta um exemplo de um espectrograma dos sons de uma bateria, composto a partir dos instrumentos bumbo, caixa e chimbal. Os gráficos (a) e (b) mostram o sinal no domínio do tempo e seu respectivo espectrograma, isto é, as componentes de frequência presentes no sinal, em cada instante de tempo. Quanto mais as cores se aproximam do vermelho, maiores as amplitudes das faixas de frequências indicadas no eixo vertical. Já os gráficos (c), (d) e (e) apresentam, respectivamente, o espectrograma do bumbo, da caixa e do chimbal.

Figura 1: Sinal e espectrograma dos sons de uma bateria



Em relação ao segundo parâmetro, quanto maior o valor do número máximo de iterações, maior a probabilidade de se obter uma separação mais bem definida das fontes. Porém, haverá o aumento do esforço computacional e, consequentemente, do tempo total de processamento. Vale ressaltar que uma melhor aproximação entre as matrizes V e A não é necessariamente obtida à medida que se aumenta o número de iterações.

O terceiro parâmetro, o número de fontes, recebe um valor inteiro positivo que se refere exatamente ao número de fontes que compõem o som original.

Os passos realizados pelo método NMF para a fatoração da matriz V, conforme [8], são os seguintes:

- Inicializar as matrizes W e H com valores aleatórios, entre 0 e 1.
- Atualizar a matriz W, aplicando-se as regras de atualização a serem apresentadas na Subseção 1.2.
- Atualizar a matriz H, conforme as regras de atualização a serem apresentadas na Subseção 1.2.
- Repetir esse processo até que o número máximo de iterações seja alcançado.

1.2 Regras de Atualização

As regras de atualização são empregadas para minimizar a diferença entre o espectrograma original e o estimado [9].

Lee e Seung (1999) investigaram a aplicação do método NMF para o reconhecimento facial e a análise semântica de textos, comparando-o com as técnicas PCA e VQ. Essa comparação mostrou que o método NMF obteve melhores resultados. Nesse trabalho, foram investigadas duas funções de custo: a Distância Euclidiana Quadrática e a Divergência de *Kullback-Leibler*, apresentadas a seguir [7].

A função de custo baseada na Distância Euclidiana Quadrática é expressa por

$$f(W,H) = \frac{1}{2}||V - WH||_F^2, \tag{2}$$

onde a parcela $||V - WH||_F$ é calculada pela Norma de Frobenius, dada por [10]:

$$||A||_F = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{M} a_{ij}^2} \tag{3}$$

sendo $A \in \mathbb{R}^{N\times M}$. Com base na expressão (3), a Equação (2) pode ser reescrita da forma

$$f(W,H) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M}(v_{i,j} - a_{i,j})^2 \tag{4}$$

A função de custo com base na Divergência de *Kullback-Leibler* é calculada por meio da expressão

$$D(V||A) = \sum_{i=1}^{N}\sum_{j=1}^{M}(v_{i,j}\log\frac{v_{i,j}}{a_{i,j}} - v_{i,j} + a_{i,j}), \tag{5}$$

onde $v_{i,j}$ e $a_{i,j}$ são elementos dos espectrogramas original V e estimado A, respectivamente.

Segundo [9], as duas funções de custo (2) e (5) são convexas[1] em H ou W, mas não em ambas. Logo,

[1] Uma função $f(x)$ é dita convexa em um intervalo $[a,b]$ se, para quaisquer dois pontos x_1 e x_2 em $[a,b]$ e para qualquer λ, $0 < \lambda < 1$, $f[\lambda x_1 + (1-\lambda)x_2] <= \lambda f(x_1) + (1-\lambda)f(x_2)$ [11].

deve-se utilizar um método de otimização para que se possa encontrar um mínimo local. Dentre os métodos existentes, o gradiente descendente é indicado como uma das técnicas que apresenta maior facilidade para a implementação computacional.

O método do gradiente descendente, quando utilizado para minimização do erro ou desvio entre duas grandezas, obtém um decréscimo da função de custo a partir de um ponto inicial, na direção de variação máxima, o gradiente, obtido pela expressão

$$\nabla f = \frac{\partial f}{\partial x_n}, \tag{6}$$

sendo $\frac{\partial f}{\partial x_n}$ a derivada parcial da função f em relação à variável x_n [12]. Com base no cálculo do gradiente da função de custo f, a Equação (7) apresenta, genericamente, o processo de minimização através do método do gradiente descendente

$$x_{n+1} = x_n + \alpha \nabla f, \tag{7}$$

onde α representa é um fator pré-definido, com implicações diretas sobre o número de iterações para a convergência ao valor mínimo local. Esse fator deve ser suficientemente pequeno a fim de garantir essa convergência [9].

Portanto, utilizando-se o algoritmo do gradiente descendente sobre uma função de custo, seja a Distância Euclidiana Quadrática ou a Distância de Kullback-Leibler, obtém-se a regra para atualização da matriz W apresentada na Equação (8)

$$W \leftarrow W + \alpha \frac{\partial E}{\partial W}, \tag{8}$$

onde E é a função de custo selecionada.

As regras de atualização das matrizes W e H, utilizando o gradiente descendente sobre a Distância Euclidiana, são dadas por:

$$W \leftarrow W \otimes \frac{VH^T}{WHH^T}, \tag{9}$$

$$H \leftarrow H \otimes \frac{W^T V}{W^T W H}, \tag{10}$$

nas quais $\otimes$ representa o produto de Hadamard. No produto de Hadamard, as matrizes devem ter as mesmas dimensões e os seus elementos são multiplicados ponto a ponto [13] da forma,

$$c_{ij} = a_{ij} b_{ij}. \tag{11}$$

A divisão também é realizada ponto a ponto, isto é, o elemento c_{ij} é resultado da divisão a_{ij}/b_{ij}, conforme a Equação (12):

$$c_{ij} = a_{ij}/b_{ij} \tag{12}$$

Para a função de custo Divergência de Kullback-Leibler, as regras de atualização das matrizes, através do gradiente descendente, são dadas por:

$$W \leftarrow W \otimes \frac{(\frac{V}{WH})H^T}{H^T}, \tag{13}$$

$$H \leftarrow H \otimes \frac{W^T(\frac{V}{WH})}{W^T}, \tag{14}$$

1.3 NMF Supervisionada

Enquanto na versão cega ou não-supervisionada do algoritmo NMF não são consideradas quaisquer informações *a priori*, em sua versão supervisionada são utilizados sinais de cada fonte para treinamento prévio, a fim de se obter conhecimento sobre os padrões espectrais antes da separação propriamente dita. Nesse sentido, a estratégia utilizada consiste em incluir na matriz base W, que contém os padrões espectrais de todas as fontes contidas na mistura, padrões espectrais extraídos dos sinais de treinamento de cada fonte [6].

A separação de fontes por meio da versão supervisionada do método NMF, é realizada em duas etapas [14]:

- Treinamento - Nessa fase assume-se que existem sinais de treinamento cujos espectrogramas, $E_1, E_2, ..., E_n$, possam ser fatorados em duas matrizes A e B que representam, respectivamente, os padrões espectrais e os seus respectivos ganhos, em cada instante de tempo. Após a fase de treinamento, existirá um conjunto matrizes $A_1, A_2, ..., A_N$ que representa os padrões espectrais das N fontes contidas nas misturas, conforme a Equação (15).

$$\begin{aligned} E_1 &\approx A_1 B_1 \\ E_2 &\approx A_2 B_2 \\ &\vdots \\ E_n &\approx A_n B_n \end{aligned} \tag{15}$$

- Separação - Uma vez realizado o treinamento, é possível executar a separação, utilizando-se os padrões espectrais $A_1, A_2, ..., A_n$ e concatenando-os na matriz W. Apenas serão atualizados os índices da matriz W que correspondam às fontes não treinadas e a matriz H, que contém os ganhos de cada fonte presente na matriz W.

  Para os casos nos quais todas as fontes são treinadas, o objetivo da separação passa a ser encontrar a matriz de ganhos H e somente as regras de atualização referentes à matriz H deverão ser executadas, seja pela Distância Euclidiana Quadrática, por meio da Equação (10), ou pela Divergência de Kullback-Leibler, pela Equação (14).

2 VALIDAÇÃO E MÉTRICAS

Nos testes realizados foram utilizados sinais de áudio sintetizados de forma a possibilitar a validação e avaliação da qualidade da separação em relação aos sinais originais de cada uma das fontes separadamente.

Também foram utilizados sinais acústicos captados através de microfones, cada um posicionado bem próximo de cada fonte a fim de minimizar a interferência de outras fontes. Esses sinais captados foram gerados utilizando-se o mesmo ritmo e tempo dos sinais sintetizados. Buscou-se, dessa forma, avaliar o desempenho dos métodos de separação quando aplicados a sinais sintetizados e captados de uma bateria acústica por meio de microfones.

A reconstrução dos sinais no domínio do tempo discreto, seja da mistura ou de cada fonte isolada, foi realizada por meio da Transformada de Fourier de Curta Duração Inversa (ISTFT - *Inverse Short-Time Fourier Transform*). A partir daí, foram realizadas análises sobre os sinais resultantes, envolvendo ambos os processos de separação e da reconstrução. Essa análise pode ser realizada de duas formas, utilizando critérios objetivos ou subjetivamente.

Para a avaliação subjetiva é necessário submeter o áudio original, as fontes separadas e o sinal restaurado à audição de várias pessoas, com questionamentos específicos sobre a qualidade e fidelidade dos sons. Esse formato de validação pode levar a resultados bastante imprecisos, com forte dependência do público alvo e, para que seja confiável, uma grande quantidade de fatores devem ser considerados, aumentando a complexidade da validação. Esse tipo de avaliação é estudada para os casos de áudio em [15].

A avaliação objetiva busca simular a percepção humana e aplicar métricas que possibilitam comparar os resultados de forma mais trivial que a avaliação subjetiva, podendo-se utilizar algoritmos baseados na Razão Sinal-Ruído (SNR - *Signal-to-Noise Ratio*) ou princípios da psicoacústica [8] [16] [17].

2.1 Algoritmos baseados na Razão Sinal-Ruído

As métricas baseadas em SNR consideram que a fonte estimada da mistura (s_j) é igual à fonte original (s_{alvo}) adicionada a interferências (e_{inter}) causadas por outras fontes e erros de quantização ou artefatos (e_{artif}) no processo de separação. Inclui-se, ainda, a parcela e_{noise} referente ao ruído contido na mistura original [16]. Essa notação é apresentada na Equação (16):

$$s_j = s_{alvo} + e_{inter} + e_{artef} + e_{noise} \tag{16}$$

As métricas propostas em [16] são apresentadas a seguir.

- Razão Fonte-Distorção (SDR - *Source to Distortion Ratio*): Visa quantificar a qualidade da separação de uma forma geral,com base na distorção existente entre a fonte estimada e a fonte original. A Equação (17) apresenta a expressão utilizada no cálculo da SDR.

$$SDR_{dB} = 10\log_{10}\frac{||s_{alvo}||^2}{||e_{inter} + e_{artef} + e_{noise}||^2} \tag{17}$$

- Razão Fonte-Interferência (SIR - *Source to Interference Ratio*): Mensura a interferência que a fonte sofreu das demais fontes presentes na mistura, dada pela expressão (18),

$$SIR_{dB} = 10\log_{10}\frac{||s_{alvo}||^2}{||e_{inter}||^2}. \tag{18}$$

- Razão Fontes-Artefatos (SAR - *Sources to Artifacts Ratio*): Essa razão visa medir a quantidade de imperfeições geradas na fase de processamento e pode ser avaliada por meio da Equação (19).

$$SAR_{dB} = 10\log_{10}\frac{||s_{alvo} + e_{inter} + e_{noise}||^2}{||e_{artef}||^2} \tag{19}$$

- Razão Fonte-Ruído (SNR - *Sources-to-Noise Ratio*) - A medida SNR trata da quantidade de ruído presente nas fontes de sinais e é calculada conforme Equação (20).

$$SNR_{dB} = 10\log_{10}\frac{||s_{alvo} + e_{inter}||^2}{||e_{noise}||^2} \tag{20}$$

Para todas essas métricas, valores mais altos são indicativos de melhor qualidade da separação. Como exemplo, um baixo valor SIR significa a presença de forte interferência das outras fontes contidas na mistura.

2.2 Algoritmos baseados em psicoacústica

Além dos algoritmos baseados em SNR, foi também aplicado um formato de avaliação objetiva baseado na psicoacústica, isto é, utilizando-se características do sistema auditivo humano. Esse modelo é denominado Avaliação Perceptiva da Qualidade de Áudio (PEAQ - *Perceptual Evaluation of Audio Quality*)[17].

Esse algoritmo recebe como entradas o arquivo de áudio sob análise e um outro, que será utilizado como referência. Como saída, o método gera uma nota final que representa quão degradado está o áudio em relação ao áudio de referência. Essa nota é chamada de Grau de Diferença Objetiva (ODG - *Objective Difference Grade*) e recebe uma classificação de degradação que vai de 0 (imperceptível) até $-4,0$ (muito incômodo).

Dentre as várias características do sistema auditivo humano, pode-se destacar cinco grandezas que são tratadas pelos modelos de avaliação de qualidade de áudio baseados em psicoacústica [18], sendo:

- *Loudness* - refere-se à intensidade do sinal de áudio, ou seja, amplitude e frequência das amostras e, dessa forma, determina quão audíveis são essas amostras.

- *Sharpness* - determinada pelas amplitudes relativas das componentes de baixa e alta frequência, refere-se à percepção auditiva de sons mais graves ou agudos.

- *Fluctuation Strength* - mede quão perceptíveis, auditivamente, são as modulações lentas, seja em amplitude ou frequência. Usualmente, são consideradas variações acima de 20 Hz.

- *Roughness* - ao contrário da característica *Fluctuation Strength*, essa grandeza mede a percepção humana em relação às variações rápidas do sinal.

- *Masking* - refere-se ao estudo de quão audível o sinal analisado é quando outro sinal sonoro (mascarante), de maior intensidade, está presente. Esse processo pode ocorrer de duas formas:

  Mascaramento simultâneo - quando o sinal analisado e o sinal mascarante ocupam o mesmo intervalo de tempo.

  Mascaramento temporal - quando os dois sinais são apresentados em intervalos diferentes de tempo, pode ser divido em pré-mascaramento, no qual o sinal analisado antecede o sinal mascarante, e em pós-mascaramento, quando o sinal analisado sucede o sinal mascarante.

3 RESULTADOS DE SIMULAÇÃO

Os algoritmos NMF apresentados na Seção 1 foram implementados em ambiente computacional e aplicados a sinais de áudio sintetizados e gravados de uma bateria com três diferentes fontes: bumbo, chimbal e caixa.

As análises comparativas foram realizadas visando à avaliação de impacto das funções de custo Distância Euclidiana Quadrática (DE) ou Divergência de Kullback-Leibler (KL) e considerando sinais sintetizados ou gravados. Em todos os casos foram utilizados os mesmos parâmetros de entrada, ou seja:

- Construção da matriz V ou espectograma do sinal: Foram utilizadas janelas de 512 amostras, com sobreposição de 256 amostras e FFTs de 1024 pontos;

- Número máximo de iterações = 100;

- Quantidade de fontes a serem separadas = 3.

Para os sinais gravados, a captação se deu utilizando um microfone por peça e cada um desses foi posicionado bem próximo de cada fonte, visando minimizar um fenômeno denominado vazamento (*leaking*) que refere-se ao fato de um microfone dedicado a uma peça captar os sinais de outras peças no entorno, por exemplo, o microfone da caixa captar os sinais dos outros instrumentos da bateria, como o bumbo [19].

3.1 Análise sobre a restauração de sinais

A primeira análise destaca o desempenho das duas funções de custo em relação à reconstrução do sinal original com todas as fontes, ou seja, após realizar a separação, a mistura foi refeita e as comparações foram realizadas.

Tabela 1: Método NMF não-supervisionado.

	SDR (dB)		ODG	
Sinal	*DE*	*KL*	*DE*	*KL*
Sinal gravado	6,759	6,759	-3,488	-3,488
Sinal sintetizado	7,657	7,657	-3,103	-3,103

No caso do método NMF não-supervisionado não houve diferenças, isto é, tanto para o sinal gravado quanto para o sinal sintetizado, as duas funções de custo mostraram os mesmos resultados, conforme apresenta a Tabela 1.

Por outro lado, os métodos supervisionados apresentaram diferentes desempenhos, conforme os resultados das métricas mostradas Tabela 2.

Tabela 2: Método NMF supervisionado - Reconstrução de Sinais.

	SDR (dB)		ODG	
Sinal	*DE*	*KL*	*DE*	*KL*
Sinal gravado	17,806	18,084	-1,283	-0,662
Sinal sintetizado	14,657	14,595	-1,302	-0,853

A Figura 2 apresenta o sinal gravado original utilizado nas simulações e na Figura 3, são mostradas as diferenças entre o sinal original e os sinais reconstruídos por meio de cada uma das funções de custo. Visualmente, as reconstruções dos sinais são similares utilizando-se ambas as funções de custo. Entretanto, segundo os dois formatos de avaliação de qualidade, baseado em SNR e por psicoacústica, a reconstrução pela Divergência de Kullback-Leibler apresentou melhor resultado que a Distância Euclidiana Quadrática, conforme destacou a Tabela 2.

Com base na avaliação por psicoacústica, a reconstrução do sinal sintetizado também mostrou resultados favoráveis à Divergência de Kullback-Leibler ao passo que, segundo a avaliação baseada em SNR, o resultado da reconstrução mostrou-se ligeiramente favorável à Distância Euclidiana Quadrática.

3.2 Análise da separação de fontes acústicas

Na segunda avaliação, buscou-se verificar a qualidade da separação dos sinais e nesse sentido, para ambos os sinais, gravado e sintetizado, o método NMF

Figura 2: Sinal original no domínio do tempo discreto.

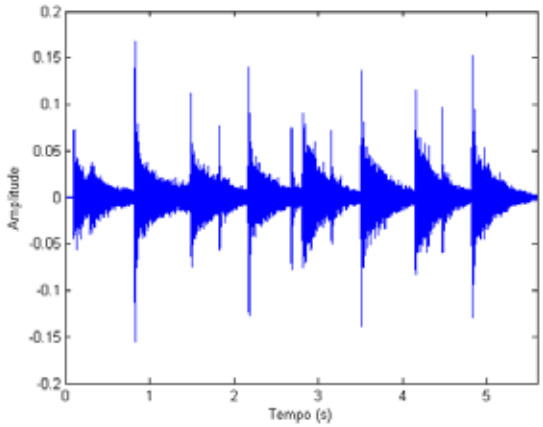
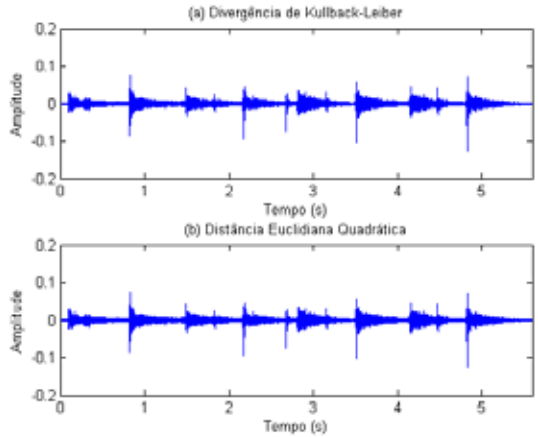


Figura 3: Diferenças entre o sinal original e os sinais reconstruídos.



não-supervisionado não apresentou diferenças para as duas funções de custo.

O método Supervisionado, por sua vez, apresentou diferenças diante da utilização das funções de custo e para ambos os tipos de sinais, sintetizado ou gravado, a Distância Euclidiana Quadrática obteve melhores resultados.

As Tabelas 3 e 4 apresentam os valores obtidos para as métricas baseadas em SNR, utilizando-se sinais gravados e aplicando-se as funções de custo Distância Euclidiana Quadrática e Divergência de Kullback-Leibler, respectivamente.

Tabela 3: Análise de fontes isoladas, utilizando a Distância Euclidiana Quadrática.

Fonte	*SDR (dB)*	*SAR (dB)*	*SIR (dB)*
Chimbal	13,824	14,199	24,809
Caixa	11,140	17,782	12,272
Bumbo	9,532	17,329	10,386

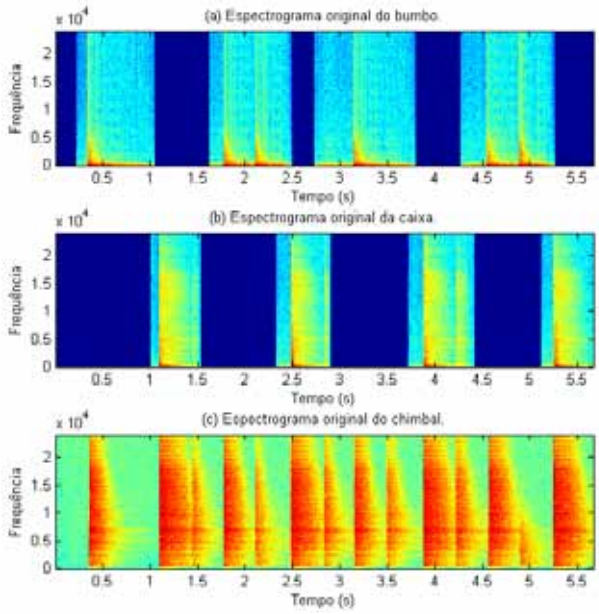
Por meio da análise das Tabelas 3 e 4 é possível verificar que a Distância Euclidiana Quadrática obteve melhores resultados segundo todas as métricas para a separação do bumbo e caixa, enquanto que, para o chimbal, a Divergência de Kullback-Leibler indicou um desempenho mais alto.

Tabela 4: Análise de fontes isoladas, utilizando a Divergência de Kullback-Leibler.

Fonte	*SDR (dB)*	*SAR (dB)*	*SIR (dB)*
Chimbal	13,876	14,203	25,431
Caixa	7,321	16,573	7,965
Bumbo	6,795	15,330	7,567

As faixas de frequências do bumbo e da caixa apresentam componentes de frequência próximas, fato que pode ser verificado na Figura 4. A aplicação da Divergência de Kullback-Leibler implicou resíduos do bumbo na caixa e vice-versa, provocando distorção e interferência em relação à fonte original. É possível verificar essa interferência e distorção no bumbo e na caixa por meio das métricas SDR e SIR, segundo as quais a Divergência de Kullback-Leiber obteve piores resultados. Essa maior interferência pode ser verificada na Figura 5.

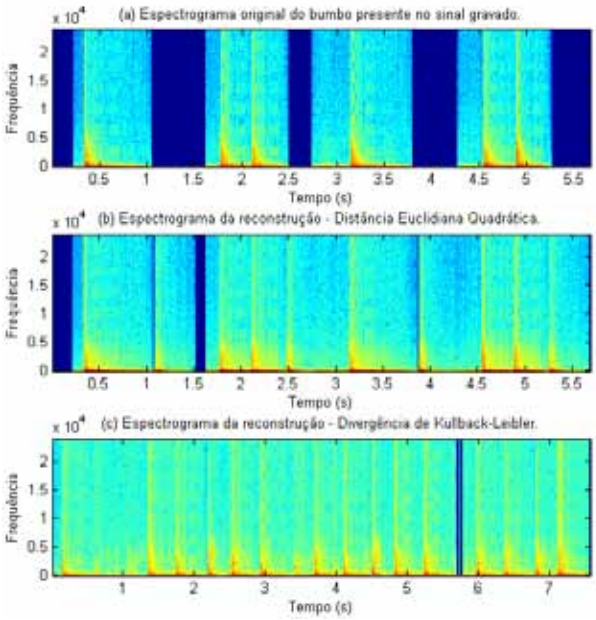
Figura 4: Espectrograma das fontes contidas no sinal gravado.



O som do chimbal foi melhor separado quando utilizado o método NMF Supervisionado com a função de custo Distância Euclidiana Quadrática, com base na Tabela 4, onde esse sinal alcançou valores SAR, SIR e SDR mais elevados.

A avaliação por psicoacústica confirmou o melhor desempenho da Distância Euclidiana Quadrática na separação do bumbo e da caixa, com valores ODG menores comparados à Divergência de Kullback-

Figura 5: Espectrograma do bumbo original contido no sinal gravado e suas reconstruções.



Leibler. O som chimbal, por sua vez, mostrou uma separação melhor segundo a Divergência de Kullback-Leibler, com nota menor do que a nota apresentada pela Distância Euclidiana Quadrática.

Finalmente, a Tabela 5 resume as comparações e enfatiza qual função de custo apresentou melhor separação para cada uma das fonte gravadas, sendo que apenas diferenças maiores que 10% são destacados.

Tabela 5: Melhor separação para cada fonte gravada.

		SDR (dB)		SAR (dB)		SIR (dB)		ODG	
Fonte	*Melhor*	*DE*	*KL*	*DE*	*KL*	*DE*	*KL*	*DE*	*KL*
Chimbal	**KL**	13,82	13,88	14,20	14,20	24,80	25,43	-2,59	-2,55
Caixa	**DE**	**11,14**	7,32	**17,78**	16,57	**12,27**	7,96	-3,91	-3,91
Bumbo	**DE**	**9,53**	6,79	**17,39**	15,33	**10,38**	7,58	-3,90	-3,91

Tabela 6: Melhor separação para cada fonte sintetizada.

		SDR (dB)		SAR (dB)		SIR (dB)		ODG	
Fonte	*Melhor*	*DE*	*KL*	*DE*	*KL*	*DE*	*KL*	*DE*	*KL*
Chimbal	**DE**	**11,18**	10,85	**11,56**	11,23	**22,21**	22,00	-1,87	**-1,65**
Caixa	**DE**	**15,03**	10,64	**16,48**	11,96	**20,62**	16,72	-3,91	-3,91
Bumbo	**DE**	**21,320**	14,24	**22,06**	19,36	**29,59**	15,89	-3,91	-3,91

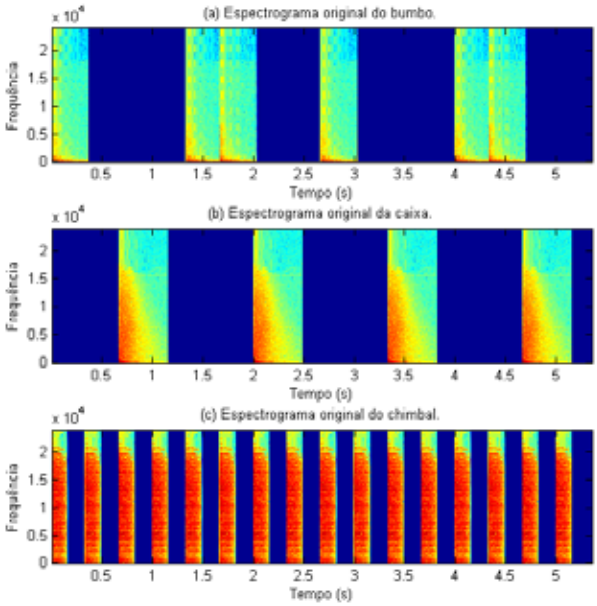
A avaliação da separação, quanto aos sinais sintetizados, apresentou resultados melhores do que no caso da separação de sinais gravados, segundo as métricas baseadas em SNR, isto é, o método NMF separou melhor os sinais sintetizados.

Considerando a avaliação via SNR, o método supervisionado, utilizando a função de custo Distância Euclidiana Quadrática, separou melhor todas as fontes sinte-

Já para a avaliação via psicoacústica, as simulações resultaram em empate para o bumbo e a caixa e apresentaram ODG melhor para a Divergência Kullback-Leibler no caso do chimbal.

Essa diferença acerca das notas ODG entre os sinais gravados e os sintetizados, ocorre pelo motivo de que o espectro da caixa sintetizada se distancia do espectro do bumbo. A caixa sintetizada possui um timbre mais agudo quando comprada com a caixa gravada, isto é, ela possui maior amplitude em componentes de frequência acima de 5000 Hz, conforme Figuras 6 e 4.

Figura 6: Espectrograma das fontes contidas no sinal sintetizado.



Para o método NMF e suas versões, valores de ODG próximos de $-4,0$ (muito incômodo) já foram apresentados na literatura atual, como em [8].

4 CONCLUSÃO

Com base nos resultados obtidos na separação de sinais acústicos, concluiu-se que a Distância Euclidiana Quadrática, como função de custo para o método NMF Supervisionado, apresenta melhor desempenho que a Divergência de Kullback-Leibler para a separação de fontes mais próximas em termos de frequência.

Quanto ao método NMF na versão não-supervisionada, não foram observadas diferenças entre os resultados quanto às funções de custo testadas. Contudo, é possível verificar que a técnica não se comporta de forma equivalente para sinais gravados e sintetizados, com melhor desempenho para esses últimos. Vale ressaltar que a separação de sinais gravados constitui um problema não resolvido, visto que boa parte dos trabalhos são direcionados aos sinais sintetizados.

LISTA DE AUTORES / *AUTHOR INDEX*

Apolinário, I. F.: 62

Biscainho, L. W.: 42, 62
Brandenburg, K.: 31

Cancela, P.: 62
Chiaramonte, M.: 52

Fernandes, B.: 47
Freitas, F. M.: 91

Costa, M.: 82

Fornari, J.: , 75,57

Goecke, D.: 31

Haddad, D. B.: 42

Johann, M. O.: 52

Klein, M. E.: 69

Liebetrau, J.: 31
Lima, M.: 42

Martins, W. A.: 42
Meneses, E. A.: 75

Noceti Filho, S.: 47

Paul, S.: 2,69
Peixoto, Z. M.: 91
Penteado, A. F.: 57
Pirotti, R.: 52

Rocamora, M.: 62

Sanches, I.: 82
Santos, M. B.: 91
Sporer, T.: 31

Velloso, R. C.: 16

Wanderley, M. M.: 75
Yefinczuk, A.: 52

AES BRASIL EXPO 2015

PROMOÇÃO/ORGANIZAÇÃO:

Francal Feiras

AFILIADA:

SINDIPROM

UBRAFE

APOIO

CAPES

CNPq

MONTADORA OFICIAL:

Estrutural Montagens

LOCAL:

EXPO CENTER NORTE

CENTRO DE EXPOSIÇÕES E CONVENÇÕES

VIAGEM:

LAN TAM

SOLIDAVIAGENS

AUDIO ENGINEERING SOCIETY

BRAZIL SECTION

Conceito gráfico e direção de arte: totum agenciatotum.com.br

Imagens e vetores: freepik