Anais do 9º Congresso de Engenharia de Áudio
15ª Convenção Nacional da AES Brasil

Proceedings of the 9th AES Brazil Conference
15th AES Brazil National Convention

AES BRASIL EXPO 2011

CONVENÇÃO E EXPOSIÇÃO DE TECNOLOGIA: ÁUDIO
VÍDEO, ILUMINAÇÃO E INSTALAÇÕES ESPECIAIS

17 A 19 | MAIO

Palácio das Convenções
do Anhembi - São Paulo - SP

Audio Engineering Society - Seção Brasil

Anais do 9º Congresso de Engenharia de Áudio da AES-Brasil

Proceedings of the 9th AES Brazil Conference

17-19 Maio/May, 2011

Palácio das Convenções do Anhembi – São Paulo – SP – Brasil

Copyright ©2011
Audio Engineering Society – Brazil Section

Congresso de Engenharia de Áudio (9.: São Paulo: 2011);
Convenção Nacional AES Brasil (15.: São Paulo: 2011)
Anais 9.Congresso de Engenharia de Audio;
15. Convenção Nacional AES Brasil / ed. R.R.A. Faria. -- São Paulo: AES Brasil, 2011.
Online www.aesbrasil.org/congressos

ISSN 2177-529X

1.Engenharia de áudio (Congressos) 2.Processamento de sinais (Congressos) 3.Computação musical (Congressos) I.Con-venção Nacional AES Brasil (15.: São Paulo, 2011) II.Áudio Engineering Society. Seção Brasil III.Faria, Regis Rossi Alves IV.t.

Os artigos publicados nestes anais foram reproduzidos 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 dos artigos sem autorização expressa da AES Brasil.

Impresso no Brasil.
Printed in Brazil.

Editoração e arte / *Publishing and art:* Leandro Morais Moreira

totum
www.agenciatotum.com.br

Organização

Apoio

Patrocínio

Realização

Audio Engineering Society
BRASIL

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:*	Aldo Ricardo Soares
Vice- Presidente / *Vice- Chairman:*	Armando Baldassarra
Secretário / *Secretary:*	José Anselmo
Tesoureiro / *Treasurer:*	Joel Vieira de Brito
Coordenador Acadêmico / *Academic Coordinator:*	Júlio Lucchi
Conselheiros / *Counselors:*	Carlos Ronconi
	Dagoberto Almeida
	David Fernandes
	Flávia Calabi
	Framklim Garrido
	João Américo
	José Augusto Mannis
	Marcelo Claret
	Regis Rossi A. Faria

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*

Coordenador da Convenção / *Convention Chair*
Aldo R. Soares (AES - Brasil)

Coordenador do Programa Técnico / *Program Committee Chair*
Julio C. Lucchi (USTJ – IMT – FEI - ITA)

Coordenadores de Artigos / *Papers chairs*
Luiz W. P. Biscainho (POLI&COPPE - URFJ)
Francisco J. Fraga da Silva (CECS - UFABC)

Coordenador Editorial / *Editorial Chair*
Regis R. A. Faria (USP)

Coordenador de Infra-estrutura / *Infrastructure Chair*
Jose Augusto Mannis (IA - UNICAMP)

Coordenador de Comunicação / *Communications Chair*
Paulo R. A. Marins (UNB)

Comitê de Programa Técnico / *Technical Program Committee*

Alexandre Leizor Szczupack (COPPE - UFRJ)

Anibal J. de S. Ferreira (Univ. Porto - Portugal)

Carlos Alberto Ynoguti (INATEL - MG)

Celso Kurashima (UFABC - SP)

Christian Herrera (CEFET - MG)

Diego Haddad (CEFET - RJ)

Fabricio de Oliveira Ourique (FENG - PUCRS)

Fernando Santana Pacheco (IFSC & LINCE - UFSC)

Jose Manuel Neto Vieira (Univ. de Aveiro - Portugal)

Leonardo de Oliveira Nunes (COPPE - UFRJ)

Marcelo G. de Queiroz (IME - USP)

Márcio da Costa P. Brandão (CIC-UnB)

Paulo Esquef (INCC)

Phillip M. S. (EP - USP)

Sidnei Noceti Filho (LINSE/CTC/UFSC)

Walter Gontijo (UNIVALI & LINCE - UFSC)

Nesta edição tivemos 25 artigos submetidos em 2 categorias: artigos completos (14 submissões) e artigos estudantis (11), demonstrando o interesse crescente da comunidade estudantil, trazendo a público os resultados de seus trabalhos de pesquisa e desenvolvimentos técnico-científicos.

Foram selecionados 16 artigos para apresentação nesta 9a. edição do congresso, nas 2 categorias. Os trabalhos abordam tópicos como computação musical, psicoacústica, áudio multicanal, acústica ambiental, análise/síntese sonora, processamento de áudio e eletrônica para áudio, contribuindo para uma real construção do conhecimento e o avanço da grande área de engenharia de áudio no Brasil.

Além disso, apresentamos nesta edição um artigo convidado sobre filtros e efeitos de áudio analógicos e digitais, de autoria dos profs. Sidnei Noceti Filho e Walter Gontijo, da Universidade Federal de Santa Catarina, e do prof. André Dalcastagnê, do Instituto Federal de Santa Catarina, aos quais agradecemos pela valiosa contribuição. Temos certeza que o tema abordado é de interesse de tantos que desejam conhecer mais e se aventurar pela mágica de fazer e processar áudio.

Esperamos que todos aproveitem ao máximo a convenção e o congresso deste ano!

Regis Rossi Faria
Coord. Editorial

Filtros e efeitos para áudio - analógicos e digitais
Terça-feira/*Tuesday*, 17 Maio/*May* 2011 13:30 - 15:00
Sala 2 / *Room 2*
Sidnei Noceti Filho (UFSC), Walter Antônio Gontijo (UFSC) e André Luís Dalcastagnê (IFSC)

Arranjos de microfones em sistemas de comunicações por voz
Terça-feira/*Tuesday*, 17 Maio/*May* 2011 15:30 - 17:00
Sala 2 / *Room 2*
Luiz Wagner Biscainho (UFRJ) e Leonardo de Oliveira Nunes (UFRJ)

Audio para video games
Terça-feira/*Tuesday*, 17 Maio/*May* 2011 16:30 - 18:30
Auditorio 8 / *Auditorium 8*
Leonard Paul (Universidade de Arte e Design Emily Carr, Vancouver)

Distorções audíveis em sistemas processados por DSPs
Terça-feira/*Tuesday*, 17 Maio/*May* 2011 17:00 - 18:30
Sala 2 / *Room 2*
Francisco Monteiro, Ruy Monteiro e Homero Sette (Studio R / Nashville)

Representações tempo-frequência para a análise de música (Idioma: Espanhol)
Quarta-feira/*Wednesday*, 18 Maio/*May* 2011 15:30 - 17:00
Sala 2 / *Room 2*
Eng. Ernesto López e Eng. Martín Rocamora (Instituto de Ingeniería Eléctrica, Facultad de Ingeniería y Escuela Universitaria de Música, Universidad de la República, Uruguay)

Painel: Tecnologias de rede na transferência de áudio
Quarta-feira/*Wednesday*, 18 Maio/*May* 2011 15:30 - 17:00
Auditorio 8 / *Auditorium 8*
Moderador: Miguel Ratton (Informus Music Center), Rafael Susin (DSPro), Sotex - Aviom – A-Net , Alex Lameira (Roland Brasil), Aldo Linares (Yamaha do Brasil)

Mixando monitores
Quarta-feira/*Wednesday*, 18 Maio/*May* 2011 17:30 - 19:00
Auditorio 8 / *Auditorium 8*
Greg Looper (Monitor Engineer e Pro-Tools Recording do Tom Petty & The Heartbreakers)

O primeiro som da minha igreja
Quinta-feira/*Thursday*, 19 Maio/*May* 2011 13:00 - 14:30
Sala 1 / *Room 1*
David Fernandes (AudioCon)

*Palestras confirmadas até a dia 16/05/2011

Sessão 1 / *Session 1 - Full Papers 1*

Terça-feira/*Tuesday*, 17 Maio/*May* 2011 10:00AM - 12:00AM
Sala 1 / *Room 1*
Coordenador de sessão / *Chair:* Julio Lucchi

Um tutorial sobre a medição in situ do coeficiente de absorção com as técnicas PP e PU

Eric Brandão e Arcanjo Lenzi (Universidade Federal de Santa Catarina)

Avaliação de medições de coeficiente de espalhamento de amostras de difusores sonoros com elementos semicilíndricos e cilíndricos concebidos a partir de design serial: influência da natureza dos elementos e da qualidade de ordenação

José A. Mannis, Christian dos Santos, Stellamaris Rolla Bertoli, Regiane Pupo (Universidade Estadual de Campinas - UNICAMP), Fernandes Oliveira e Jorge Vicente Lopes da Silva (Centro de Tecnologia da Informação Renato Archer - CTI)

Um Estudo Comparativo de Equalizadores de Áudio Gráficos Analógicos

André L. Dalcastagnê (Instituto Federal de Santa Catarina) e Sidnei Noceti Filho (Universidade Federal de Santa Catarina)

Aplicação de Controle Baseado em Passividade a Fontes de Alimentação para Amplificadores de Potência

Thiago R. de Oliveira (Instituto Federal de Minas Gerais), Pedro F. Donoso-Garcia (Universidade Federal de Minas Gerais)

Hardware Básico para Controle via MIDI

Miguel B. Ratton (Informus Music Center)

Sessão 2 / *Session 2 - Student Papers*

Quarta-feira/*Wednesday*, 18 Maio/*May,* 2011 10:00AM - 12:00AM
Sala 1 / *Room 1*
Coordenador de sessão / *Chair:* Leonardo Nunes

Proposta de Estudo do Potencial Evocado Auditivo P300 para Averiguação dos Processos Cognitivos em Indivíduos com Ouvido Absoluto

Rodrigo Munhoz, Cesar A. Mateus, Francisco J. Fraga (Universidade Federal do ABC - UFABC) e Patricia Vanzella (Universidade de Brasília - UnB)

Sistemas de Composição Audiovisuais Duais com Síntese Granular

Fernando F. de Souza, Claudiney Carrasco, Adolfo Maia Jr. (Universidade Estadual de Campinas - UNICAMP)

Dynamic Linear Downmix from Surround to Stereo Sytems

Guilherme Lunhani, José A. Mannis, Jonatas Manzolli e Adolfo Maia Jr. (Universidade Estadual de Campinas - UNICAMP)

Desenvolvimento de Um Módulo de Resposta Vocal para a Plataforma Embarcada Arduino

Julia Cristina Vieira e Fernando S. Pacheco (Instituto Federal de Santa Catarina)

Experimentos de Teleconferência Usando o Codec de Voz Speex

Douglas G. Ribeiro e Celso S. Kurashima (Universidade Federal do ABC)

Sistema de Ocultação de Dados em Áudio através de Técnicas de Compactação e Espalhamento Espectral

Luiz D. Aquino, Aurélio F. Hoppe e Paulo R. Brandt (Fundação Universidade Regional de Blumenau - FURB)

Sessão 3 / *Session 3 - Full Papers 2*

Quinta-feira/*Thursday*, 19 Maio/*May,* 2011 10:00AM - 12:00AM
Sala 1 / *Room 1*
Coordenador de sessão / *Chair:* Luiz Biscainho

Projeto de banco de filtros digitais por janelas Kaiser para identificação de notas e acordes de violão usando imagens binarizadas por limiar de Otsu

Mauricio M. Cordeiro Jr e Helio M. de Oliveira (Universidade Federal de Pernambuco)

Pitch tracking in polyphonic audio by clustering local fundamental frequency estimates

Martín Rocamora e Pablo Cancela (Universidad de la República, Uruguay)

DUETO: System for Synthesis Singing Duets in Real Time

Adriano Monteiro e Jônatas Manzolli (Universidade Estadual de Campinas - UNICAMP)

Redução de Ruído em Sinais de Voz Usando Função de Limiar SoftSoft para Minimizar a Distorção Log-Espectral

Irineu Antunes Júnior (Universidade Federal do ABC)

Implementações em downmix passivo de 5.1 para 2.0

José A. Mannis (Universidade Estadual de Campinas - UNICAMP)

Sumário / *Contents*

Editorial 5
Programação de Palestras / 15º Convenção Nacional da AES Brasil 6
Sessões de Artigos / Papers Sessions 7

Artigo Convidado / *Invited Paper*

Filtros e Efeitos de Áudio Analógicos e Digitais
Sidnei Noceti Filho, André Luís Dalcastagnê e Walter Antônio Gontijo 12

Artigos Completos / *Full Papers*

Um tutorial sobre a medição in situ do coeficiente de absorção com as técnicas PP e PU
Eric Brandão e Arcanjo Lenzi (Universidade Federal de Santa Catarina) 39

Avaliação de medições de coeficiente de espalhamento de amostras de difusores sonoros com elementos semicilíndricos e cilíndricos concebidos a partir de design serial: influência da natureza dos elementos e da qualidade de ordenação
José A. Mannis, Christian dos Santos, Stellamaris Rolla Bertoli, Regiane Pupo , Fernandes Oliveira e Jorge Vicente Lopes da Silva 45

Um Estudo Comparativo de Equalizadores de Áudio Gráficos Analógicos
André L. Dalcastagnê e Sidnei Noceti Filho 53

Aplicação de Controle Baseado em Passividade a Fontes de Alimentação para Amplificadores de Potência
Thiago R. de Oliveira, Pedro F. Donoso-Garcia 61

Hardware Básico para Controle via MIDI
Miguel B. Ratton (Informus Music Center) 69

Projeto de banco de filtros digitais por janelas Kaiser para identificação de notas e acordes de violão usando imagens binarizadas por limiar de Otsu
Mauricio M. Cordeiro Jr e Helio M. de Oliveira 74

Pitch tracking in polyphonic audio by clustering local fundamental frequency estimates
Martín Rocamora e Pablo Cancela 80

DUETO: System for Synthesis Singing Duets in Real Time
Adriano Monteiro e Jônatas Manzolli 88

Redução de Ruído em Sinais de Voz Usando Função de Limiar SoftSoft para Minimizar a Distorção Log-Espectral
Irineu Antunes Júnior 95

Implementações em downmix passivo de 5.1 para 2.0
José A. Mannis 102

Artigos Estudantis / *Student Papers*

Proposta de Estudo do Potencial Evocado Auditivo P300 para Averiguação dos Processos Cognitivos em Indivíduos com Ouvido Absoluto
Rodrigo Munhoz, Cesar A. Mateus, Francisco J. Fraga e Patricia Vanzella 108

Sistemas de Composição Audiovisuais Duais com Síntese Granular
Fernando F. de Souza, Claudiney Carrasco, Adolfo Maia Jr. 112

Dynamic Linear Downmix from Surround to Stereo Sytems
Guilherme Lunhani, José A. Mannis, Jonatas Manzolli e Adolfo Maia Jr. 116

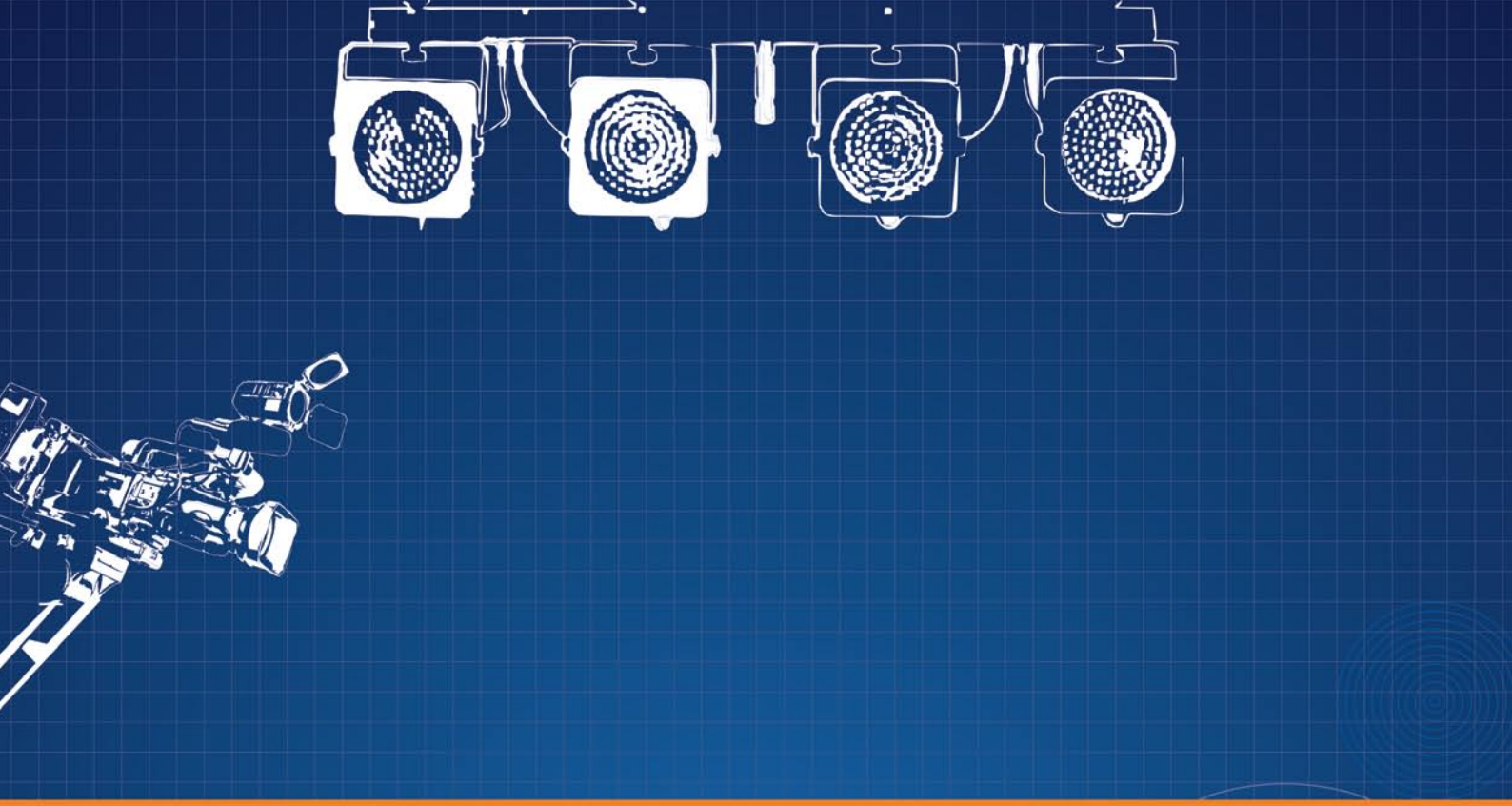
Desenvolvimento de Um Módulo de Resposta Vocal para a Plataforma Embarcada Arduino
Julia Cristina Vieira e Fernando S. Pacheco 120

Experimentos de Teleconferência Usando o Codec de Voz Speex
Douglas G. Ribeiro e Celso S. Kurashima 124

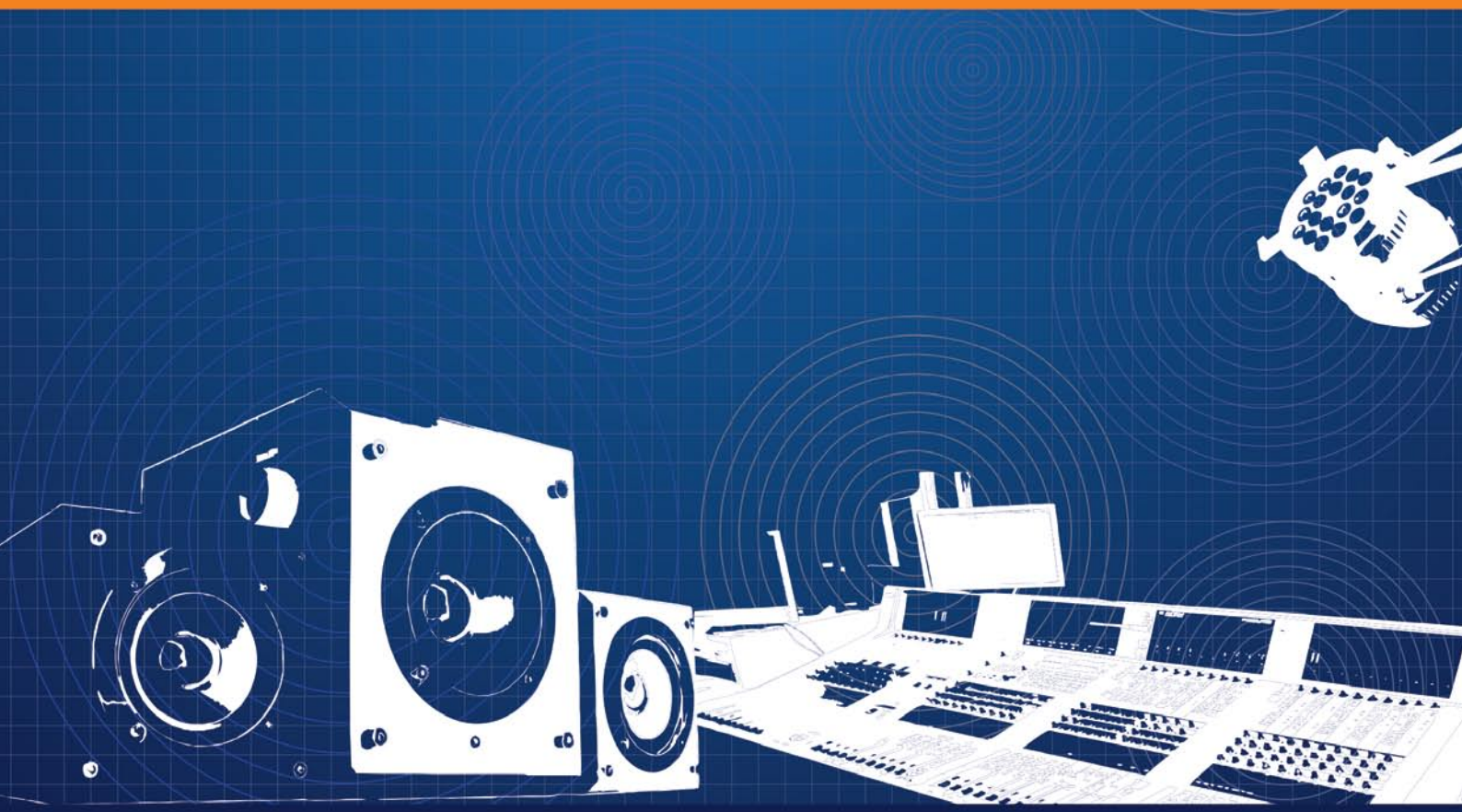
Sistema de Ocultação de Dados em Áudio através de Técnicas de Compactação e Espalhamento Espectral
Luiz D. Aquino, Aurélio F. Hoppe e Paulo R. Brandt 128

Índice de Autores / Author Index 132

Lista de Revisores / Reviewers 133



ARTIGO CONVIDADO



Sociedade de Engenharia de Áudio
Artigo de Congresso

Apresentado no 9º Congresso de Engenharia de Áudio
15ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2011, 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.

Filtros e Efeitos de Áudio Analógicos e Digitais

Sidnei Noceti Filho[1], André Luís Dalcastagnê[2] e Walter Antônio Gontijo[1]
[1] Universidade Federal de Santa Catarina
LINSE – Laboratório de Circuitos e Processamento de Sinais
Florianópolis, SC, 88040-900, Brasil
[2] Instituto Federal de Santa Catarina
Florianópolis, SC, 88040-900, Brasil
sidnei@linse.ufsc.br, andreld@ifsc.edu.br, walter@linse.ufsc.br

RESUMO
Este tutorial apresenta os principais efeitos de áudio e seus diagramas de blocos para implementações nas formas analógicas e digitais. Relativamente à forma analógica, são apresentados alguns circuitos básicos, além de considerações práticas. Não são apresentados circuitos prontos de efeitos. Entretanto, o material apresentado dá condições aos projetistas de criarem seus próprios circuitos. Relativamente à forma digital, são apresentados o processador digital de sinais, os diagramas de blocos dos principais efeitos e os algoritmos correspondentes. Os arquivos de testes e os algoritmos implementados estão disponíveis para download.

ABSTRACT
This tutorial presents the main audio effects and their block diagrams for implementation in analog and digital forms. In the analog form, some basic circuits and practical considerations are presented. Effects circuits are not presented. However, the material presented gives conditions for designers to create their own circuits. With regard to the digital form, the DSP processor, block diagrams of the main effects, and the corresponding algorithms are presented. Test files and the implemented algorithms are available for download.

0 INTRODUÇÃO

Os efeitos de áudio podem ser classificados em dois grandes grupos. No primeiro, sons de voz ou de instrumentos musicais são processados por um circuito e têm os seus parâmetros, tais como amplitude, fase e conteúdo espectral, variados. No segundo, sons dos mais diversos são produzidos por sintetizadores e, neste caso, não há necessidade de um instrumento musical. Neste tutorial, são discutidos os efeitos do primeiro grupo [1]-[3].

Os circuitos modificam o som original e, obviamente, têm a função de tornar os sons mais interessantes. Dois dos efeitos mais usados em sinais de voz e em sinais produzidos por instrumentos musicais (guitarra, por exemplo) são o eco e a reverberação. O eco se constitui em um conjunto de repetições do som original, sendo que o ouvido percebe a existência de um tempo entre o som original e as repetições. Esse fenômeno acontece na natureza quando um anteparo refletor de som está distante a pelo menos 8,5 m da fonte sonora (correspondente a um atraso de 50 ms) ou mais. Já a reverberação consiste em uma grande quantidade de reflexões, sendo que o ouvinte não percebe um silêncio entre a emissão e os sons refletidos. Isso acontece, por exemplo, quando se fala dentro de uma igreja. Outros efeitos, tais como *flanger*, *chorus* e *phaser*, são usados principalmente por guitarristas, o que não impede que outros instrumentistas os utilizem.

Diversas são as formas de se produzir um efeito. São vários os diagramas de blocos e a forma de implementá-los. Mesmo definido um diagrama de blocos específico e um circuito particular, diversos parâmetros são disponibilizados para que o músico obtenha o efeito que mais lhe agrade. Assim, não é possível expor em um texto limitado a infinidade de soluções apresentadas na literatura. Exemplificando, considere um resistor

controlado por um oscilador como um destes parâmetros variáveis. Encontram-se na literatura os resistores variáveis sendo implementados por foto-acopladores (par LDR e LED), por transistores de efeito de campo de junção (JFETs) e por amplificadores operacionais de transcondutância (OTAs). Além disso, diversas implementações de filtros e de osciladores com diferentes formas de onda são usadas para controlar os resistores variáveis.

Este trabalho tem como objetivo auxiliar projetistas e/ou *hobistas* a desenvolverem projetos a partir de diagramas de blocos e alguns circuitos básicos. Ele está organizado da forma descrita a seguir. Na Seção 1, são apresentados os diagramas de blocos de alguns dos principais efeitos analógicos. Na Seção 2, são mostrados alguns dos circuitos auxiliares usados na implementação dos blocos. A Seção 3 apresenta o controle de tonalidade (função de transferência do tipo *shelving)* e equalizadores analógicos gráficos, paramétricos e paragráficos (função de transferência do tipo *bump*). A Seção 4 aborda os processadores de sinais digitais, célula básica em circuitos de efeitos digitais. Na Seção 5, são apresentados os diagramas de blocos de alguns dos principais efeitos digitais. A Seção 6 apresenta os equalizadores digitais. A Seção 7 comenta os resultados experimentais dos efeitos implementados na forma digital. Por fim, na Seção 8, são apresentadas as considerações finais.

1 DIAGRAMAS DE BLOCOS DOS EFEITOS ANALÓGICOS

1.1 Vibrato

O vibrato é um efeito que soa ao ouvido como uma modulação nas freqüências do sinal. A variação das freqüências é naturalmente obtida quando um guitarrista, acionando uma alavanca, altera a tensão nas cordas da guitarra. Em um teclado, tal efeito é obtido com a variação da freqüência de oscilação dos geradores de sinais responsáveis pela geração das formas de onda. Para sinais de freqüência fixa, como ocorre de modo geral nos pedais, um efeito semelhante à modulação em freqüência é obtido com o artifício da modulação da fase de um filtro passa-tudo (*all-pass*) de primeira ordem ou com a cascata de dois deles. A modulação é obtida com um oscilador de baixa freqüência (LFO). O sinal do LFO altera o valor da resistência responsável pela característica de fase do circuito passa-tudo. A Figura 1(a) mostra o diagrama de blocos. A Figura 1(b) mostra a variação da fase obtida com a cascata de dois filtros passa-tudo de primeira ordem.

Um dos controles principais do vibrato é o *rate*, que altera a freqüência de oscilação do LFO e, conseqüentemente, a velocidade com que as semitonações ocorrem. O outro é o controle *depth*, que altera a amplitude na saída do LFO e, conseqüentemente, o quanto varia o atraso proporcionado pelo filtro passa-tudo, ocasionando variações na aparente semitonação. Efeitos interessantes de vibrato são obtidos com freqüências do LFO de 4 a 8 Hz.

1.2 Tremolo

O efeito de tremolo corresponde a uma modulação da amplitude de um sinal. As variações são obtidas com um LFO. As Figuras 2(a), 2(b) e 2(c) mostram três possíveis diagramas de blocos para a obtenção desse efeito. Nos primeiro e segundo diagramas, podem ser usados circuitos com OTAs ou amplificador (amp. op.) mais resistor controlado implementado com foto-acoplador ou ainda JFETs. No terceiro, usa-se um VCA (amplificador controlado por tensão) disponível na forma integrada. A Figura 2(d) mostra a variação do ganho que ocorre em quaisquer dos circuitos.

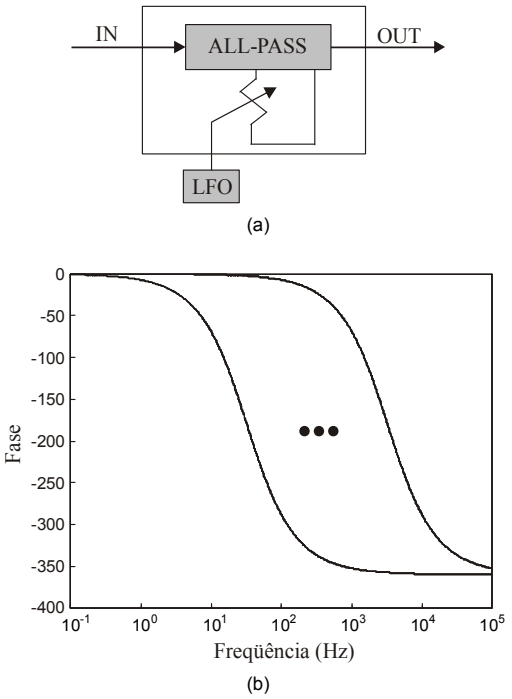


Figura 1 Efeito vibrato. (a) Diagrama de blocos da implementação do efeito. (b) Fase obtida ao se utilizar dois filtros passa-tudo.

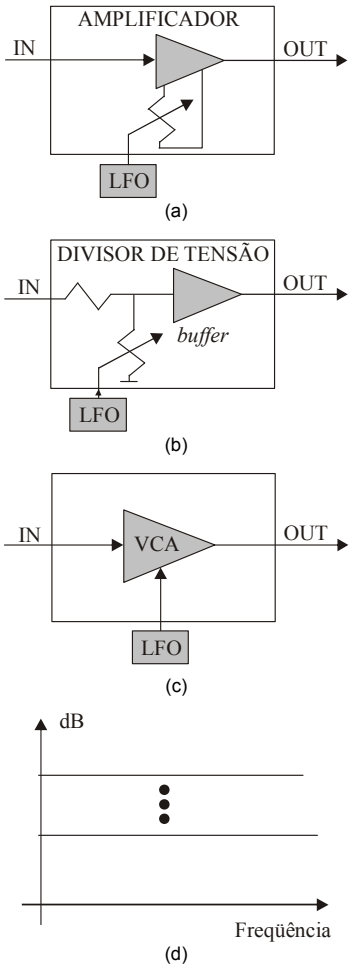


Figura 2 Obtenção do efeito tremolo. (a) Primeira possibilidade de diagrama de blocos. (b) Segunda possibilidade de diagrama de blocos. (c) Terceira possibilidade de diagrama de blocos. (d) Ganho das etapas.

Um dos controles principais é o *rate*, que comanda a freqüência de oscilação do LFO e, conseqüentemente, a velocidade com que o volume varia. O outro é o *depth*, que comanda o quanto varia o ganho de qualquer um dos tipos de estágios mostrados. Efeitos interessantes de tremolo são obtidos com freqüências do LFO na faixa de 1 a 10 Hz.

1.3 Phaser

Os diagramas de blocos do *phaser* são mostrados na Figura 3. O diagrama da Figura 3(a) é semelhante ao diagrama de blocos do vibrato, pois também usa uma cascata de filtros passa-tudo controlados por um LFO. A realimentação mostrada na Figura 3(b) pode ser feita entre estágios passa-tudo e não necessariamente no primeiro estágio.

Como já dito, no efeito vibrato, um ou dois filtros passa-tudo são suficientes para a obtenção do efeito. No caso do *phaser*, efeitos significativos são obtidos com quatro a doze estágios. Outra diferença em relação ao vibrato é que aqui são obtidos efeitos mais significativos quando se soma o sinal atrasado com o sinal puro.

Na saída do circuito do *phaser*, ocorrem interferências construtivas e destrutivas. Tais interferências, ao contrário do que acontece no efeito *flanger* (discutido mais adiante), ocorrem em freqüências que não apresentam uma relação de números inteiros umas com as outras.

Um dos controles principais do *phaser* é o *rate*, que altera a freqüência do LFO e, conseqüentemente, a velocidade com que ocorrem as variações de fase. O outro controle é o *depth*, que altera a amplitude na saída do LFO e, conseqüentemente, o quanto varia o atraso proporcionado pelos filtros passa-tudo. Efeitos interessantes de *phaser* são obtidos com freqüências do LFO na faixa de 3 a 8 Hz.

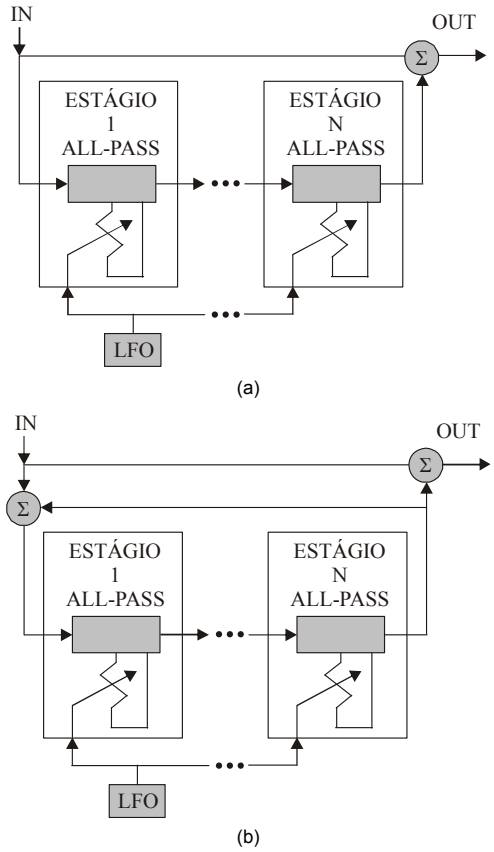


Figura 3 Diagramas de blocos do efeito *phaser*. (a) Sem realimentação. (b) Com realimentação.

1.4 Chorus

O efeito *chorus* é obtido com a soma de um sinal que foi semitonado com o sinal puro, conforme mostra a Figura 4. O sinal semitonado pode ser obtido com um circuito analógico amostrado (um BBD – *bucket-brigade device*, apresentado na Seção 2.3, por exemplo). Um sinal de *clock* com freqüência variável controla os tempos de atraso entre a entrada e a saída que variam de 10 a 30 ms. Ela é alterada com o auxílio de um LFO com freqüência da ordem de 1 Hz ou menor.

O diagrama do *chorus* da Fig. 4 é semelhante ao diagrama do *phaser* da Fig. 3 (a). No diagrama da Figura 4, os blocos "COMP." e "EXP." são opcionais. Eles realizam uma compressão e uma expansão do sinal, respectivamente, para diminuir a faixa dinâmica do sinal a ser processado pelo BBD.

Um dos controles principais do *chorus* é o *rate*, que comanda a freqüência de oscilação do LFO. Essa freqüência controla a velocidade das variações. O controle *depth* é o comando de quanto varia o atraso do registrador de deslocamento, BBD, proporcionando diferentes efeitos. Pode ser colocado também um controle no somador de saída para "dosar" o quanto de sinal puro é misturado com o sinal atrasado.

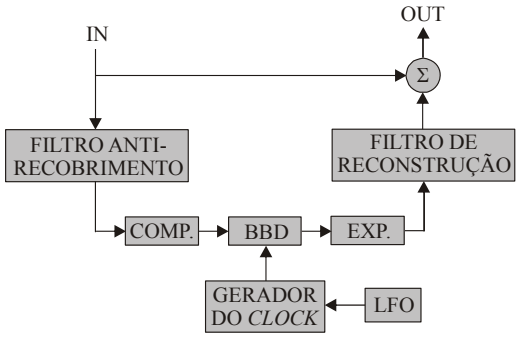


Figura 4 Diagrama de blocos do efeito *chorus*.

1.5 Flanger

O efeito *flanger* também é obtido com a soma de um sinal que foi semitonado por um BBD com o sinal puro, conforme mostra a Figura 5. As diferenças em relação ao efeito *chorus* são:

i. O sinal semitonado é realimentado ao registrador de deslocamento analógico.
ii. Os tempos de atraso entre a entrada e a saída do registrador de deslocamento são menores do que aqueles usados no efeito *chorus* (de 0,1 a 10 ms).

Um dos controles principais do *flanger* é o *rate*, que é o controle da freqüência do LFO, sendo da ordem de 2 Hz ou menos. Assim, como nos casos anteriores, outro controle importante é o *depth*, que controla o quanto varia o atraso do registrador de deslocamento.

O diagrama de blocos do *flanger* (ver Figura 5) é semelhante ao diagrama do *phaser* da Figura 3(b). Como no *phaser*, na saída do circuito do *flanger*, ocorrem interferências construtivas e destrutivas. Mas aqui ocorrem em freqüências que apresentam uma relação de números inteiros umas com as outras. Sendo assim, em uma escala logarítmica, as freqüências ficam igualmente espaçadas.

A diferença fundamental que faz com que o efeito *phaser* soe diferente dos efeitos *chorus* e *flanger* é que o filtro passa-tudo é um circuito analógico contínuo que apresenta uma fase não-linear (a fase é uma função arco-tangente). Conseqüentemente, os atrasos variam com a freqüência do sinal. Nos efeitos *chorus* e *flanger*, o circuito atrasador é um circuito analógico amostrado que

proporciona atrasos independentes da freqüência (fase linear) para uma freqüência de *clock* constante. No entanto, nos circuitos *chorus* e *flanger*, a freqüência do *clock* é variável. Mas devido a essas diferentes características de fase, os efeitos obtidos soam diferentes. Por operar com sinais analógicos contínuos, os circuitos do *phaser* dispensam filtros anti-recobrimento e de reconstrução.

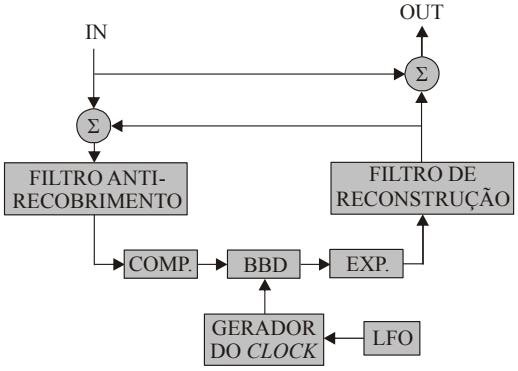


Figura 5 Diagrama de blocos de um *flanger*.

1.6 Eco

O efeito eco é usado tanto para sinais de guitarra como para sinais de voz. Ele consiste em simular as repetições do som, como mencionado, que ocorrem quando existe um refletor a uma distância de pelo menos 8,5 m, o que corresponde a um atraso de cerca de 50 ms entre o som original e o refletido. Mas esse é aproximadamente o tempo mínimo para que o som original e o refletido soem distintamente. Efeitos mais significativos e interessantes são obtidos para tempos de atraso maiores.

O diagrama de blocos de um efeito de eco (ver Figura 6) é semelhante ao do *flanger*. Mas, além da diferença significativa entre os tempos de atraso envolvidos, aqui a freqüência do *clock* é constante.

Os principais controles do ECO são: o *delay time*, que é o tempo entre uma amostra e outra; o *feedback*, que controla a quantidade de sinal que é realimentada (e conseqüentemente o número de repetições); e o *mixer*, que dosa a quantidade do sinal atrasado com o sinal original

Outro efeito interessante semelhante ao eco é o *reverber*. Nesse caso, não apenas uma, mas inúmeras reflexões ocorrem em diferentes instantes de tempo de tal forma que não é percebido um tempo de atraso entre o som original e os vários sons refletidos. O som persiste, diminuindo de amplitude até se extinguir. Mas devido ao grande número de reflexões e inúmeros tempos de atraso que precisam ser simulados, não é prático obter o efeito *reverber* com circuitos eletrônicos analógicos. Efeitos interessantes são obtidos com processadores digitais ou com transdutores eletromecânicos e molas, por exemplo.

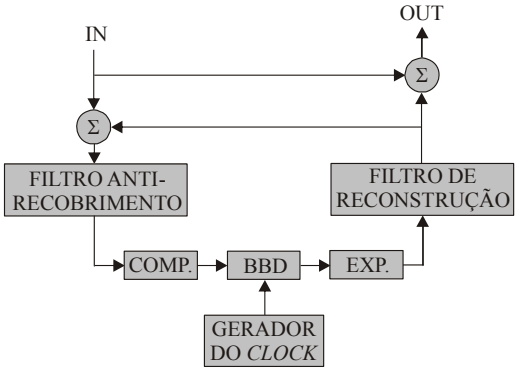


Figura 6 Diagrama de blocos para obtenção do efeito de eco.

1.7 Wah-Wah

O efeito wah-wah tenta imitar o som feito pela boca quando seu nome é pronunciado. Ele é obtido quando o conteúdo espectral é variado, contínua ou discretamente (ver Figura 7), com o uso de um dos seguintes tipos de filtros:

i. Filtro seletor passa-faixa (com um zero na origem e um zero no infinito) com freqüência central variando para cima e para baixo com o uso de um pedal, ou automaticamente, com o uso de um detector de envelope.

ii. Filtro passa-baixa com fator de qualidade $Q >> 0{,}707$, para apresentar sobrepassamento em freqüência e, conseqüentemente, um efeito tipo passa-faixa.

iii. Equalizador *bump* (ou "*peaking filter*").

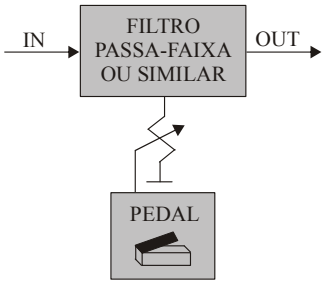


Figura 7 Diagrama de blocos do efeito wah-wah.

Os três tipos de curvas são mostrados na Figura 8. Naturalmente, os efeitos soam diferentes quando os diferentes tipos de filtros são usados. Efeitos mais significativos são obtidos com um filtro passa-faixa. Variações típicas da freqüência de pico variam de 300 Hz a 3 kHz.

É possível obter um efeito wah-wah, sem o uso de um pedal, através de um detector de envelope excitado pelo sinal da guitarra (ver Figura 9). O sinal resultante na saída do detector é aplicado em um resistor dependente da tensão, o qual é responsável pela variação da freqüência do filtro utilizado. Naturalmente, a quantidade de efeito vai depender do desempenho (dinâmica) do músico. Esse tipo de efeito é comumente chamado de auto wah-wah. Essa idéia, a princípio, poderia ser tentada em outros tipos de efeitos.

No wah-wah pode-se controlar o fator de qualidade, as freqüências mínimas e máximas onde ocorrem os picos de amplitude e quanto o fator de qualidade varia quando as freqüências de pico são variadas (controle da largura de banda). Uma variação do efeito descrito consiste em usar filtros em paralelo que operam com freqüências centrais diferentes. Também é possível criar um controle de quanto de sinal puro é misturado com o sinal que passa pelo filtro.

1.8 Distorção

Os efeitos de distorção são obtidos quando se obtém uma saturação forçada de um ou mais estágios do circuito que processa o sinal, objetivando a criação de harmônicos. Enquanto em todos os outros tipos de efeitos as saturações não são desejadas, aqui elas são propositadamente produzidas.

Na literatura, esse é provavelmente o efeito que apresenta o maior número de variações em termos de circuitos. A quantidade e a proporção de harmônicos pares e ímpares dependem da topologia e, principalmente, se as saturações são obtidas com amp. op., transistores bipolares, JFET, MOSFET e válvulas. Diodos também são usados em conjunto com todos esses dispositivos.

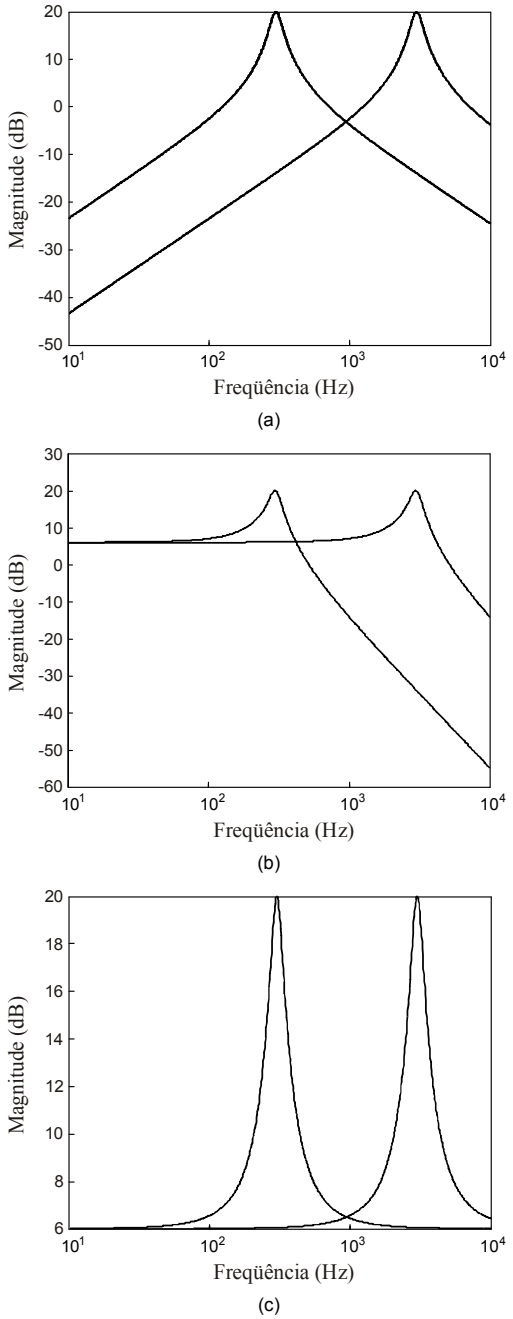


Figura 8 Três tipos de filtros usados no efeito wah-wah. (a) Filtro passa-faixa. (b) Filtro passa-baixa com alto fator de qualidade. (c) Filtro com função do tipo *bump*.

Três dos mais conhecidos tipos de pedais que proporcionam esse efeito são o *overdrive*, o *distortion* e o *fuzz*. No efeito *overdrive,* tenta-se simular as distorções que ocorrem em amplificadores valvulados saturados. Um forte componente desses sinais são harmônicos de ordem par, que soam mais suaves ao ouvido humano. No efeito *distortion*, harmônicos pares e ímpares são obtidos. Logo, em relação ao *overdrive*, tal efeito soa menos suave ao ouvido humano. No efeito *fuzz*, procura-se realçar os harmônicos ímpares, que aparecem naturalmente quando os sinais são aproximados a ondas quadradas. Por isso, dentre os três efeitos de distorção, este soa o menos suave.

Nos três tipos de pedais, pode-se controlar a quantidade de distorção (nível de saturação) e o nível do sinal na saída. Alguns pedais possibilitam alguma forma de equalização, para ressaltar e/ou atenuar certas faixas de freqüência.

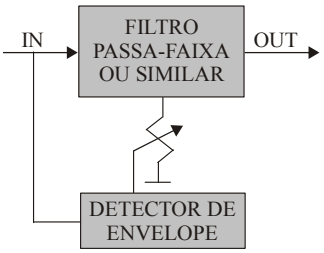


Figura 9 Diagrama de blocos do efeito auto wah-wah.

2 CIRCUITOS AUXILIARES USADOS EM PEDAIS DE EFEITOS ANALÓGICOS

2.1 Resistores Variáveis – Aspectos Práticos

Os resistores variáveis têm aplicação em diversos tipos de pedais de efeitos. Alguns exemplos são descritos a seguir.

a) LDR

O LDR (*ligth dependent resistor*) é um resistor construído com sulfeto de cádmio (CdS) ou seleneto de cádmio (CdSe) que, como o próprio nome indica, apresenta uma resistência que varia com a intensidade de luz. A Figura 10 apresenta o aspecto geral de um LDR. A Figura 11 apresenta diferentes símbolos encontrados na literatura.

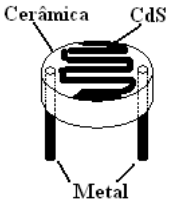


Figura 10 Aspecto geral de um LDR.

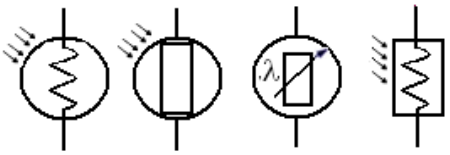


Figura 11 Possíveis símbolos do LDR.

Encontram-se no comércio LDRs com as mais variadas características. As resistências na ausência de luz variam aproximadamente de 100 kΩ a 20 MΩ. As resistências no claro variam aproximadamente de 100 Ω a 100 kΩ. As potências máximas variam de 50 a 500 mW e as tensões máximas variam de 100 a 400 V. Essas características dependem do tipo de material fotossensível usado, do diâmetro do LDR, do tipo de geometria (os três tipos mostrados na Figura 12) e da espessura e do comprimento das trilhas.

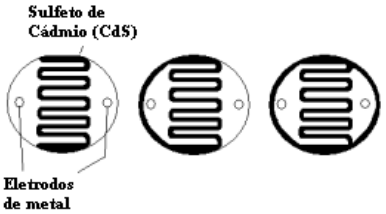


Figura 12 Três possíveis geometrias empregadas em LDRs comerciais.

A resistência de um LDR varia com a iluminância E dada em Lux (**1 Lux = 1 lúmen/m^2**) de acordo com

$$R \cong A.E^{-\alpha} \quad (1)$$

onde A e α ($0{,}5 \le \alpha \le 0{,}7$) são constantes que dependem, principalmente, se o material é CdS ou CdSe. Usando escalas logarítmicas, uma curva típica de resistência em função da iluminância é mostrada na Figura 13.

Outro parâmetro dos LDRs é o tempo de resposta da resistência a uma variação instantânea de luminosidade, as quais são diferentes na mudança de claro para escuro e de escuro para claro. Esses tempos são cerca de 10 vezes maiores quando o material fotossensível é o CdSe. Podem-se encontrar tempos de resposta de até 0,1 s. Isso limita o uso do LDR quando este é submetido a variações de luz com freqüências da ordem de dezenas de hertz.

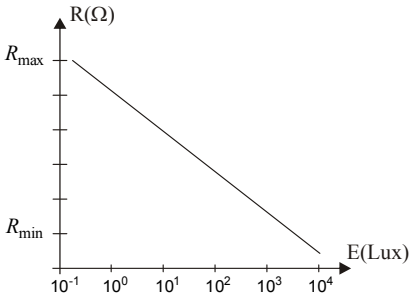


Figura 13 Resistência do LDR em função da Iluminância *E*.

O uso de LDRs em pedais de efeitos é feito usando este em um foto-acoplador, onde a fonte de luz é, por exemplo, um LED. Uma curva típica de resposta espectral de um LDR é mostrada na Figura 14, onde a sensibilidade máxima ocorre para $\lambda \approx 550$ nm. Mas é possível encontrar outros valores próximos a este, dependendo do fabricante. Pode-se excitar, conforme a Figura 14, o LDR com fontes de luz de cor branca (que contém todas as freqüências do espectro), luz verde, amarela ou laranja. A Figura 15 mostra duas curvas típicas de intensidade luminosa relativa de dois LEDs verdes que apresentam uma intersecção com a curva de resposta da Figura 14. Os LEDs verdes são feitos de vários materiais, tais como InGaN, GaP e InGaAlP, e os mais variados valores de comprimento de onda, tais como 505 nm, 555 nm e 574 nm, podem ser encontrados.

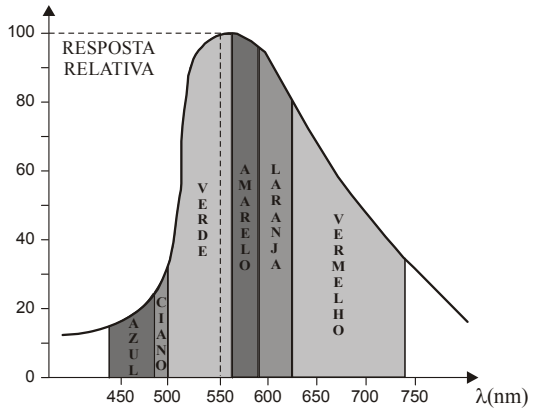


Figura 14 Curva típica da sensibilidade de um LDR em relação ao comprimento de onda da luz incidente.

No caso de LEDs de potência brancos, pode-se encontrar uma região de saturação de corrente e, conseqüentemente, de luminosidade em função da tensão (ver Figura 16). Isso pode levar a um funcionamento inadequado de um foto-acoplador.

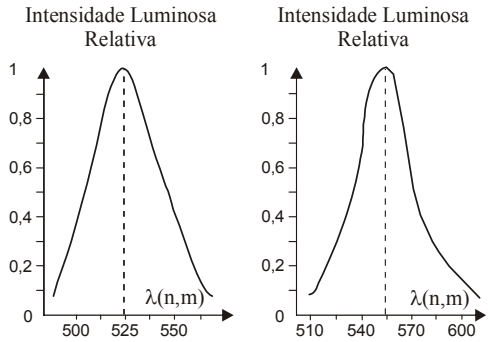


Figura 15 Curvas típicas de intensidade luminosa relativa de dois LEDs verdes.

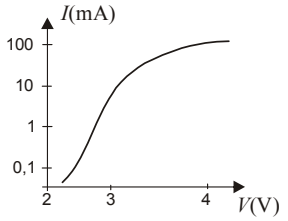


Figura 16 Curva característica de um tipo particular de LED de potência branco.

Apesar de existirem foto-acopladores comerciais, a Figura 17 apresenta uma forma de construção artesanal de um foto-acoplador. Uma forma prática de se construir, artesanalmente, um foto-acoplador é utilizar tubo termo-retrátil para fazer o acoplamento mecânico entre o LDR e o LED. Em vista da grande variedade de características de LDRs e LEDs, sugere-se aos *hobistas* que sejam feitas medidas de resistência para vários valores de corrente no LED. A partir dos dados obtidos, o pedal de efeitos pode ser projetado.

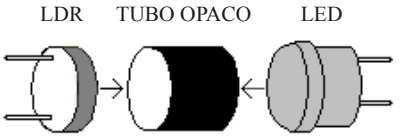


Figura 17 – Foto-acoplador artesanal.

b) Resistores Baseados em OTA

Os amplificadores operacionais de transcondutância (OTAs) são fontes de corrente controladas por tensão. Esses dispositivos apresentam a possibilidade de controle da transcondutância g_m por meio da corrente de polarização I_B [$g_m = f(I_B)$]. Normalmente, a relação entre g_m e I_B é descrita na folha de dados do fabricante, assim como aquela entre V_B, R_B e I_B [ver Figura 18(a)]. A Figura 18(b) apresenta o símbolo e o modelo linearizado de um OTA.

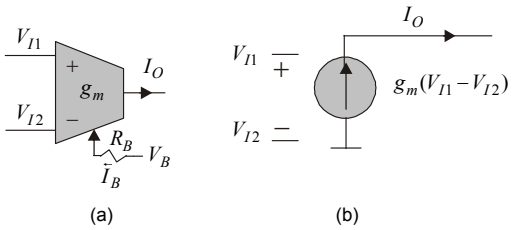


Figura 18 Amplificador de transcondutância (OTA). (a) Símbolo de um OTA ideal. (b) Modelo ideal.

Para se trabalhar com níveis de tensão de entrada da ordem de grandeza da tensão de alimentação, técnicas especiais de projeto são necessárias para minimizar as não-linearidades, entre elas o uso de circuitos não-lineares no estágio diferencial de entrada e resistores de degeneração, conforme mostrado na Figura 19. Nesse tipo de solução, o preço que se paga é que o g_m equivalente diminui na mesma proporção que aumenta o valor da tensão máxima admissível. O g_m equivalente é dado por

$$g_{mEQUI} = g_m \alpha \tag{2}$$

onde $\alpha = R_2/(R_1 + R_2)$.

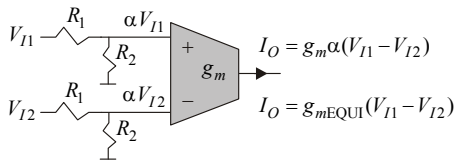


Figura 19 Uso de resistores de degeneração para aumentar a linearidade.

Na Figura 20, são mostrados dois resistores variáveis com tensão. Na Figura 20(a), o resistor é aterrado. Já na Figura 20(b), o resistor é flutuante.

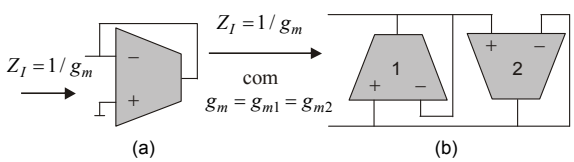


Figura 20 Resistor variável com a tensão. (a) Aterrado. (b) Flutuante.

A Figura 21 mostra o circuito para o cálculo da resistência vista entre os terminais A e B do resistor aterrado. Como $I' = 0$,

$$I = -I_0 = g_m V \,. \tag{3}$$

Logo,

$$R_I = V / I = V / g_m V = 1 / g_m \,. \tag{4}$$

Deve ser salientado que o OTA deve ser usado com resistores de degeneração, conforme mostrado na Figura 19.

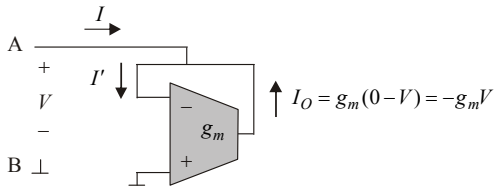


Figura 21 Circuito para o cálculo da resistência do resistor aterrado.

c) Resistores Baseados em JFET

Os JFETs podem operar como resistores aproximadamente lineares controlados por tensão, para pequenos valores de V_{DS}, em uma sub-região da chamada região ôhmica, onde $V_{DS} < |V_P|$, (tensão de *pinch-off*). Na região de resistência aproximadamente linear, os valores de V_{DS} são da ordem de poucas centenas de milivolts e dependem do JFET usado.

A Figura 22 mostra uma curva típica da corrente de dreno I_D em função da tensão V_{DS} de um JFET canal N. A Figura 23 mostra detalhes da região ôhmica próxima a $V_{DS} = 0$. Essa região se estende para valores de $V_{DS} < 0$ porque aqui o JFET apresenta um comportamento aproximadamente simétrico.

A resistência aproximadamente linear r_D (ver Figura 23) é dada por

$$r_D \cong \frac{r_0}{\left(1 - \frac{V_{GS}}{V_P}\right)^2} \tag{5}$$

onde r_0 é a resistência para $V_{GS} = 0$ (ver Figura 24).

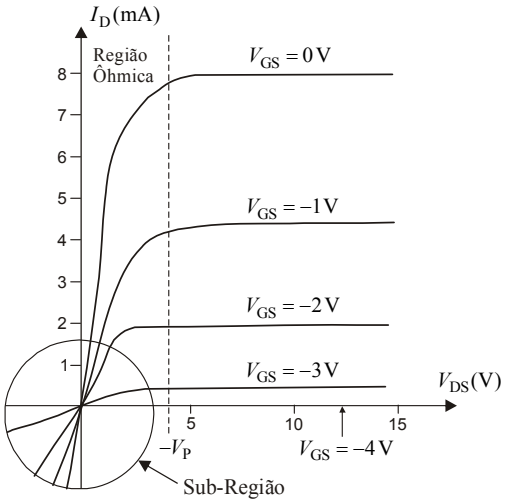


Figura 22 Curvas de corrente de dreno I_D em função da tensão V_{DS}, para vários valores de tensão V_{GS}, para um JFET canal n.

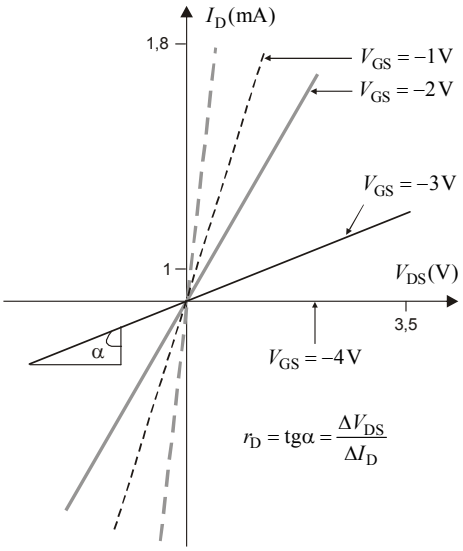


Figura 23 Detalhe das curvas na região ôhmica.

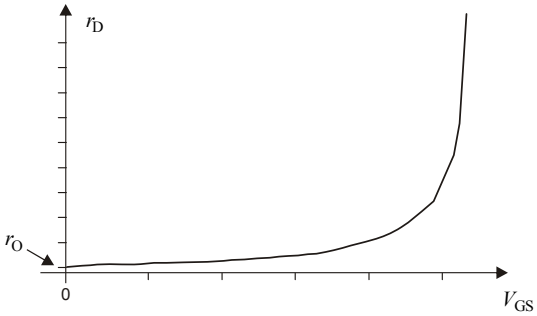


Figura 24 Resistência r_D em função da tensão V_{GS}.

A Figura 25(a) mostra um amplificador com ganho controlado por tensão que pode ser aplicado em um tremolo, por exemplo. A Figura 25(b) mostra um JFET operando como resistor variável em um filtro passa-tudo com aplicação em um vibrato ou em um *phaser*, por exemplo. A Figura 25 (c) mostra uma forma alternativa de implementar um resistor controlado por tensão. O circuito *RC* série entre o dreno e o *gate* é usado para minimizar distorções. A única condição é que $R >> r_{\mathrm{Dmax}}$ e $1 / \omega C << R$ nas freqüências de trabalho.

2.2 Osciladores

Formas de ondas quadradas, senoidais e triangulares são úteis em vários tipos de pedais de efeito analógicos, como *phaser*, *chorus*, *flanging*, vibrato, tremolo, wah-wah, etc. Dois tipos de osciladores têm sido usados em pedais de efeitos analógicos. Interessam, em particular, somente os osciladores cuja variação da freqüência é obtida com a variação de um único resistor.

Existem diversas sugestões de osciladores na rede. Aqui é mostrado um tipo simples de oscilador não-linear formado pelo laço de um integrador e um biestável. Nesse caso, uma onda quadrada e outra triangular são geradas simultaneamente (ver Figura 26).

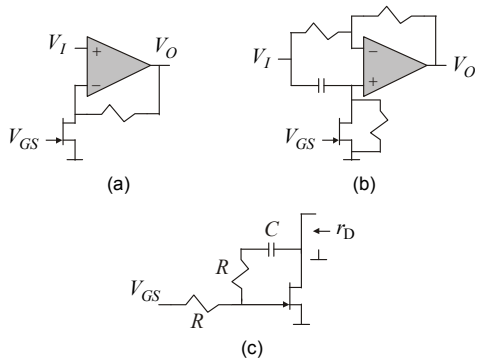


Figura 25 (a) Amplificador com ganho controlado por tensão. (b) Resistor variável aplicado em um filtro passa-tudo com fase controlada por tensão. (c) Resistor variável para uso geral.

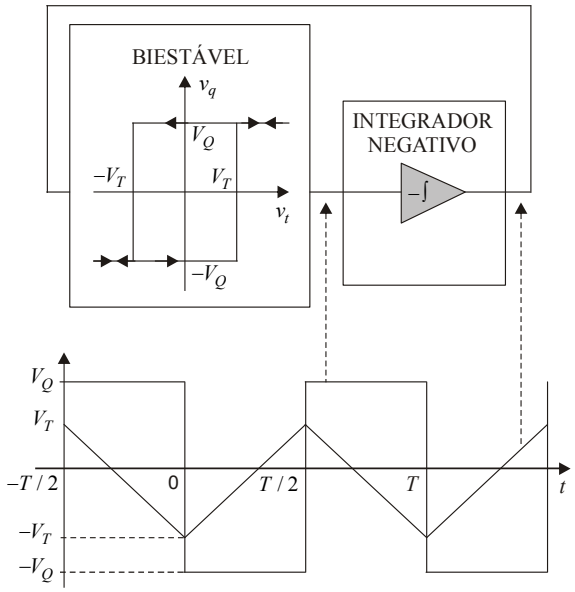


Figura 26 Oscilador de onda quadrada e triangular.

Um circuito simples que implementa o diagrama de blocos da Figura 26 é apresentado na Figura 27.

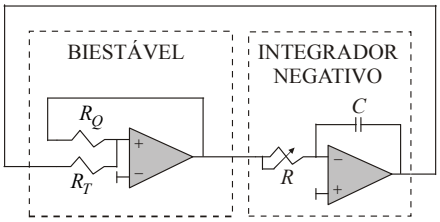


Figura 27 Simples implementação do gerador de onda quadrada e triangular.

Vamos supor que a tensão na saída do biestável seja $-V_Q$. Assim, na na saída do integrador negativo, a tensão é dada por

$$v_T(t) = -\sigma_0 \int -V_Q \, dt = \sigma_0 V_Q t - v_T(0) = \sigma_0 V_Q t - V_T \quad (6)$$

onde $\sigma_0 = 1/RC$. Como

$$v_T(T/2) = \sigma_0 V_Q (T/2) - V_T = V_T \quad (7)$$

e

$$v_T(0) = -V_T \quad (8)$$

deduz-se que

$$\sigma_0 V_Q (T/2) = 2V_T \,. \quad (9)$$

Logo, a freqüência de oscilação é dada por

$$f_0 = \frac{1}{T} = \frac{V_Q \sigma_0}{4V_T} = \frac{V_Q}{V_T} \times \frac{1}{4RC} \,. \quad (10)$$

Observa-se no circuito da Figura 27 que, no momento da troca de estado do biestável, tem-se:

$$\frac{V_Q}{R_Q} = -\frac{V_T}{R_T}. \quad (11)$$

Assim,

$$\left|\frac{V_Q}{V_T}\right| = \frac{R_Q}{R_T}. \quad (12)$$

A tensão V_Q no circuito da Figura 27 é a tensão de saturação do amplificador, cerca de 1 V abaixo da tensão de alimentação. A tensão de pico da onda triangular é

$$|V_T| = |V_Q|(R_T / R_Q) \,. \quad (13)$$

2.3 O Dispositivo *Bucket-Brigade* (BBD)

O BBD é um registrador de deslocamento (*shift register*) analógico que opera como uma linha de atraso. Ele tem aplicações em pedais de eco, *chorus* e *flanger*. A razão do nome é uma analogia com o transporte de água por meio de uma brigada de pessoas transferindo água de um balde para outro subseqüente. Analogamente, no caso do BBD, as cargas dos capacitores são transferidas para os capacitores subseqüentes. A Figura 28 mostra a operação de um BBD de uma forma simplificada. O tempo de atraso do sinal de saída depende do número de estágios (que é igual ao número de capacitores) e do período de *clock T*. Por exemplo, para 1024 estágios e $T = 0{,}1$ ms, correspondente a uma freqüência de amostragem de 10 kHz, o tempo de atraso de uma amostra na saída é de

$$1024 \times (0{,}1\,\mathrm{ms} / 2) = 51{,}2\,\mathrm{ms} \,. \quad (14)$$

A divisão por dois é necessária porque cada amostra é processada na metade do período. Por serem sistemas analógicos amostrados, os sistemas com BBD necessitam de filtros anti-recobrimento e de reconstrução, na entrada e na saída, respectivamente.

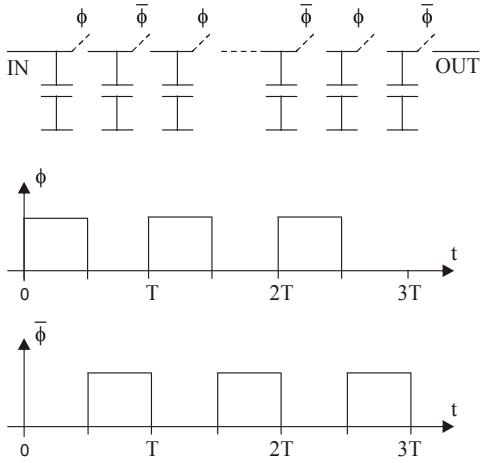


Figura 28 forma simplificada de operação de um BBD.

2.4 Filtros

a) Introdução

Uma das formas mais simples de projetar filtros ativos [4] para aplicações em áudio em geral, e particularmente em pedais de efeitos, é aquela que faz uso de amp. op., capacitores e resistores discretos. São os chamados filtros ativos-RC. Suas principais vantagens são:

i. Dispensam o uso de indutores.
ii. Os componentes são facilmente disponíveis.
iii. Em grande parte dos casos, o projeto é simples (sem influência da impedância do alto-falante no caso de *crossovers* ativos).
iv. Estágios podem ser cascateados sem alteração das funções de transferência individuais. Assim, funções de ordem maior do que dois podem ser obtidas a partir do projeto de redes de segunda e primeira ordem.
v. Em relação aos filtros passivos, a montagem pode ser mais compacta, especialmente no caso de filtros para freqüências mais baixas.
vi. Em relação aos *crossovers* passivos, filtros ativos com três amplificadores possibilitam a utilização de amplificadores com menor potência.

As topologias ativas-RC mais usadas que realizam uma função com pólos reais são mostradas na Figura 29.

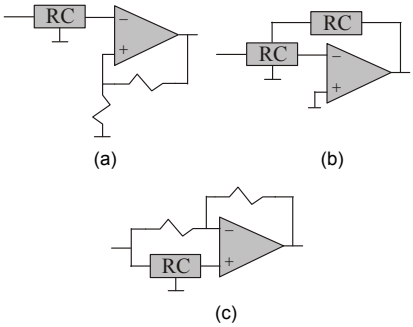


Figura 29 Estruturas de primeira ordem. (a) Filtro passa-baixa ou passa-alta com ganho positivo. (b) Filtro passa-baixa ou passa-alta com ganho positivo. (c) Filtro passa-tudo.

Os filtros ativos de segunda ordem permitem associação em cascata e a conseqüente obtenção de filtros de ordem alta. O valor da ordem é limitado a 10 ou 11 devido aos desvios estatísticos e/ou determinísticos ocasionados pelas sensibilidades e variabilidades não nulas dos componentes.

Os filtros ativos-RC de segunda ordem mais econômicos são os que fazem uso de apenas um amp.op. São os chamados SAB (*single-amplifier biquad*). As estruturas mais usadas são a estrutura Sallen-Key, na qual o amp. op. é utilizado para realizar um amplificador de ganho positivo *K*, e a estrutura *multiple feedback* (MFB), onde o elemento ativo é o próprio amp. op., cujo ganho de tensão ideal tende ao infinito.

As duas topologias básicas são mostradas na Figura 30.

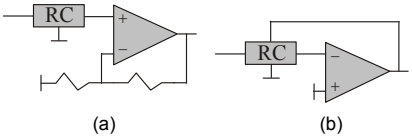


Figura 30 Estruturas para obtenção de pólos e zeros complexos. (a) Rede Sallen-Key; (b) Rede MFB.

As duas estruturas proporcionam diferentes graus de liberdade no que diz respeito à variação da freqüência (ω_0), do fator de qualidade dos pólos (*Q*), do ganho da função (*K*) e, no caso de filtros passa-faixa, da razão ω_0/Q (banda). A opção por uma ou por outra topologia deve levar em conta também as diferentes características de sensibilidade, a inversão do sinal (caso da estrutura MFB), a complexidade do projeto, o número de componentes necessários, o nível de impedância de entrada, a independência do ganho em relação aos outros parâmetros, etc. Aqui são apresentadas sugestões para o cálculo dos valores dos elementos em função dos coeficientes das funções de transferência. Como é simples a relação entre os coeficientes e os parâmetros ω_0, *Q* e *K*, as relações podem ser facilmente modificadas para que os elementos sejam encontrados em função destes últimos. Algumas relações são sugeridas, porém várias combinações são possíveis, uma vez que o número de componentes das redes é maior do que o número de coeficientes das funções. Outras combinações possíveis podem ser buscadas com o intuito, por exemplo, de diminuir a sensibilidade ou o espalhamento no valor dos elementos.

b) Filtros Passa-Baixa

Os filtros Passa-Baixa têm como principais aplicações:

i. Célula base de um equalizador *shelving* para reforço e atenuação de graves (nesse caso, também podem ser utilizadas estruturas mais simples de primeira ordem).
ii. Filtros anti-recobrimento e de reconstrução em pedais de eco, reverberadores, *flanger* e *chorus* implementados com BBD e circuitos digitais.
iii. Filtros para eliminação de eventuais ruídos e/ou interferências eletromagnéticas de altas freqüências em todos os tipos de pedais.
iv. Proteção para os *drivers* de alta freqüência (filtro ultra-sônico).
v. *Crossovers*.

Filtros de primeira ordem podem ser realizados por estruturas não-inversoras ou inversoras, mostradas na Figura 31.

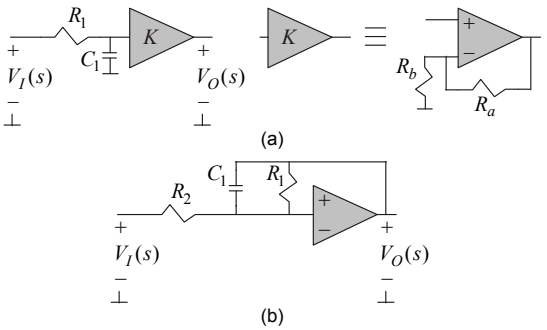


Figura 31 Filtros passa-baixa de primeira ordem. (a) Configuração não-inversora. (b) Configuração inversora.

As funções de transferência das redes das Figuras 31(a) e 31(b) (que são constantes na origem e zero no infinito) são dadas por

$$T_a(s) = T_b(s) = \frac{K\,\sigma_0}{s+\sigma_0} \tag{15}$$

onde

$$\sigma_0 = 1/R_1C_1 \tag{16}$$

e

$$K = (R_a + R_b)/R_b \tag{17}$$

para a rede da Figura 31(a), ou

$$K = -R_1/R_2 \tag{18}$$

para a rede da Figura 31(b).

A função passa-baixa de segunda ordem (que é constante na origem e zero no infinito) é dada por

$$T_{\mathrm{PB}}(s) = \frac{K\omega_0^2}{s^2 + (\omega_0/Q)s + \omega_0^2} = \frac{A}{s^2 + a_1 s + a_0} \tag{19}$$

onde K é o ganho em baixa freqüência, ω_0 é o módulo dos pólos (para pólos complexos) e Q é o fator de qualidade dos pólos. A Figura 32 mostra a magnitude para $\omega_0 = 1$ rad/s, $K = 1$ e Q variável.

A Figura 33 mostra em (a) e (b) os filtros SAB passa-baixa, Sallen-Key e MFB, respectivamente.

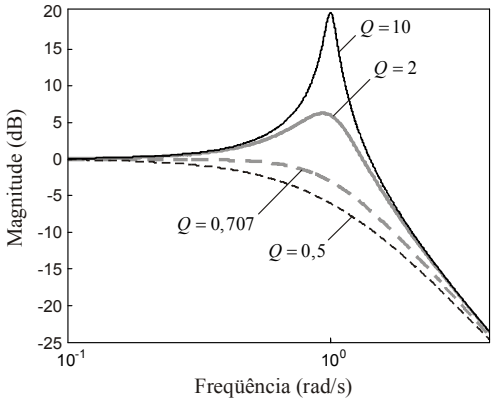


Figura 32 Funções PB normalizadas com Q variável.

Os elementos de um filtro passa-baixa Sallen-Key podem ser encontrados em função dos coeficientes a_1 e a_0:

$$C_1 = C_2 = \frac{1}{b\sqrt{a_0}} \tag{20}$$

$$R_1 = R_2 = R_b = b \tag{21}$$

$$R_a = 2b - \frac{ba_1}{\sqrt{a_0}} \tag{22}$$

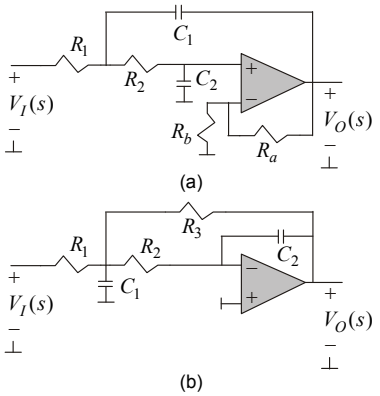


Figura 33 Filtros PB. (a) Rede SK. (b) Rede MFB.

O ganho resultante em baixas freqüências é

$$K = 3 - (1/Q) \tag{23}$$

ou

$$K = 3 - (a_1/\sqrt{a_0})\,. \tag{24}$$

Para

$$Q = \sqrt{a_0}/a_1 \geq 0{,}5 \tag{25}$$

(pólos complexos) tem-se $K \geq 1$. Por exemplo, para $Q = 0{,}5$, $K = 1$. Para $Q = 1$, $K = 2$. Se for desejado abaixar o ganho sem alterar o valor de ω_0 e de Q, pode-se usar um divisor de tensão resistivo na entrada, como mostra o exemplo prático 5. O valor de b em (20), (21) e (22) pode ser atribuído pelo projetista da seguinte forma: escolhe-se um valor comercial para C_1; o valor de b é

$$b = 1/C_1\sqrt{a_0}\,. \tag{26}$$

Se o ganho em baixa freqüência for unitário ($K = 1$), faz-se $R_b \to \infty$ e $R_a = 0$ e os valores são dados por:

$$C_1 = \frac{2}{ba_1} \tag{27}$$

$$C_2 = \frac{a_1}{2ba_0} \tag{28}$$

$$R_1 = R_2 = b. \tag{29}$$

Os elementos de um filtro passa-baixa MFB podem ser encontrados em função dos coeficientes a_1, a_0 e A:

$$C_1 = \frac{2}{ba_1} + \frac{a_0}{ba_1|A|} \tag{30}$$

$$C_2 = \frac{1}{bC_1|A|}\ \mathrm{F} \tag{31}$$

$$R_1 = R_2 = b \tag{32}$$

$$R_3 = \frac{b|A|}{a_0}. \tag{33}$$

c) Filtros Passa-Alta

Os filtros Passa-Alta têm como principais aplicações:

i. Célula base de um equalizador *shelving* para reforço e atenuação de agudos (nesse caso, podem ser utilizadas estruturas mais simples de primeira ordem).

ii. Proteção para os alto-falantes de baixa freqüência (filtro subsônico).
iii. Desacoplamento DC. Permite que estágios sejam acoplados sem que as polarizações sejam afetadas.
iv. *Crossovers*.

Filtros passa-alta de primeira ordem podem ser realizados pelas estruturas não-inversoras ou inversoras, mostradas na Figura 34. Essas estruturas podem ser usadas na cascata de filtros de ordem ímpar.

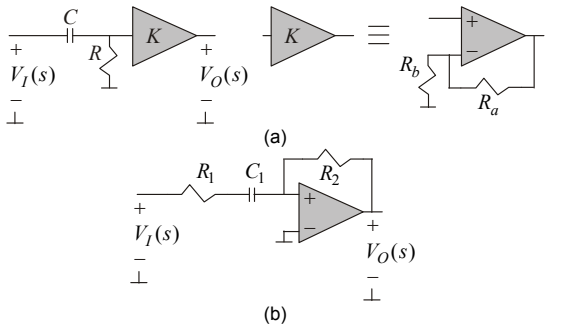


Figura 34 Filtros passa-alta de primeira ordem. (a) Configuração não-inversora. (b) Configuração inversora.

As funções de transferência das redes da Figuras 34(a) e 34(b) (que são zero na origem e constantes no infinito) são dadas por

$$T_a(s) = T_b(s) = K\frac{s}{s+\sigma_0} \tag{34}$$

onde

$$\sigma_0 = 1/R_1C_1 \tag{35}$$

e

$$K = (R_a + R_b)/R_b \tag{36}$$

para a rede da Figura 34(a), ou

$$K = -R_1/R_2 \tag{37}$$

para a rede da Figura 34(b).

Os filtros passa-alta de segunda ordem têm sua função de transferência geral dada por

$$T(s) = \frac{Ks^2}{s^2 + (\omega_0/Q)s + \omega_0^2} = \frac{Ks^2}{s^2 + a_1 s + a_0} \tag{38}$$

onde K é o ganho em alta freqüência, ω_0 é o módulo dos pólos (para pólos complexos) e Q é o fator de qualidade dos pólos.

As Figuras 35(a) e 35(b) mostram os filtros SAB passa-alta, Sallen-Key e MFB, respectivamente.

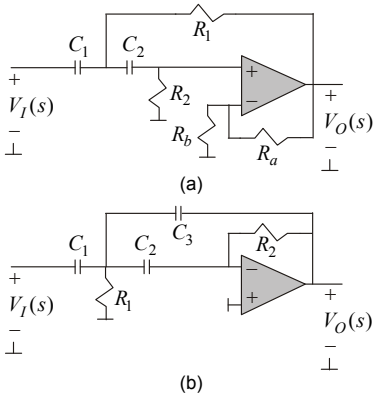


Figura 35 Filtros passa-alta. (a) Rede Sallen-Key. (b) Rede MFB.

Os elementos de um filtro passa-alta Sallen-Key podem ser encontrados em função dos coeficientes a_1 e a_0:

$$C_1 = C_2 = \frac{1}{b\sqrt{a_0}} \tag{39}$$

$$R_1 = R_2 = R_b = b \tag{40}$$

$$R_a = 2b - \frac{ba_1}{\sqrt{a_0}}. \tag{41}$$

O ganho resultante em altas freqüências é

$$K = 3 - (1/Q) \tag{42}$$

ou

$$K = 3 - (a_1/\sqrt{a_0}). \tag{43}$$

Para

$$Q = \sqrt{a_0}/a_1 \geq 0{,}5 \tag{44}$$

(pólos complexos) resulta em $K \geq 1$. Por exemplo, para $Q = 0{,}5$, $K = 1$. Para $Q = 1$, $K = 2$. Se for desejado abaixar o ganho sem alterar ω_0 e Q, pode-se usar um divisor de tensão capacitivo na entrada, como mostrado no exemplo prático 5.

Se o ganho em alta freqüência for unitário ($K = 1$), faz-se $R_b \to \infty$ e $R_a = 0$ e os valores são dados por

$$C_1 = C_2 = \frac{1}{b} \tag{45}$$

$$R_1 = \frac{ba_1}{2a_0} \tag{46}$$

$$R_2 = \frac{2b}{a_1}. \tag{47}$$

Os elementos de um filtro passa-alta MFB podem ser encontrados em função dos coeficientes a_1, a_0 e K:

$$C_1 = C_2 = \frac{1}{b\sqrt{a_0}} \tag{48}$$

$$C_3 = \frac{1}{b\sqrt{a_0}\,|K|} \tag{49}$$

$$R_1 = \frac{ba_1/\sqrt{a_0}}{2 + 1/|K|} \tag{50}$$

$$R_2 = \frac{b\sqrt{a_0}}{a_1}(2|K| + 1). \tag{51}$$

A Figura 36 mostra a magnitude para $\omega_0 = 1$, $K = 1$ e Q variável.

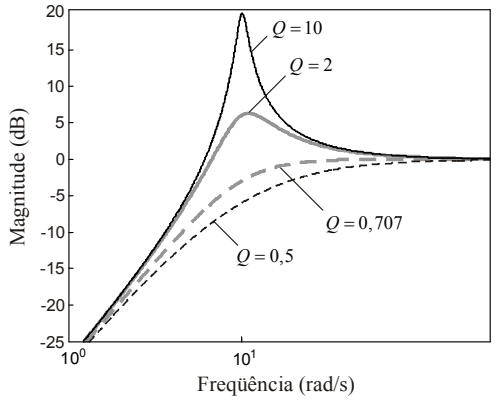


Figura 36 Funções passa-alta normalizadas com fator de qualidade variável.

Exemplo prático 1: *Quando dois circuitos quaisquer são colocados em cascata, não é desejado que as respectivas polarizações sejam afetadas. Por isso deve ser usado um acoplamento AC, ou seja, um acoplamento capacitivo. Conseqüentemente, aparece um efeito passa-alta devido à criação de um zero na origem e um pólo finito, cujo valor depende da capacitância e da resistência vista pelos seus terminais. A Figura 37 mostra um exemplo onde* R_{IN} *representa a impedância de entrada do circuito na saída de um pedal de efeitos,* R_{VAR} *representa um eventual controle de volume de saída e* R_{PROT} *protege o circuito de eventuais curtos-circuitos na saída.*

A freqüência do pólo criado em baixa freqüência é dada por

$$f_p = (1/2\pi C(R_{IN} + R_{OUT})) . \quad (52)$$

Supondo que é desejada uma faixa plana (dentro de ≈−0,1 dB*) para um sinal de guitarra, cuja freqüência mais baixa é* f_{min} = 82,41 Hz *(para um sinal de contra baixo,* f_{min} = 41,2 Hz*), o pólo deve ser cerca de 15 % de* f_{min}*, ou seja,* f_p = 12,36 Hz.

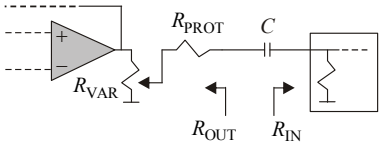


Figura 37 Acoplamento típico entre pedais de efeito.

O pior caso sob o ponto de vista do valor do capacitor é quando $R_{IN} + R_{OUT}$ *é mínimo. O mínimo valor de* R_{OUT} *é* $R_{IN} = R_{OUT}$*. Alguns fabricantes recomendam um valor mínimo de impedância de entrada para o circuito a ser excitado de* R_{IN} = 10 kΩ*. Para* R_{PROT} = 470 Ω *(valor típico usado), o valor mínimo do capacitor deve ser:*

$$C = \frac{1}{2\pi f_p (R_{IN} + R_{OUT})} = \frac{1}{2\pi \times 12.36 \times 10.470} = 1{,}23\,\mu\text{F} \quad (53)$$

Se N pedais forem colocados em cascata, a queda de sinal em f_{min} *é aproximadamente* $N \times$ 0,1 dB*. Assim, com 6 acoplamentos, a queda em* f_{min} *é aproximadamente −0,6 dB. A Figura 38 mostra um zoom (em torno de* f_{min}*) na magnitude da resposta em freqüência em função do número de acoplamentos (N = 1,...,6). Porém, muitos fabricantes projetam seus pedais com impedância de entrada da ordem de centenas de quiloohms. Então, normalmente, essa não deve ser uma preocupação de guitarristas. Mas hobistas devem estar atentos a esses detalhes.*

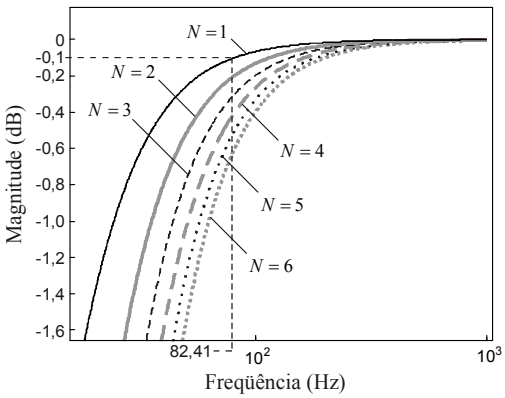


Figura 38 *Zoom* em torno de f_{min} para vários números de estágios.

d) Filtros Passa-Faixa

Os filtros Passa-Faixa têm como principais aplicações:

i. Célula base em pedal de efeito wah-wah.
ii. Célula base de equalizadores gráficos, paramétricos e paragráficos.
iii. *Crossovers*.

A função passa-faixa de segunda ordem, que é zero na origem e zero no infinito, é dada por

$$T(s) = \frac{K(\omega_0 / Q)\, s}{s^2 + (\omega_0 / Q)s + \omega_0^2} = \frac{KB\, s}{s^2 + Bs + \omega_0^2} = \frac{A\, s}{s^2 + Bs + a_0} \quad (54)$$

onde K é o ganho na freqüência central ω_0 e Q é o fator de qualidade dos pólos.

Nas funções dos filtros passa-faixa, tem-se que:

$$Q = \omega_0 / (\omega_2 - \omega_1) \quad (55)$$

e

$$\omega_0 = (\omega_1 \omega_2)^{1/2} \quad (56)$$

onde ω_2 e ω_1 são as freqüências de corte de 3 dB superior e inferior, respectivamente, e a banda B é dada por

$$B = \omega_2 - \omega_1 . \quad (57)$$

Note na Figura 39 e na equação que relaciona Q e B que, à medida que o fator $Q \to \infty$, a largura de banda $B \to 0$.

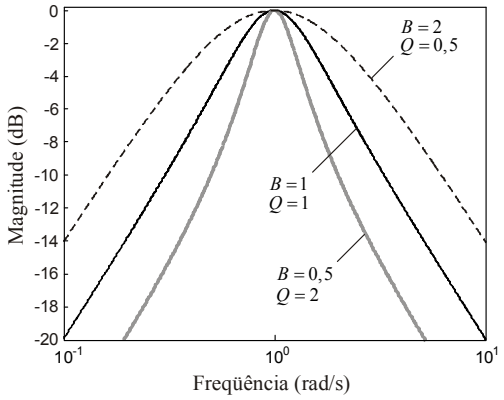


Figura 39 Funções PF normalizadas com *Q* e *B* variáveis.

A Figura 40 mostra os filtros SAB passa-faixa, Sallen-Key [Figura 40(a)] e MFB[Figura 40(b)].

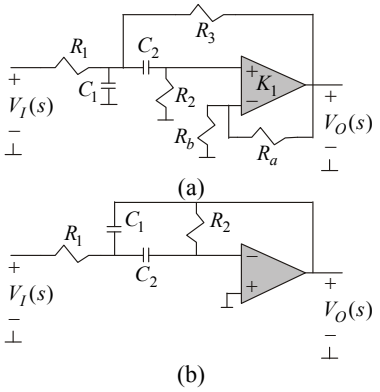


Figura 40 Filtros passa-faixa. (a) Rede Sallen-Key. (b) Rede MFB.

Os elementos de um filtro passa-faixa Sallen-Key podem ser encontrados em função dos coeficientes A, B e a_0. O ganho máximo obtido na freqüência ω_0 é $K = A/B$.

$$C_1 = \frac{1}{b\sqrt{a_0}} \tag{58}$$

$$C_2 = \frac{2}{bR_2\sqrt{a_0}} \tag{59}$$

$$R_1 = R_3 = b \tag{60}$$

$$R_2 = \frac{b\sqrt{a_0}}{A + B - 2{,}5\sqrt{a_0}} \tag{61}$$

$$K_1 = \frac{A}{\sqrt{a_0}} = \frac{R_a + R_b}{R_b} \tag{62}$$

Se o valor de $A = KB$ não for suficientemente grande de modo que

$$A + B - 2{,}5\sqrt{a_0} > 0\ , \tag{63}$$

deve-se forçar um aumento de K para que $R_2 > 0$. Depois o ganho K deve ser diminuído com a colocação de um divisor de tensão na entrada (ver Exemplo Prático 5).

Os elementos de um filtro passa-faixa MFB podem ser encontrados em função dos coeficientes B e a_0:

$$C_1 = C_2 = \frac{B}{2ba_0} \tag{64}$$

$$R_1 = b \tag{65}$$

$$R_2 = \frac{4ba_0}{B^2} \tag{66}$$

O módulo do ganho obtido na freqüência ω_0 é

$$|K| = 2a_0 / B^2 = 2Q^2\ . \tag{67}$$

Assim, para pólos com $Q < 0{,}707$, obtém-se $|K| < 1$. Se for desejado abaixar o ganho sem alterar ω_0 e Q, pode-se usar um divisor de tensão resistivo na entrada (ver Exemplo Prático 5).

Exemplo prático 2: *Foi comentado que filtros passa-baixa e passa-alta podem ser usados para realizar filtros ultra-sônicos e subsônicos, respectivamente. Mas porque não usar um passa-faixa? As razões são o aumento da sensibilidade e a diminuição da faixa dinâmica. Vamos ver um exemplo onde é desejado um filtro passa-alta com* $f_A = 20$ Hz *e* $Q = 0{,}707$, *e um filtro passa-baixa com* $f_B = 20$ kHz *e* $Q = 0{,}707$.

Em termos do produto de uma função passa-baixa multiplicada por uma função passa-alta, a função passa-faixa é:

$$T_{\mathrm{PF}}(s) = \frac{\omega_B^2}{s^2 + (\omega_B / Q)s + \omega_B^2} \times \frac{s^2}{s^2 + (\omega_A / Q)s + \omega_A^2} \tag{68}$$

$$T_{\mathrm{PF}}(s) = \frac{(2\pi.2\times10^4)^2}{s^2 + \left(\dfrac{2\pi.2\times10^4}{0{,}707}\right)s + (2\pi.2\times10^4)^2} \times \frac{s^2}{s^2 + \left(\dfrac{2\pi.20}{0{,}707}\right)s + (2\pi.20)^2} \tag{69}$$

Como o ganho em baixa freqüência e em alta freqüência é unitário, faz-se $R_b \to \infty$ *e* $R_a \to 0$ *nas duas redes (passa-baixa Sallen-Key e passa-alta Sallen-Key) e o projeto é feito usando as equações já apresentadas para o cálculo dos componentes. Para a rede passa-baixa com* $b = 1$, *tem-se:* $C_1 = 11{,}254$ μF, $C_2 = 5{,}627$ μF *e* $R_1 = R_2 = 1\ \Omega$. *Para a rede passa-alta, tem-se* $C_1 = C_2 = 1$ F, $R_1 = 0{,}005627\ \Omega$ e R2 = 0,011254 Ω.

A Figura 41 mostra a cascata das duas redes. A ordem dos dois estágios é proposital. O estágio passa-baixa na entrada minimiza problemas de slew-rate, e o estágio passa-alta na saída minimiza problemas de offset.

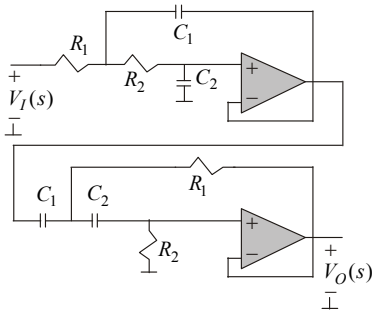


Figura 41 Cascata de redes passa-baixa e passa-alta.

Em termos do produto de uma função passa-faixa multiplicada por outra função passa-faixa, a função é:

$$T_{\mathrm{PF}}(s) = \frac{K(\omega_A / Q)\, s}{s^2 + (\omega_A / Q)s + \omega_0^2} \times \frac{K(\omega_B / Q)\, s}{s^2 + (\omega_B / Q)s + \omega_0^2} \tag{70}$$

$$T_{\mathrm{PF}}(s) = \frac{\sqrt{500}(2\pi.20 / 0{,}707)\, s}{s^2 + (2\pi\times 20 / 0{,}707)s + (2\pi.20)^2} \times \frac{\sqrt{500}(2\pi.2.10^4 / 0{,}707)\, s}{s^2 + (2\pi.2.10^4 / 0{,}707)s + (2\pi.2.10^4)^2} \tag{71}$$

No caso da rede MFB passa-faixa, o módulo do ganho na freqüência central é dado por

$$|K| = 2Q^2 = 1. \tag{72}$$

Nesse caso, como o ganho necessário é muito mais alto, ou seja, $\sqrt{500}$ *, opta-se pelas redes Sallen-Key passa-faixa, apresentadas na Figura 42.*

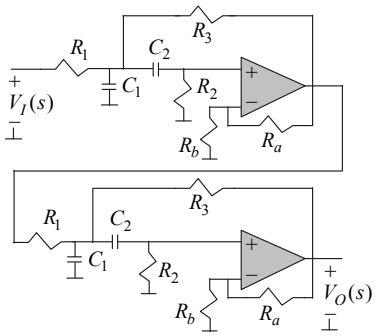


Figura 42 Cascata de duas redes passa-faixa.

Usando as equações já apresentadas para o cálculo dos componentes, para $b = 1$, *obtêm-se* $C_1 = 7{,}957$ mF, $C_2 = 0{,}486$ F, $R_1 = R_3 = 1\ \Omega$ e $R_2 = 32{,}74$ mΩ. Assim, $K_1 = 31{,}62$ *e* $R_b = 1\ \Omega$. Logo, $R_a = 30{,}623\ \Omega$.

Com exceção dos capacitores, que devem possuir um valor 1000 vezes menor, os elementos do filtro PF2 são idênticos aos do filtro PF1.

A Figura 43 apresenta as magnitudes das respostas de todos os estágios individuais e também das duas cascatas. Para as cascatas, atribuiu-se, para todos os amp. op.,

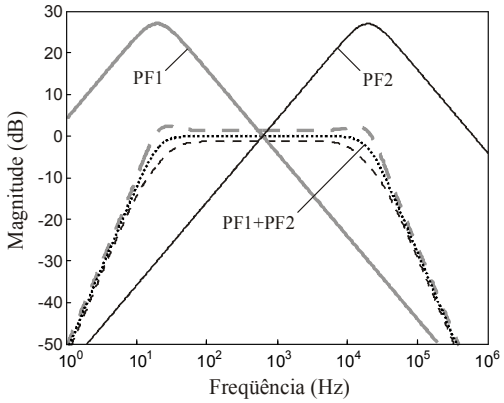
GB = 6,5 MHz *e* $R_0 = 0$*, com tolerância de 30 % para o GB, 1 % para os resistores e 10 % para os capacitores. Para a rede formada por dois passas-faixas (PF-PF), encontram-se desvios de até* ±4 dB *na banda de passagem. Para a rede passa-baixa e passa-alta (PB-PA), encontram-se desvios de aproximadamente* 0 dB *na banda de passagem e um máximo de* ±1,23 dB *fora da banda. Isso se deve ao alto ganho exigido dos estágios individuais na rede PF-PF. Porém, essa não é a única desvantagem da cascata PF-PF em relação à cascata PB-PA. A análise da faixa dinâmica mostrou que, para uma entrada de* 10 V_{PICO}*, a rede PB-PA apresenta uma tensão máxima de saída de* 10 V_{PICO}*. A rede PF-PF apresenta uma tensão máxima de saída "teórica" de* 223,6 V_{PICO}*. Logo, para uma saída máxima de* 10 V_{PICO}*, a máxima tensão admissível na entrada da rede PF-PF é* 0,4472 V_{PICO}*, reduzindo a faixa dinâmica em* 27 dB.

Exemplo prático 3: Este é um contra-exemplo em relação ao Exemplo prático 2. Vamos comparar novamente uma cascata de redes passa-faixa (PF-PF) com a cascata de uma rede passa-baixa e outra passa-alta (PBPA), só que agora as redes passa-faixa 1 e 2 apresentam um ganho unitário na freqüência central normalizada de 1 rad/s e fator de qualidade Q = 0,707. A função da rede PFPF é

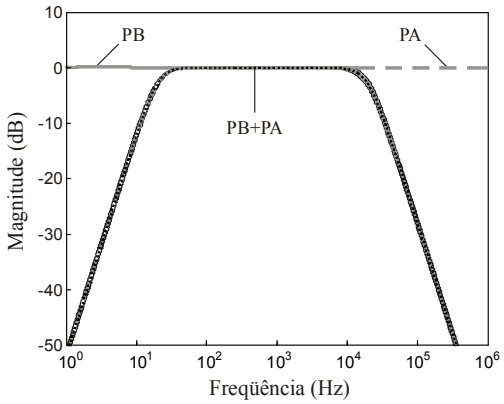
$$T_{\mathrm{PFPF}}(s) = \frac{1\times\sqrt{2}\,s}{s^2+\sqrt{2}s+1}\times\frac{1\times\sqrt{2}\,s}{s^2+\sqrt{2}s+1} \tag{73}$$

e a função da rede PBPA é dada por

$$T_{\mathrm{PBPA}}(s) = \frac{\sqrt{2}\times 1}{s^2+\sqrt{2}s+1}\times\frac{\sqrt{2}\,s^2}{s^2+\sqrt{2}s+1}. \tag{74}$$



(a)



(b)

Figura 43 Magnitude da resposta dos estágios individuais e também das duas cascatas. (a) Rede PF-PF. (b) Rede PB-PA.

Como neste caso, para a rede passa-faixa MFB, tem-se que

$$|K| = 2Q^2 = 1 \tag{75}$$

o projeto será realizado com essa topologia. Para b = 1, usando as equações de projeto já apresentadas, obtém-se para as duas redes passa-faixa MFB: $R_1 = 1\ \Omega$, $R_2 = 2\ \Omega$ e $C_1 = C_2 = 0{,}707$ F.

Usando as equações de projeto para a rede passa-baixa MFB, obtém-se: $R_1 = R_2 = 1\ \Omega$, $R_3 = \sqrt{2}\,\Omega$, $C_1 = 1{,}91$F e $C_2 = 0{,}369$ F.

Usando as equações de projeto para a rede passa-alta MFB, obtém-se: $C_1 = C_2 = 1$ F, $C_3 = 0{,}707$ F, $R_1 = 0{,}52\ \Omega$ e $R_2 = 2{,}707\ \Omega$.

As Figuras 44 e 45 apresentam as redes PF-PF e PB-PA (MFB), respectivamente.

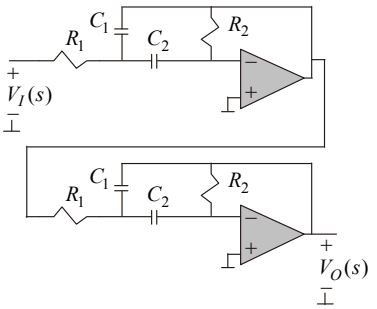


Figura 44 Rede PF-PF.

Para a rede PBPA, encontra-se um desvio de ±1,2 dB*. A análise da faixa dinâmica mostrou que, para uma entrada de* 10 V_{PICO}*, a rede PFPF apresenta uma tensão máxima de saída de* 10 V_{PICO}*.*

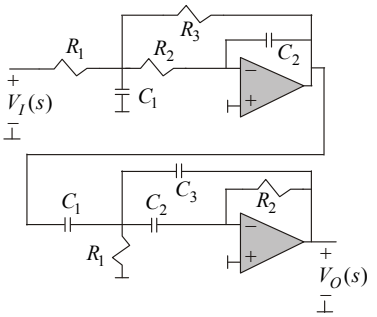


Figura 45 Rede PB-PA.

A rede PBPA apresenta uma tensão máxima de saída "teórica" de 14,14 V_{PICO}*, reduzindo a faixa dinâmica em* 3 dB*. Então, ao contrário do Exemplo prático 2, aqui a cascata PFPF apresentou um melhor desempenho. Isso porque, nas redes PBPA, os ganhos individuais das funções passa-baixa e passa-alta nas bandas de passagem são maiores do que 1 e também porque apresentam um número maior de elementos passivos.*

e) Filtros Passa-Tudo.

Um filtro passa-tudo apresenta uma função de transferência da seguinte forma:

$$T(s) = K\,D(-s)\,/\,D(s)\,. \tag{76}$$

Esse tipo de filtro é comumente aplicado:

i. Em células básicas em pedais de vibrato e de *phaser*.

ii. Na linearização da fase (minimização na distorção de imagens e de interferência intersimbólica).

iii. Na obtenção de atrasos da ordem de milissegundos na faixa de áudio, necessários, por exemplo, no alinhamento mecânico de alto-falantes.

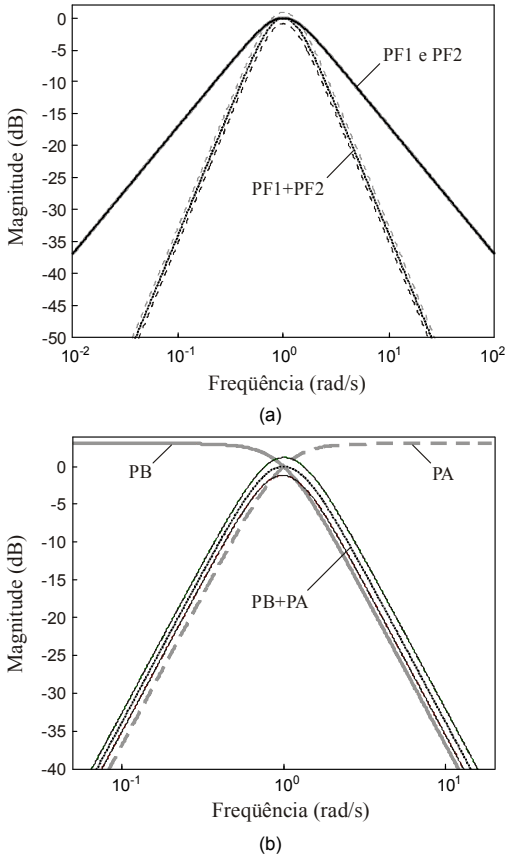


Figura 46 Magnitude da resposta das duas cascatas. (a) Rede PFPF. (b) Rede PBPA.

As estruturas da Figura 47 realizam filtros passa-tudo de primeira ordem, cujas funções de transferência são dadas por

$$T(s) = K\frac{-s+\sigma_0}{s+\sigma_0} \tag{77}$$

onde σ_0 $1/RC$ e K é o ganho em toda a faixa de freqüência. Para a rede da Figura 47(a), $K = -1$ e os desvios estocásticos da magnitude tendem a zero em freqüências acima do pólo. Para a rede da Figura 47(b), $K = 1$ e os desvios estatísticos da magnitude tendem a zero em freqüências abaixo do pólo. Os desvios estocásticos da fase são semelhantes.

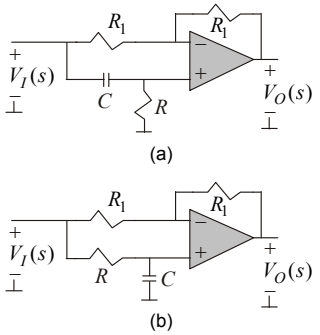


Figura 47 Filtros passa-tudo de primeira ordem. (a) Com resistor aterrado. (b) Com capacitor aterrado.

A estrutura da Figura 48 realiza um filtro passa-tudo de segunda ordem, cuja função de transferência é

$$T(s) = \frac{K(s^2-(\omega_0/Q)s+\omega_0^2)}{s^2+(\omega_0/Q)s+\omega_0^2} = \frac{K(s^2-a_1s+a_0)}{s^2+a_1s+a_0} \tag{78}$$

onde K é o ganho em toda a faixa de freqüências, ω_0 é o módulo dos pólos ou zeros (quando as singularidades são complexas) e Q é o fator de qualidade dessas singularidades.

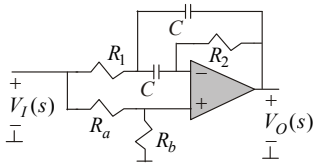


Figura 48 Filtro passa-tudo de segunda ordem.

Os valores de R_a e R_b do filtro passa-tudo da Figura 48 devem ser calculados de tal forma que o ganho resultante seja

$$K = \frac{a_0}{a_0+a_1^2} = \frac{R_b}{R_a+R_b} = \frac{Q^2}{Q^2+1}. \tag{79}$$

Os demais elementos, normalizados em impedância, podem ser encontrados em função dos coeficientes a_1 e a_0:

$$R_2 = R_a = b \tag{80}$$

$$R_1 = \frac{0{,}25ba_1^2}{a_0} \tag{81}$$

$$C = \frac{1}{b\sqrt{a_0R_1}} \tag{82}$$

$$R_b = \frac{0{,}25b}{R_1}. \tag{83}$$

O ganho K resultante é sempre menor do que 1. Por exemplo, para $Q = 10$, o ganho resultante é $K = 0{,}99$. Para $Q = 1$, o ganho resultante é $K = 0{,}5$.

A Figura 49 mostra que diferentes formas da fase podem ser obtidas em função do Q, em diferentes posições do eixo ω (em função de ω_0). Por isso, um dos usos dos filtros passa-tudo é colocá-los em cascata com sistemas que precisam ter suas fases corrigidas.

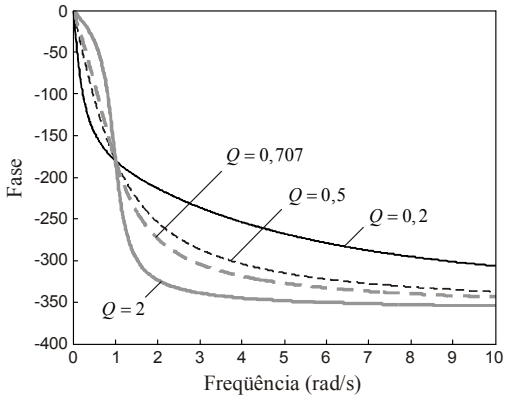


Figura 49 Fase em função da freqüência (em escala linear) de um filtro passa-tudo de segunda ordem.

As células básicas usadas nos circuitos que proporcionam o efeito de vibrato e de *phaser* analógicos são constituídas por filtros passa-tudo de primeira ordem [ver Figura 47(a)], em que o resistor aterrado é substituído

por um resistor cujo valor depende da tensão de saída de um LFO (foto-acoplador, JFET ou OTA). A Figura 50 ilustra a variação da fase da resposta em freqüência de um filtro passa-tudo de primeira ordem, normalizado em freqüência. Estruturas de segunda ordem não são usadas em pedais de efeitos porque a variação de apenas uma resistência (como no caso do passa-tudo de primeira ordem) provocaria o não atendimento da equação

$$K = a_0 / (a_0 + a_1^2) \tag{84}$$

e o circuito deixaria de ser um passa-tudo.

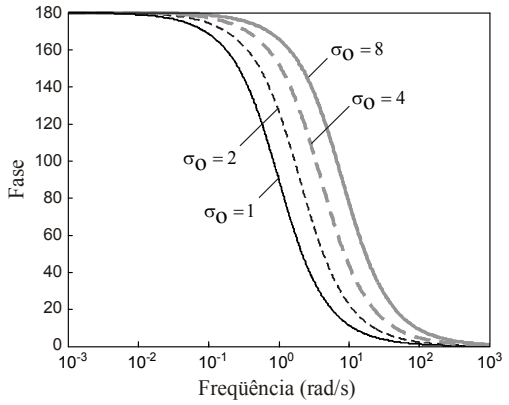


Figura 50 - Fase em função da freqüência do filtro passa-tudo da Figura 47 para vários valores de σ_0.

Exemplo prático 4: *A seguir, é discutido porque não é prático obter efeitos de eco (ou de reverberação) com sistemas analógicos contínuos.*

Para a obtenção de um atraso praticamente constante em uma faixa de freqüência limitada com um passa-tudo pode-se proceder da seguinte forma: escolhe-se uma função passa-baixa polinomial que apresente atraso de fase (ou de grupo) plano em uma faixa desejada; a seguir, constrói-se o polinômio do numerador fazendo

$$N_n(s) = D_n(-s)\ . \tag{85}$$

Assim, a função passa-tudo terá a seguinte função de transferência:

$$T_n(s) = \frac{KN_n(s)}{D_n(s)} = \frac{KD_n(-s)}{D_n(s)}\ . \tag{86}$$

Neste exemplo, vamos trabalhar com uma função Ulbrith-Piloty (UP) [5] de ordem 10, que apresenta atrasos de grupo e de fase aproximadamente planos (ripple de δ_{PICO} = 0,05 s *na função normalizada original). A razão de se optar por um filtro UP, em vez de, por exemplo, um filtro Bessel, é que com este último o mesmo atraso seria obtido somente até uma freqüência de aproximadamente* 6 kHz.

A função UP desnormalizada obtida é dada pela expressão abaixo, cujos coeficientes a_{1i} *e* a_{0i} *estão apresentados na Tabela 1.*

$$T_{10}(s) = \frac{K\,D_{10}(-s)}{D_{10}(s)} = \prod_{i=1}^{5} \frac{K(s^2 - a_{1i}s + a_{0i})}{s^2 + a_{1i}s + a_{0i}}\ . \tag{87}$$

Tabela 1 Coeficientes das cinco biquadráticas do filtro Ulbrith-Piloty passa-tudo de ordem 10

i	a_{1i}	$a_{0i}(\times 10^{-6})$
1	20347,7809	3853,3747
2	27413,5568	2579,08249
3	30079,1435	1489,27709
4	31232,0956	707,800515
5	31687,8925	303,037403

Com o filtro UP passa-baixa foi obtido um atraso

$$\tau(0)\ (@\,10\,\mathrm{kHz}) = 0{,}1848\,\mathrm{ms} \tag{88}$$

(ver Figura. 51) com uma queda de 200 dB/dec na banda de rejeição. Com o filtro UP AP foi obtido um atraso

$$\tau(0)\ (@\,10\,\mathrm{kHz}) = 0{,}3696\,\mathrm{ms} \tag{89}$$

o dobro do atraso obtido com o filtro passa-baixa, agora com a magnitude da resposta em freqüência plana.

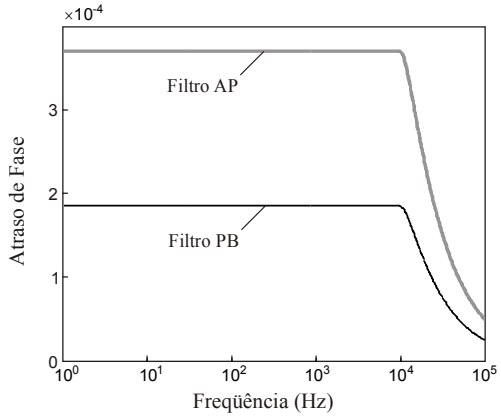


Figura 51 Atrasos dos filtros Ulbrith-Piloty, passa-baixa e passa-alta de ordem 10.

Usando filtros passa-tudo UP de ordem 10, para se obter um atraso de 50,2656 ms *em uma faixa de 10 kHz, seriam necessários 136 estágios iguais ao projetados. Então, a ordem total do sistema analógico seria igual a 1360. Esse exemplo mostra que não é prático obter os atrasos necessários em efeitos de eco (ou de reverberação) com sistemas analógicos contínuos.*

Exemplo prático 5: *As estruturas mais simples de filtros, redes com um único amp. op., apresentam uma desvantagem no que diz respeito à variação dos seus parâmetros. Normalmente, o ganho K, a freqüência* ω_0 *e o fator de qualidade Q das funções de transferência não podem ser independentemente variados quando se varia o valor de um único componente. Mas a diminuição do ganho pode ser obtida por meio de um divisor de tensão na entrada. Observe a Figura 52. Se a impedância vista pelo nó (i) for mantida constante, a diminuição do ganho pode ser feita independentemente dos outros parâmetros. Para isso, deve-se forçar a igualdade* $Z = Z' \| Z''$.

$$\frac{V_0'(s)}{V_I(s)} = \left(\frac{Z''}{Z' + Z''}\right)\frac{V_O(s)}{V_I(s)} \tag{90}$$

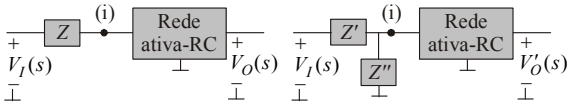


Figura 52 (a) Rede com impedância *Z* na entrada. (b) Diminuição do ganho sem alteração dos demais parâmetros.

Suponha que um ganho Y deve ser reduzido para um ganho X qualquer. Se o elemento na entrada é um resistor, então:

$$\frac{X}{Y} = \frac{R''}{R' + R''} \tag{91}$$

e

$$R = \frac{R'R''}{R' + R''}. \tag{92}$$

Assim, é possível mostrar que

$$R' = (Y / X)R \tag{93}$$

e

$$R'' = RR' / (R' - R)\ . \tag{94}$$

Se o elemento na entrada é um capacitor então:

$$\frac{X}{Y} = \frac{C'}{C' + C''} \tag{95}$$

e

$$C = C' + C''. \tag{96}$$

Assim, é possível mostrar que

$$C' = (X / Y)C \tag{97}$$

e

$$C'' = C - C'\ . \tag{98}$$

3 EQUALIZADORES DE ÁUDIO ANALÓGICOS

Equalizadores de áudio podem ser fixos ou variáveis. No caso dos fixos, os parâmetros de controle não são ajustáveis pelo usuário. Como exemplo, podem-se citar os equalizadores RIAA, usados tanto na gravação quanto na reprodução de discos de vinil. Nessa categoria, também se enquadram algumas próteses auditivas, cujo ajuste dos parâmetros é feito pelo fabricante após a realização de um exame áudio-métrico no paciente. Entretanto, são os equalizadores com parâmetros variáveis o foco principal deste tutorial, circuitos que permitem, dependendo do tipo de equalizador, o ajuste de parâmetros, tais como ganho/atenuação, freqüência central e largura de banda (fator de qualidade). Uma das aplicações mais importantes de equalizadores de áudio variáveis é a correção da resposta em freqüência de ambientes. Por exemplo, equalizadores de áudio são comumente empregados na eliminação de realimentações acústicas, popularmente conhecidas como microfonias. Além disso, os equalizadores de áudio variáveis são muito importantes para o ajuste da reprodução de um sistema de som ao gosto do ouvinte.

3.1 Tipos de Equalizadores

a) Equalizador Controle de Tonalidade (*Shelving*)

O equalizador mais simples possível é o equalizador controle de tonalidade, adequado para ajustar o nível de graves e agudos de um sistema de som. Esse tipo de equalizador é comumente usado no estágio de pré-amplificação de amplificadores de potência bem como na primeira e na última seção de equalizadores gráficos ou paramétricos, explicados a seguir, para o ajuste dos graves e agudos, respectivamente. Sendo assim, ele possui duas seções: uma para o ajuste de graves e outra para o de agudos. A Figura 53 apresenta exemplos de magnitude de resposta em freqüência de um equalizador controle de tonalidade para diferentes níveis de amplificação e atenuação, tanto para a seção ajuste de graves [Figura 53(a)] quanto para a seção ajuste de agudos [Figura 53(b)]. Tais respostas são planas nas baixas e nas altas freqüências, daí o nome *shelving* (prateleira).

b) Equalizador Gráfico

O equalizador gráfico [6] permite ajustar as faixas de freqüências fixas, cada uma podendo apresentar um ganho ou uma atenuação de valor K (em geral, expresso em decibéis). Cada seção de um equalizador gráfico possui uma função de transferência do tipo *bump*, que será apresentada em detalhes adiante. A Figura 54 apresenta um exemplo de resposta em freqüência de uma seção de um equalizador gráfico, com freqüência central $f_0 = 1$ kHz e fator de qualidade $Q = 4{,}318$ (largura de banda) [1]. O usuário não tem controle sobre esses parâmetros, que são fixos. Ele simplesmente pode ajustar o ganho que deseja na freqüência f_0, que, em geral, pode variar na faixa de −12 a +12 dB. O nome equalizador gráfico deve-se à disposição dos controles de ganho (quando verticais), os quais representam de forma aproximada a curva de resposta em freqüência do equalizador.

Muitos equalizadores gráficos utilizam o padrão ISO de freqüências centrais em que o espectro é dividido em freqüências relacionadas por um fator multiplicativo k. Por exemplo, para um equalizador de 31 bandas, comumente usado em aplicações profissionais, também chamado de 1/3 de oitava, a cada oitava deve haver três freqüências centrais. Assim, se entre uma determinada freqüência f_0 e uma freqüência uma oitava acima ($2f_0$) devem existir três freqüências centrais, pode-se deduzir que

$$k^3 f_0 = 2 f_0 \ \Rightarrow\ k = 2^{1/3} \cong 1{,}2599. \tag{99}$$

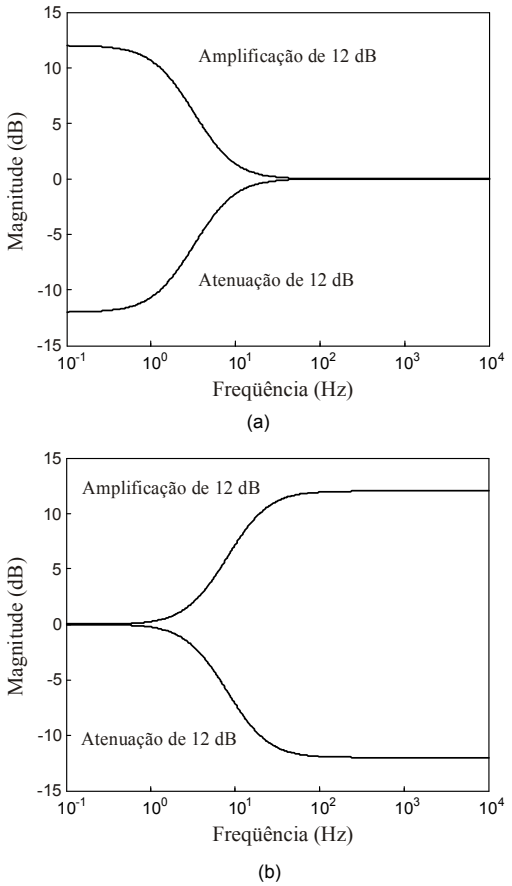


Figura 53 Exemplos de respostas em freqüência de um equalizador controle de tonalidade. (a) Seção controle de graves. (b) Seção controle de agudos.

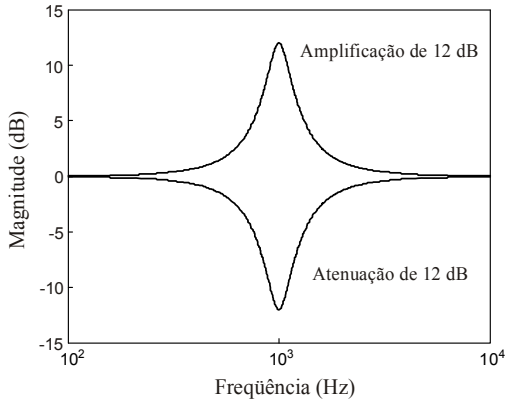


Figura 54 Exemplos de respostas em freqüência de uma seção *bump* de um equalizador gráfico, ambas com freqüência central de 1 kHz e fator de qualidade 4,318.

A Tabela 2 apresenta o conjunto de freqüências centrais f_0 adotado para um equalizador de 31 bandas, considerando 1000 Hz como freqüência de referência e $k = 2^{1/3}$.

Tabela 2 Freqüências centrais de um equalizador gráfico de 31 bandas, adotando $k = 2^{1/3}$ e 1000 Hz como freqüência de referência

Seção	Freqüência central f_0	Seção	Freqüência central f_0
1	$1000k^{-17} = 19{,}686$ Hz	17	$1000k^{-1} = 793{,}7005$ Hz
2	$1000k^{-16} = 24{,}803$ Hz	18	1000 Hz
3	$1000k^{-15} = 31{,}25$ Hz	19	$1000k = 1259{,}9$ Hz
4	$1000k^{-14} = 39{,}3725$ Hz	20	$1000k^{2} = 1587{,}40$ Hz
5	$1000k^{-13} = 49{,}6063$ Hz	21	$1000k^{3} = 2000$ Hz
6	$1000k^{-12} = 62{,}50$ Hz	22	$1000k^{4} = 2519{,}80$ Hz
7	$1000k^{-11} = 78{,}7451$ Hz	23	$1000k^{5} = 3174{,}80$ Hz
8	$1000k^{-10} = 99{,}2126$ Hz	24	$1000k^{6} = 4000$ Hz
9	$1000k^{-9} = 125$ Hz	25	$1000k^{7} = 5039{,}70$ Hz
10	$1000k^{-8} = 157{,}4901$ Hz	26	$1000k^{8} = 6349{,}60$ Hz
11	$1000k^{-7} = 198{,}4251$ Hz	27	$1000k^{9} = 8000$ Hz
12	$1000k^{-6} = 250$ Hz	28	$1000k^{10} = 10{,}079$ kHz
13	$1000k^{-5} = 314{,}9803$ Hz	29	$1000k^{11} = 12{,}699$ kHz
14	$1000k^{-4} = 396{,}8503$ Hz	30	$1000k^{12} = 16$ kHz
15	$1000k^{-3} = 500$ Hz	31	$1000k^{13} = 20{,}158$ kHz
16	$1000k^{-2} = 629{,}9605$ Hz		

Como já mencionado, além da freqüência central, cada seção de um equalizador gráfico deve possuir um determinado fator de qualidade Q [4]. A fim de deduzi-lo, considere a Figura 55, que demonstra parte do espectro de freqüências de um equalizador gráfico qualquer. A freqüência central f_0 está relacionada com as freqüências de corte f_i e f_s (freqüências onde o ganho decai 3 dB) por

$$f_0^2 = f_i f_s = f_x k f_x = k f_x^2. \tag{100}$$

De acordo com a Figura 55, $f_i = f_x$ e $f_s = kf_x$. Logo,

$$f_0^2 = f_x k f_x = k f_x^2 \tag{101}$$

de onde se determina que

$$f_0 = \sqrt{k}\, f_x\,. \tag{102}$$

Além disso, a largura de banda, em hertz, é

$$B = k f_x - f_x = (k-1) f_x\,. \tag{103}$$

Por definição, o fator de qualidade é

$$Q = f_0 / B\,. \tag{104}$$

Substituindo (102) e (103) em (104), pode-se deduzir:

$$Q = \frac{\sqrt{k}}{k-1}\,. \tag{105}$$

Logo, para o equalizador de 1/3 de oitava, para o qual $k = 2^{1/3}$, o fator de qualidade de cada seção é $Q \approx 4{,}318$.

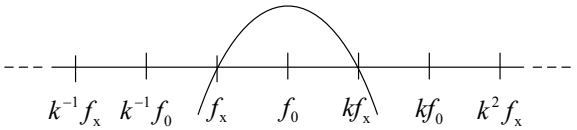


Figura 55 Espectro de freqüências de um equalizador gráfico qualquer, usado para determinar o fator de qualidade Q.

c) Equalizador Paramétrico

Este tipo de equalizador também possui função de transferência do tipo *bump*. Entretanto, diferentemente do equalizador gráfico, o equalizador paramétrico possui, além do controle sobre o ganho em cada seção, controle sobre o valor da freqüência central f_0 e do fator de qualidade Q de cada seção. Esses equalizadores possuem grande precisão, sobretudo para correção da resposta em freqüência do ambiente, quando se deseja corrigir a resposta de uma determinada faixa de freqüência. O nome (paramétrico) deve-se justamente ao fato deste permitir o ajuste de todos os parâmetros de equalização. A Figura 56 apresenta exemplos de controle da freqüência central f_0 [Figura 56(a)] e do fator de qualidade Q [Figura 56(b)].

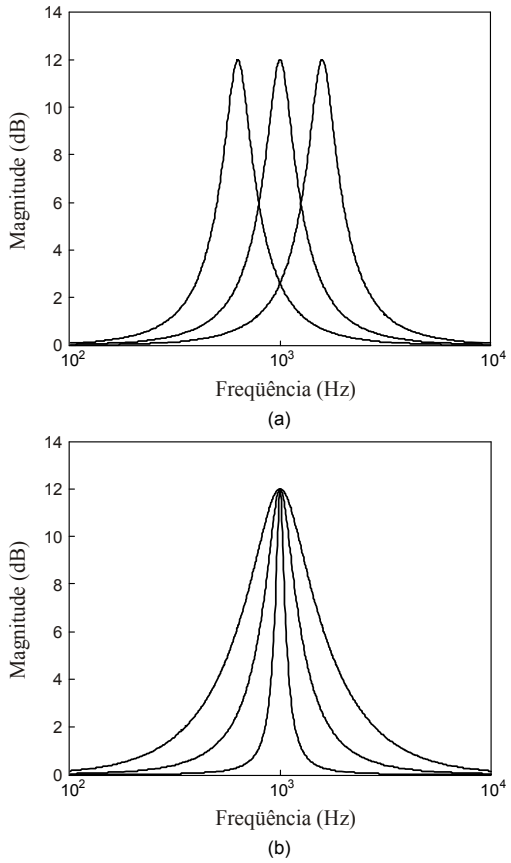


Figura 56 Exemplos de controle de um equalizador paramétrico. (a) Controle do valor da freqüência central f_0. (b) Controle do fator de qualidade Q.

Têm-se também os equalizadores semi-parametricos onde o fator de qualidade Q é pré-fixado (não ajustável), porém pode-se variar a freqüência central e o ganho da célula. Esse tipo de equalizador é muito comum em mesas de som.

d) Equalizador Paragráfico

Como o próprio nome sugere, este equalizador é uma junção dos equalizadores gráfico e paramétrico. Do equalizador gráfico, possui as freqüências centrais no padrão ISO (ou outras freqüências pré-determinadas). Do equalizador paramétrico, possui o ajuste da freqüência central e pode ou não possuir ajuste do fator de qualidade Q (largura de banda). Resumindo, trata-se de um equalizador gráfico com freqüências e fatores de qualidade variáveis. Devido ao seu maior custo, conseqüência natural de sua maior flexibilidade de ajustes, não se oferece comercialmente equalizadores paragráficos de 1/3 oitava. Porém, pode-se encontrar equalizadores paragráficos de 1 oitava.

3.2 Funções de Transferência

a) Funções de Transferência Controle de Tonalidade (*Shelving*)

Como já mencionado, um equalizador controle de tonalidade possui duas seções: uma para o controle de graves e outra, para o de agudos. Assim, são necessárias duas funções de transferência, uma para cada seção.

A função de transferência da seção de controle de graves é

$$T_{\mathrm{CG}}(s) = \frac{s+a}{s+b}. \tag{106}$$

Nesse caso, o ganho em altas freqüências ($s \to \infty$) é unitário (0 dB), mas, em baixas freqüências ($s \to 0$), ele é dado por $20\log(a/b)$. Assim, se $a > b$, em baixas freqüências tem-se uma amplificação (*boost*). Se $a < b$, em baixas freqüências tem-se uma atenuação (*cut*). A Figura 53(a) ilustra essas duas situações, para os casos de $a = 39{,}81$ rad/s e $b = 10$ rad/s (amplificação de 12 dB em baixas freqüências) e para $a = 10$ rad/s e $b = 39{,}81$ rad/s (atenuação de 12 dB em baixas freqüências).

No caso da seção controle de agudos, a função de transferência é dada por

$$T_{\mathrm{CA}}(s) = \frac{cs+1}{ds+1} = \frac{c}{d} \times \frac{s+(1/c)}{s+(1/d)}. \tag{107}$$

Nesse caso, o ganho em baixas freqüências ($s \to 0$) é unitário (0 dB), mas, em altas freqüências ($s \to \infty$), ele é dado por $20\log(c/d)$. Assim, para $c > d$, em altas freqüências tem-se uma amplificação (*boost*). Para $c > d$, em altas freqüências tem-se uma atenuação (*cut*). A Figura 53(b) ilustra essas duas situações, para os casos de $c = 0{,}03981$ rad/s e $d = 0{,}01$ rad/s e para $c = 0{,}01$ rad/s e $d = 0{,}03981$ rad/s.

b) Função de Transferência *Bump*

Uma função de transferência do tipo *bump* é dada por

$$T_{\mathrm{bump}}(s) = \frac{s^2 + (\omega_0 / Q_{\mathrm{Z}})s + \omega_0^2}{s^2 + (\omega_0 / Q_{\mathrm{P}})s + \omega_0^2}. \tag{108}$$

onde ω_0 é o módulo dos pólos e dos zeros da função, Q_{Z} é o fator de qualidade dos zeros da função e Q_{P} é o fator de qualidade dos pólos da função. Substituindo s por jω, obtém-se a resposta em freqüência da função *bump*, cuja magnitude é

$$\left|T_{\mathrm{bump}}(\mathrm{j}\omega)\right|_{\mathrm{dB}} = 20\log\left(\sqrt{\frac{\omega_0^2 - \omega^2 + \omega^2\left(\omega_0 / Q_{\mathrm{Z}}\right)^2}{\omega_0^2 - \omega^2 + \omega^2\left(\omega_0 / Q_{\mathrm{P}}\right)^2}}\right). \tag{109}$$

Assim, podem-se observar três situações:

- Em baixa freqüência ($\omega \to 0$), o ganho tende a 0 dB.
- Em alta freqüência ($\omega \to \infty$), o ganho tende a 0 dB.
- Na freqüência ω_0, o ganho é $20\log(Q_{\mathrm{P}}/Q_{\mathrm{Z}})$.

Logo, na freqüência ω_0, há três possíveis situações:

- Se $Q_{\mathrm{P}} > Q_{\mathrm{Z}}$, obtém-se um ganho (*boost*).
- Se $Q_{\mathrm{P}} = Q_{\mathrm{Z}}$, o ganho é nulo (resposta *flat*).
- Se $Q_{\mathrm{P}} < Q_{\mathrm{Z}}$, obtém-se uma atenuação (*cut*).

A Figura 54 ilustra a resposta em freqüência de uma função de transferência *bump*, com $f_0 = 1$ kHz ($\omega_0 = 2\pi f_0$) e $Q = 4{,}318$, tanto para uma amplificação de 12 dB ($Q_{\mathrm{P}}/Q_{\mathrm{Z}} = 3{,}981$) quanto para uma atenuação de 12 dB ($Q_{\mathrm{P}}/Q_{\mathrm{Z}} = 0{,}251$).

3.3 Obtenção das Funções de Transferência Controle de Tonalidade e *Bump*

Um esquema possível, muito usado por fabricantes, para a implementação de equalizadores de áudio é mostrado na Figura 57. Dependendo do tipo de bloco de filtro seletor utilizado, $T_{\mathrm{FS}}(s)$, obtém-se as funções de transferência controle de tonalidade (controle de graves ou agudos) ou *bump*. Equalizadores de várias bandas podem ser obtidos com esse mesmo esquema, bastando para isso colocar quantos blocos internos de filtros seletores quanto forem o número de bandas desejadas, juntamente com os seus respectivos ganhos K.

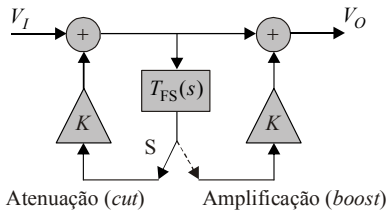


Figura 57 Diagrama de blocos para a implementação de funções *shelving* e *bump*.

Através da Figura 57, pode-se deduzir que com a chave S na posição amplificação, obtém-se a função de transferência

$$T(s) = \frac{V_{\mathrm{O}}(s)}{V_{\mathrm{I}}(s)} = 1 + KT_{\mathrm{FS}}(s). \tag{110}$$

Já com a chave na posição atenuação, implementa-se a seguinte função

$$T(s) = \frac{V_{\mathrm{O}}(s)}{V_{\mathrm{I}}(s)} = \frac{1}{1 + KT_{\mathrm{FS}}(s)}. \tag{111}$$

a) Caso 1: Função $T_{\mathrm{FS}}(s)$ do tipo passa-baixa

Considere o uso de uma função passa-baixa no diagrama da Figura 57, do tipo

$$T_{\mathrm{PB}}(s) = \frac{\sigma_0}{s + \sigma_0}. \tag{112}$$

Se a chave S estiver na posição amplificação e substituindo (14) em (12), implementa-se a seguinte função de transferência

$$T(s) = \frac{s + (1+K)\sigma_0}{s + \sigma_0} = \frac{s+a}{s+b}. \tag{113}$$

Ou seja, está se implementando uma função controle de graves com $a > b$ (amplificação). Já com a chave na posição atenuação, implementa-se

$$T(s) = \frac{s + \sigma_0}{s + (1+K)\sigma_0} = \frac{s+a}{s+b}. \quad (114)$$

Nesse caso, implementa-se a função controle de graves com $a < b$ (atenuação).

Uma possível rede que implementa a função $T_{PB}(s)$ é mostrada na Figura 58. A sua função de transferência é dada por

$$T_{PB}(s) = \frac{1/RC}{s + 1/RC} = \frac{\sigma_0}{s+\sigma_0}. \quad (115)$$

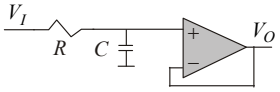


Figura 58 Filtro passa-baixa usado para implementar uma seção controle de graves.

b) Caso 2: Função $T_{FS}(s)$ do tipo passa-alta

Considere o uso de uma função passa-alta no diagrama da Figura 57, do tipo

$$T_{PA}(s) = \frac{s}{s+\sigma_0}. \quad (116)$$

Se a chave S estiver na posição amplificação e substituindo (18) em (12), implementa-se a seguinte função de transferência

$$T(s) = \frac{\left[(1+K)/\sigma_0\right]s + 1}{(1/\sigma_0)s + 1} = \frac{cs+1}{ds+1}. \quad (117)$$

Ou seja, está se implementando uma função controle de agudos com $c > d$ (amplificação). Já com a chave na posição atenuação, implementa-se

$$T(s) = \frac{(1/\sigma_0)s + 1}{\left[(1+K)/\sigma_0\right]s + 1} = \frac{cs+1}{ds+1}. \quad (118)$$

Nesse caso, implementa-se a função controle de graves com $c < d$ (atenuação).

Uma possível rede que implementa a função $T_{PA}(s)$ é mostrada na Figura 59. A sua função de transferência é dada por

$$T_{PA}(s) = \frac{s}{s + 1/RC} = \frac{s}{s+\sigma_0}. \quad (119)$$

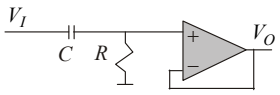


Figura 59 Filtro passa-alta usado para implementar uma seção controle de agudos.

c) Caso 3: Função $T_{FS}(s)$ do tipo passa-faixa

Considere o uso de uma função passa-faixa no diagrama da Figura 57 [7], do tipo

$$T_{PF}(s) = \frac{(\omega_0/Q)s}{s^2 + (\omega_0/Q)s + \omega_0^2}. \quad (120)$$

Se a chave S estiver na posição amplificação e substituindo (22) em (12), implementa-se a seguinte função de transferência

$$T(s) = \frac{s^2 + (1+k)(\omega_0/Q)s + \omega_0^2}{s^2 + (\omega_0/Q)s + \omega_0^2} = \frac{s^2 + (\omega_0/Q_Z)s + \omega_0^2}{s^2 + (\omega_0/Q_P)s + \omega_0^2}. \quad (121)$$

Ou seja, está se implementando uma função *bump* com $Q_P > Q_Z$ (amplificação). Já com a chave na posição atenuação, obtém-se

$$T(s) = \frac{s^2 + (\omega_0/Q)s + \omega_0^2}{s^2 + (1+k)(\omega_0/Q)s + \omega_0^2} = \frac{s^2 + (\omega_0/Q_Z)s + \omega_0^2}{s^2 + (\omega_0/Q_P)s + \omega_0^2}. \quad (122)$$

Nesse caso, implementa-se a função *bump* com $Q_P < Q_Z$ (atenuação).

Uma possível rede que implementa a função $T_{PF}(s)$ é mostrada na Figura 60, uma rede *multiple feedback* (MFB). A sua função de transferência é dada por

$$T_{PF}(s) = G\frac{(\omega_0/Q)s}{s^2 + (\omega_0/Q)s + \omega_0^2}. \quad (123)$$

onde

$$G = -\frac{R_2}{2R_1} \quad (124)$$

$$\omega_0 = \sqrt{\frac{1}{R_1 R_2 C^2}} \quad (125)$$

$$Q = \frac{\sqrt{R_1 R_2}}{2R_1} \quad (126)$$

com $R_1 = R_{1S} // R_{1P}$. O projeto dessa rede pode ser visto no item 2.4 deste tutorial.

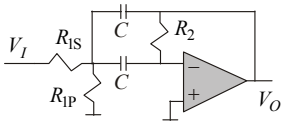


Figura 60 Filtro passa-faixa usado para implementar uma seção *bump*.

3.4 Diagramas de Equalizadores de Áudio

Uma maneira de implementar cada seção da Figura 57 é usar dois somadores inversores e um potenciômetro com tap central aterrado (*grounded center tap*), estrutura mostrada na Figura 61 para o caso particular de uma seção *bump*, na qual se utiliza um filtro passa-faixa. Note que foi colocado um inversor em cascata com o filtro passa-faixa, pois o filtro aqui sugerido (filtro MFB da Figura 60) é inversor. Caso se utilize uma estrutura não-inversora, tal como a Sallen-Key [4], o bloco inversor não deve ser utilizado.

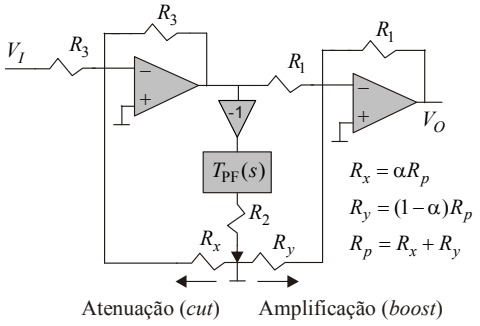


Figura 61 Implementação de uma seção de um equalizador.

Para implementar equalizadores de várias seções, basta acrescentar outros blocos $T_{FS}(s)$ em paralelo de acordo com a necessidade, com todas as entradas ligadas juntas e com cada saída ligada a um potenciômetro com *tap* central aterrado, cujos terminais devem ser conectados aos somadores. Como exemplo, a Figura 62 apresenta o diagrama de blocos de um equalizador composto por *n* seções *bump*.

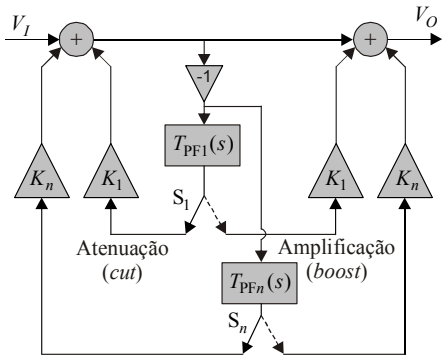


Figura 62 Diagrama de blocos de um equalizador de *n* bandas.

Exemplo prático 6: *Neste exemplo, é mostrada a comparação da magnitude da resposta em freqüência quando se tenta obter um reforço de graves e agudos com um equalizador gráfico e com dois equalizadores shelving. A Figura 63 mostra o caso de um reforço* +12 dB. *Os controles de amplitude são feitos de tal forma que seja obtida uma queda de* 3 dB *em* 71,13 Hz *e* 7 kHz. *Pode ser observado que quando se usa o equalizador gráfico, ondulações indesejadas ocorrem em baixas e altas freqüências.*

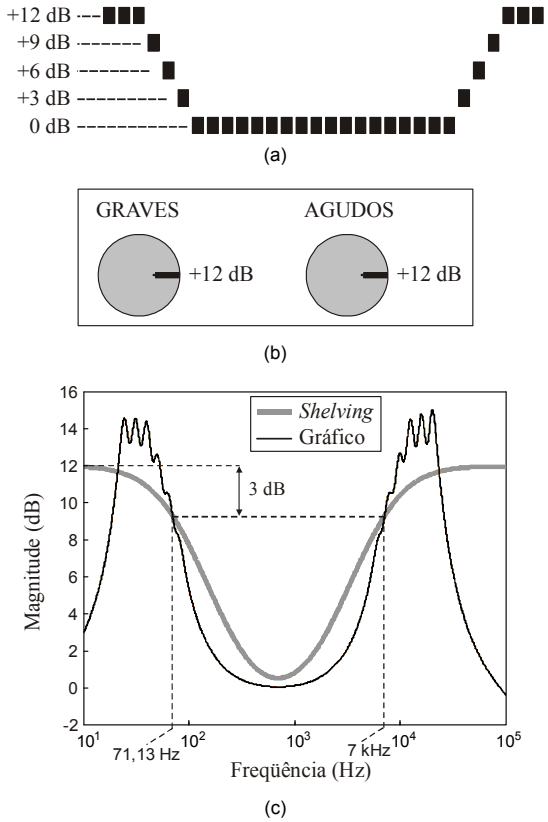


Figura 63 Comparação entre duas formas de se obter reforço de graves e agudos. (a) Posição dos potenciômetros *sliders* de um equalizador gráfico. (b) Posição *knobs* dos equalizadores shelving. (c) Magnitude da resposta em freqüência nos dois casos.

4 PROCESSADOR DIGITAL DE SINAIS (DSP)

Basicamente, DSP é um processador com uma arquitetura interna customizada para a execução das operações usuais de processamento digital de sinais (PDS), tais como: soma, atrasos, multiplicação com acumulação, etc. Normalmente, a arquitetura do DSP é composta por: unidades computacionais (ALU, SHIFTER, MAC), de geração de endereços, de controle de programa, memória de dados, de programas, diversos barramentos, amplo conjunto de periféricos, etc. Essa variedade de recursos diferencia o DSP dos processadores de uso geral bem como possibilita ao DSP atender aos requisitos computacionais de desempenho dos algoritmos de PDS. A Figura 64 apresenta os principais blocos que compõem o DSP.

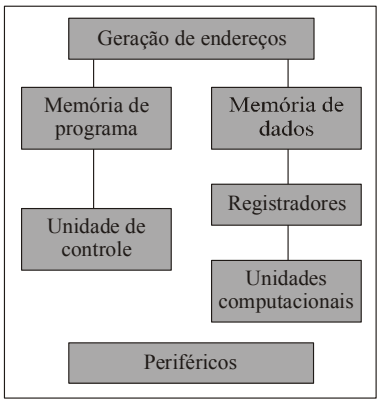


Figura 64 Principais blocos do processador DSP.

A ANALOG DEVICES (AD) [8] é um dos principais fabricantes de processadores DSP e possui diferentes opções de dispositivos de ponto fixo e ponto flutuante. Esse tutorial utiliza como dispositivo alvo o processador DSP BF-533 de ponto fixo (16bits) da família Blackfin da AD [9].

Para facilitar a implementação e testes de novas aplicações, a AD disponibiliza um ambiente de desenvolvimentos integrado (IDE) e kits de desenvolvimento. O IDE da AD é denominado Visualdsp [10] e os kits de desenvolvimento de EZ-KIT Lite [11]. Deve-se ressaltar que o VisualDSP permite a implementação de algoritmos utilizando as linguagens de programação "C" [12] e "Assembly" [13]. Para exemplificar a implementação de algoritmos usando tais linguagens, é apresentado, na Tabela 3, um trecho de código cuja função executada é equivalente.

Tabela 3 – Trecho de código equivalente em "C" e "Asm"

Código em "C"	Código em "Asm"

for (j=1;j<10;j++)
{
Acc=L_mac(Acc,*pa++,
*pi--);
}
	a0=0||r1.l=w[i1++]||r2.l=w[i0--];
	Loop loop_3 lc1 = P1;
	Loop_begin loop_3;
	a0+=r1.l*r2.l||r1.l=
	w[i1++]||r2.l=w[i0--];
	Loop_end loop_3;

Com relação ao EZ-KIT Lite, suas principais características são [11]:

- Processador DSP BF 533 (756 MHz),
- Memória SDRAM (64 MB),
- Memória flash (2 MB),
- 4 entradas e 6 saídas (96KHz/48KHz),
- Leds,
- Push buttons.

A Figura 65 mostra o diagrama de blocos do EZ-KIT Lite com o conversor analógico digital (CAD), o conversor digital analógico (CDA) e o DSP BF-533.

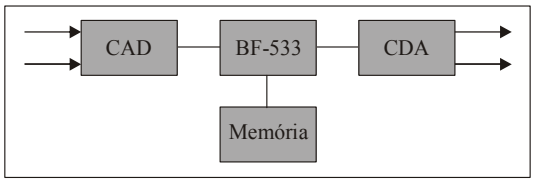


Figura 65 Diagrama de blocos do EZ-KIT Lite.

Na seqüência, são apresentados alguns exemplos de algoritmos que serão posteriormente implementados e executados em tempo real no EZ-KIT Lite.

5 EFEITOS DIGITAIS – DIAGRAMA DE BLOCOS

5.1 Introdução

O *delay* (atraso) é o bloco básico usado na obtenção de diversos efeitos como, por exemplo, vibrato, *flanger*, *chorus*, *reverb*, eco, etc.

O *delay* atrasa um sinal digital de um número de amostras desejadas. Dependendo de qual sinal é atrasado, obtém-se um *Comb filter* FIR (sinal de entrada) ou *Comb filter* IIR (sinal de saída). A equação de um *Comb filter* FIR [14] é dada por:

$$y(n) = x(n) + ax(n-D) \tag{127}$$

onde $y(n)$ é a amostra de saída, $x(n)$ é a amostra de entrada, a é o fator de ponderação, D é o valor do atraso em amostras e $x(n-D)$ é a amostra de entrada atrasada de D amostras.

O diagrama de blocos do *Comb filter* FIR é mostrado na Figura 66 e sua correspondente resposta em freqüência é apresentada na Figura 67.

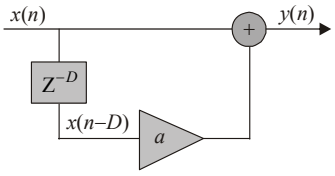


Figura 66 Diagrama de blocos do *Comb filter* FIR.

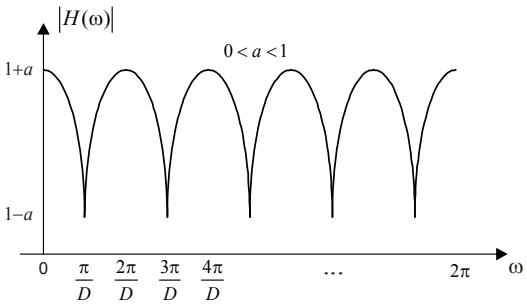


Figura 67 Resposta em freqüência do *Comb filter* FIR [14].

A equação recursiva de um *Comb filter* IIR é dada por [14]:

$$y(n) = x(n) + ay(n-D) \tag{128}$$

onde $y(n-D)$ é a amostra de saída atrasada de D amostras.

O diagrama de blocos do *Comb filter* IIR é apresentado na Figura 68 e sua correspondente resposta em freqüência na Figura 69.

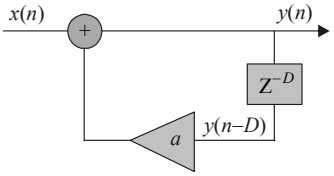


Figura 68 Diagrama de blocos do *Comb filter* IIR.

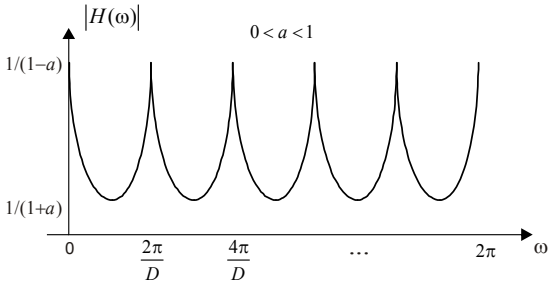


Figura 69 Resposta em freqüência do *Comb filter* IIR [14].

5.2 Vibrato

O valor escolhido para o atraso D em amostras (de forma equivalente, em segundos) altera o som percebido bem como o tipo de efeito de áudio. Por exemplo, o efeito vibrato é implementado na forma digital variando o atraso D continuamente no tempo [14]-[16]. O diagrama de blocos que representa tal implementação é mostrado na Figura 70.

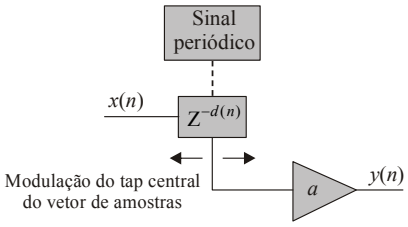


Figura 70 Diagrama de blocos do vibrato.

A equação correspondente do vibrato é dada por [14]:

$$y(n) = ax(n-d(n)) \tag{129}$$

onde $y(n)$ é a amostra de saída, $d(n)$ é o atraso variante no tempo, a é o fator de ponderação e $x(n-d(n))$ é a amostra de entrada atrasada de $d(n)$ amostras.

O atraso variante no tempo ($d(n)$) é obtido a partir de um oscilador de baixa freqüência (LFO) que gera um sinal periódico, dado por [14]:

$$d(n) = \frac{D}{2}[1-\cos(2\pi.n.f_{cy})] \tag{130}$$

onde D é o valor máximo do atraso em amostras e f_{cy} é a freqüência do LFO.

Na prática, o valor máximo utilizado para o atraso está entre 0 ms e 3 ms. Considerando, por exemplo, que a freqüência de amostragem é 8kHz, obtém-se 24 amostras para o máximo atraso D. Efeitos interessantes são obtidos quando a freqüência do LFO é menor que 5Hz.

Para a implementação em DSP do algoritmo do vibrato, deve-se alocar um vetor de tamanho D para armazenar as amostras de entrada bem como calcular o atraso $d(n)$. Normalmente, existem duas opções para se obter esse atraso. A primeira considera uma tabela com os valores anteriormente armazenados. Na segunda opção, é utilizada

uma equação recursiva para calcular o sinal periódico. Normalmente, devido à baixa freqüência do LFO (menor que 5Hz), há uma relação de compromisso entre complexidade computacional e utilização de recursos de memória, Na primeira, tem-se uma menor complexidade computacional com maior recurso de memória, enquanto na segunda opção tem-se uma maior complexidade computacional com menor recurso de memória [17].

Outra consideração relevante é sobre o valor obtido para o atraso $d(n)$, o qual pode ser fracionário. Pode-se perceber que tal situação resulta em um problema, dado que $d(n)$ corresponde a um índice do vetor de amostras. Normalmente, a solução usada em tal situação é considerar a interpolação linear entre duas amostras nos instantes de tempo $M+1$ e M, conforme apresentado por [18]:

$$y(n) = x(n-[M+1])\,frac + x(n-M)(1-frac) \qquad (131)$$

onde $frac$ corresponde ao valor fracionário do atraso.

5.3 Flanger/Chorus

Os efeitos *flanger*/*chorus* podem ser implementados de forma equivalente ao *Comb filter* FIR, ou seja, adicionando o sinal de entrada com uma réplica atrasada no tempo, conforme diagrama de blocos apresentado na Figura 71. Equivalente ao vibrato, a réplica é obtida de um atraso variante no tempo ($d(n)$), que é controlado por um LFO. Normalmente, no *flanger*, o atraso é menor que 10ms e a freqüência do LFO é menor ou igual a 2Hz. Já, para o *chorus*, o atraso deve ficar entre 10ms e 30ms e a freqüência do LFO menor ou igual a 1Hz [18-19].

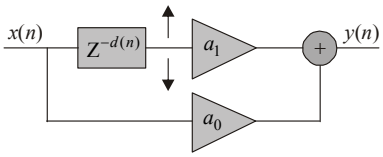


Figura 71 Diagrama de blocos *flanger*/*chorus*.

A equação geral para os efeitos *flanger*/*chorus* é dada por [14]:

$$y(n) = a_0 x(n) + a_1 x(n-d(n)) \qquad (132)$$

onde $y(n)$ é a amostra de saída, $d(n)$ é o atraso variante no tempo, a_0 e a_1 são os fatores de ponderação e $x(n-d(n))$ é a amostra de entrada atrasada de $d(n)$ amostras.

A correspondente resposta em freqüência para $a_0 = 1$ e $0 < a_1 < 1$ é mostrada na Figura 72. Pode-se observar em tal figura que a variação de $d(n)$ causa a atenuação de diferentes componentes de freqüência.

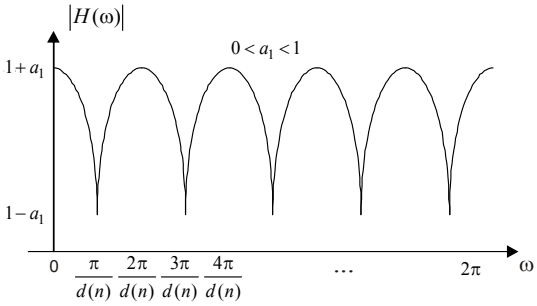


Figura 72 Resposta em freqüência do *flanger*/*chorus*.

Uma outra estratégia de implementação do *flanger*/*chorus* foi proposta por Dattorro [19]. A Figura 73 mostra o diagrama de blocos de tal estratégia.

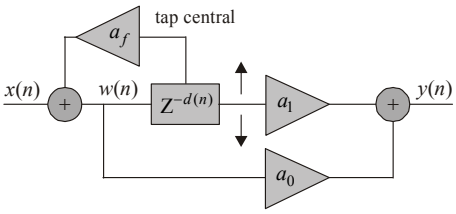


Figura 73 Diagrama de blocos do *flanger*/*chorus* proposto em [19].

As equações que correspondem a implementação da Figura 73 são dadas por:

$$w(n) = x(n) - a_f w(n - \frac{D}{2}) \qquad (133)$$

$$y(n) = a_0 w(n) + a_1 w(n-d(n)). \qquad (134)$$

Deve-se observar que o diagrama apresentado na Figura 73 permite representar os efeitos vibrato, *flanger* e *chorus*. Para cada efeito, deve-se escolher os valores adequados para os coeficientes, tamanho do atraso, freqüência e tipo do sinal usado no LFO, conforme apresentado na Tabela 4 [19].

Tabela 4 Representação dos efeitos vibrato, *flanger* e *chorus* usando a estrutura proposta por Dattorro

Efeito	Coeficientes	Atraso D (ms)	LFO (Hz)	Tipo de sinal usado no LFO
Vibrato	$a_f = 0$ $a_0 = 0$ $a_1 = 1$	0 – 3	0.1 até 5Hz	Senoidal
Flanger	$a_f = 0.7$ $a_0 = 0.7$ $a_1 = 0.7$	D < 10	<2Hz	Senoidal
Chorus	$a_f = -0.7$ $a_0 = 0.7$ $a_1 = 1$	10<D<30	< 1Hz	Senoidal Ruído filtrado banda estreita

5.4 Tremolo

O tremolo é implementado digitalmente multiplicando o sinal de entrada pela saída do LFO. Tal operação equivale a modular em amplitude a entrada, conforme diagrama em blocos apresentado na Figura 74.

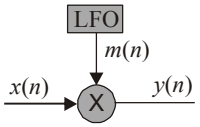


Figura 74 Diagrama em blocos do tremolo.

Na prática, a freqüência de saída do LFO é ajustada entre 1Hz e 10 Hz. A equação para o cálculo de tal freqüência é dada por [17]:

$$m(n) = \sin(2\pi . n . f_{cy}) \qquad (135)$$

onde $m(n)$ é o sinal de modulação e f_{cy} é a freqüência do sinal de modulação.

O sinal na saída do tremolo é dado por:

$$y(n) = m(n)x(n) \qquad (136)$$

onde $x(n)$ é o sinal de entrada e $y(n)$ é o sinal de saída.

5.5 Eco

O eco pode ser implementado digitalmente a partir do *Comb filter* FIR com a especificação do valor desejado do atraso D. Por exemplo, considerando que a freqüência de amostragem seja 8kHz e que se deseja um atraso máximo de 0,1segundos, deve-se adotar o valor 800 para o atraso D e alocar um tamanho correspondente para o vetor de atrasos $x(n-D)$.

A equação para a implementação do eco é dada por [14]:

$$y(n) = x(n) + ax(n-D) \tag{137}$$

onde $y(n)$ é a amostra de saída, $x(n)$ é a amostra de entrada, a é o fator de ponderação, D é o valor do atraso em amostras e $x(n-D)$ é a amostra de entrada atrasada de D amostras.

O diagrama de blocos para a implementação do eco é mostrado na Figura 75.

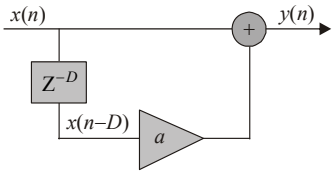


Figura 75 Diagrama de blocos do eco.

5.6 Reverberação

Conforme apresentado na Seção 1.6, o efeito de reverberação ou *reverb* simula a reflexão do som em um ambiente fechado, produzindo repetições do som original. O som produzido na reverberação é classificado em três componentes: som direto, reflexões iniciais (*early*) e reflexões finais (*late*). As reflexões iniciais ocorrem de 10ms a 100ms após o som direto. Já as reflexões finais são maiores que 100ms [14]. A Figura 76 mostra os componentes da reverberação.

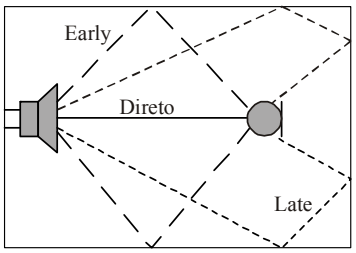


Figura 76 Componentes da reverberação.

A implementação digital da reverberação deve, necessariamente, considerar os três componentes citados, de forma a produzir um efeito mais próximo do "real". Os algoritmos mais utilizados para a implementação da reverberação foram propostos por Schoreder [20] e Moorer [21], consistindo em uma seqüência de *Comb filter* IIR (C1 a C4) e filtros passa-tudo (A1 e A2), conforme apresentado no diagrama de blocos da Figura 77.

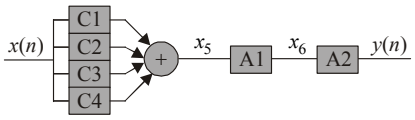


Figura 77 Diagrama de blocos da reverberação.

As equações dos *Comb filter* IIR são dadas por:

$$y_{C1}(n) = x(n) + a_1 y(n-D_1) \tag{138}$$

$$y_{C2}(n) = x(n) + a_2 y(n-D_2) \tag{139}$$

$$y_{C3}(n) = x(n) + a_3 y(n-D_3) \tag{140}$$

$$y_{C4}(n) = x(n) + a_4 y(n-D_4). \tag{141}$$

Já as equações correspondentes aos filtros passa-tudo são dadas por:

$$y_{A1}(n) = a_5 y(n-D_5) - a_5 x_5(n) + x_5(n-D_5) \tag{142}$$

$$y_{A2}(n) = a_6 y(n-D_6) - a_6 x_6(n) + x_6(n-D_6). \tag{143}$$

5.7 Wah-Wah

O efeito wah-wah é implementado na forma digital adicionando a saída de um filtro passa faixa de banda estreita com o sinal de entrada, conforme apresentado na Figura 78. Normalmente, o filtro passa-faixa utilizado é projetado com variáveis de estado e sua freqüência central é variada continuamente [18] de 300Hz a 3000Hz.

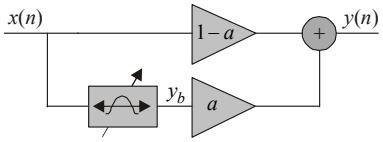


Figura 78 Diagrama de blocos do wah-wah.

Na Figura 79, é apresentado o diagrama de blocos do filtro com variáveis de estado.

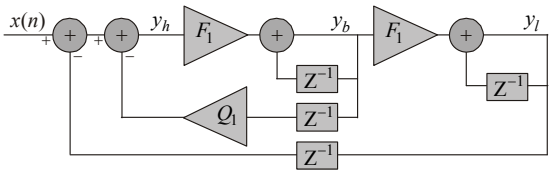


Figura 79 Diagrama de blocos do filtro com variáveis de estado.

As equações usadas para o filtro com variáveis de estado para o passa-baixa, passa-faixa e passa-alta são respectivamente:

$$y_l(n) = F_1 y_b(n) + y_l(n-1) \tag{144}$$

$$y_b(n) = F_1 y_h(n) + y_b(n-1) \tag{145}$$

$$y_h(n) = x(n) - y_l(n-1) - Q_1 y_b(n-1). \tag{146}$$

A equação final para a implementação do wah-wah é:

$$y(n) = (1-a)x(n) + ay_b(n). \tag{147}$$

5.8 Phaser

Diferentemente do uso de filtros passa-tudo utilizados nos circuitos analógicos, na forma digital, o efeito *phaser* é implementado adicionando a saída de um filtro *notch* com o sinal de entrada, conforme apresentado na Figura 80.

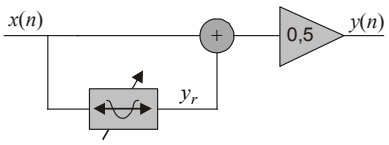


Figura 80 Diagrama de blocos do *phaser*.

A função de transferência do *phaser* é dada por [18]:

$$H(z) = \frac{1}{2}[1 + A_2(z)]. \tag{148}$$

São utilizadas estruturas de segunda ordem para o filtro *notch*, cuja função de transferência é dada por [18]:

$$A_2(z) = \frac{-a + (d - da)z^{-1} + z^{-2}}{1 + (d - da)z^{-1} - az^{-2}}. \tag{149}$$

Os parâmetros "a" e "d" são obtidos a partir das especificações de freqüência de corte (f_c), freqüência de amostragem (f_s) e largura de banda (f_b) do filtro por [18]:

$$a = \frac{\tan((\pi f_b / f_s) - 1)}{\tan((2\pi f_b / f_s) + 1)} \tag{150}$$

$$d = -\cos(2\pi f_c / f_s). \tag{151}$$

A equação discreta para a implementação do filtro *notch* é:

$$\begin{aligned} y_r(n) = &-ax(n) + d(1-a)x(n-1) + x(n-2) \\ &- d(1-a)y_1(n-1) + ay(n-2). \end{aligned} \tag{152}$$

A equação final para a implementação do *phaser* é:

$$y(n) = 0.5[x(n) + y_r(n)]. \tag{153}$$

6 EQUALIZADOR

Diferentemente dos equalizadores analógicos, que usam N seções *Bump*, o equalizador digital é normalmente implementado por uma seqüência de seções *Shelving* e *Bump*, conforme apresentado na Figura 81. Essa estratégia de implementação evita o surgimento das oscilações indesejáveis mostradas no exemplo prático 6. As seções *Shelving* são utilizadas no primeiro e no último estágio do equalizador com a função de controle de graves e de agudos, respectivamente. Já as seções *bump* têm a função de controle das médias freqüências.

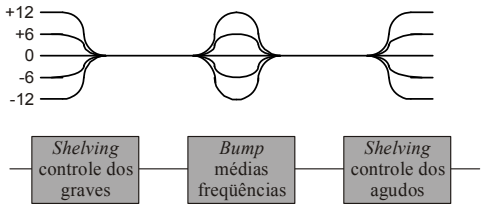


Figura 81 Conexão entre seções *shelving* e *bump* [22].

6.1 Seção Shelving

As seções *Shelving* reforçam (*boost*) ou atenuam (*cut*) os componentes de baixa e alta freqüência pela variação nos parâmetros de freqüência de corte (f_c) e ganho (H_o) do filtro. As equações do *shelving* para controle dos graves e agudos de primeira ordem [22], respectivamente, são:

$$H(z) = 1 + \frac{H_o}{2}[1 + A(z)] \tag{154}$$

$$H(z) = 1 + \frac{H_o}{2}[1 - A(z)] \tag{155}$$

onde H_o é o ganho e $A(z)$ é a função de transferência de um filtro passa-tudo de primeira ordem dada por:

$$A(z) = \frac{z^{-1} + a}{1 + az^{-1}}. \tag{156}$$

O parâmetro "a" é obtido a partir da especificação da freqüência de corte (f_c) e da freqüência de amostragem (f_s). As equações do parâmetro "a" para o controle das baixas freqüências são dadas por [15]:

$$a = \frac{\tan((\pi f_c / f_s) - 1)}{\tan((\pi f_c / f_s) + 1)}[Boost] \tag{157}$$

$$a = \frac{\tan((\pi f_c / f_s) - V_o)}{\tan((\pi f_c / f_s) + V_o)}[Cut]. \tag{158}$$

As equações de "a" para o controle das altas freqüências são [22]:

$$a = \frac{\tan((\pi f_c / f_s) - 1)}{\tan((\pi f_c / f_s) + 1)}[Boost] \tag{159}$$

$$a = \frac{V_o \tan((\pi f_c / f_s) - 1)}{V_o \tan((\pi f_c / f_s) + 1)}[Cut] \tag{160}$$

onde V_o é obtido a partir da especificação do ganho em dB por:

$$V_o = 10^{G/20} \tag{161}$$

$$H_o = V_o - 1. \tag{162}$$

O diagrama de blocos para a implementação do *shelving* é mostrado na Figura 82.

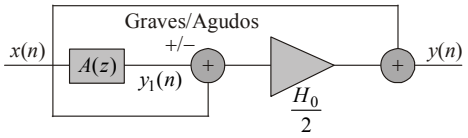


Figura 82 Diagrama para a implementação do *shelving* [18].

A equação discreta para a implementação do filtro passa-tudo é dada por:

$$y_1(n) = ax(n) + x(n-1) - ay_1(n-1). \tag{163}$$

Já a equação de saída do *shelving* (controle de graves/controle de agudos) é dada por:

$$y(n) = \frac{H_o}{2}[x(n) \pm y_1(n)] + x(n). \tag{164}$$

Caso os requisitos do *shelving* exijam que o filtro passa-tudo seja de segunda ordem, podem ser usados os procedimentos apresentados em [18].

6.2 Seção Bump (Peak)

Conforme apresentado na Figura 81, as seções *bump* atuam sobre as médias freqüências, variando os parâmetros de freqüência de corte (f_c), largura de banda (f_b) e ganho (H_o). São utilizadas estruturas de segunda ordem cuja FT [22] é dada por:

$$H(z) = 1 + \frac{H_o}{2}[1 - A_2(z)] \tag{165}$$

onde $A_2(z)$ é a função de transferência de um filtro passa-tudo de segunda ordem dada por:

$$A_2(z) = \frac{-a + (d - da)z^{-1} + z^{-2}}{1 + (d - da)z^{-1} - az^{-2}}. \tag{166}$$

Os parâmetros "a" e "d" são obtidos a partir das especificações de freqüência de corte (f_c), freqüência de

amostragem (f_s), ganho (H_o) e largura de banda (f_b) do filtro, conforme detalhado em [18]. O diagrama de blocos é mostrado na Figura 83.

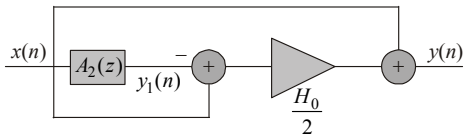


Figura 83 Diagrama de blocos para a implementação do *bump*[11].

A equação discreta para a implementação do passa-tudo é dada por:

$$\begin{aligned} y_1(n) &= -ax(n) + d(1-a)x(n-1) + x(n-2) \\ &\quad - d(1-a)y_1(n-1) + ay(n-2). \end{aligned} \tag{167}$$

Já a equação de saída da seção *peak* é dada por:

$$y(n) = \frac{H_o}{2}[x(n) - y_1(n)] + x(n). \tag{168}$$

7 RESULTADOS EXPERIMENTAIS

Os algoritmos aqui considerados foram implementados no Matlab e no ambiente do VisualDSP. Nos testes realizados, cada algoritmo foi submetido a arquivos de testes, gerando os correspondentes arquivos de saída. Os arquivos de testes e os algoritmos implementados encontram-se disponíveis para download no endereço apresentado em [23]. Deve-se observar que os arquivos foram gerados para uma freqüência de amostragem de 8 kHz e 16bits de codificação e que podem ser manuseados pela ferramenta de edição de áudio ocenaudio [24].

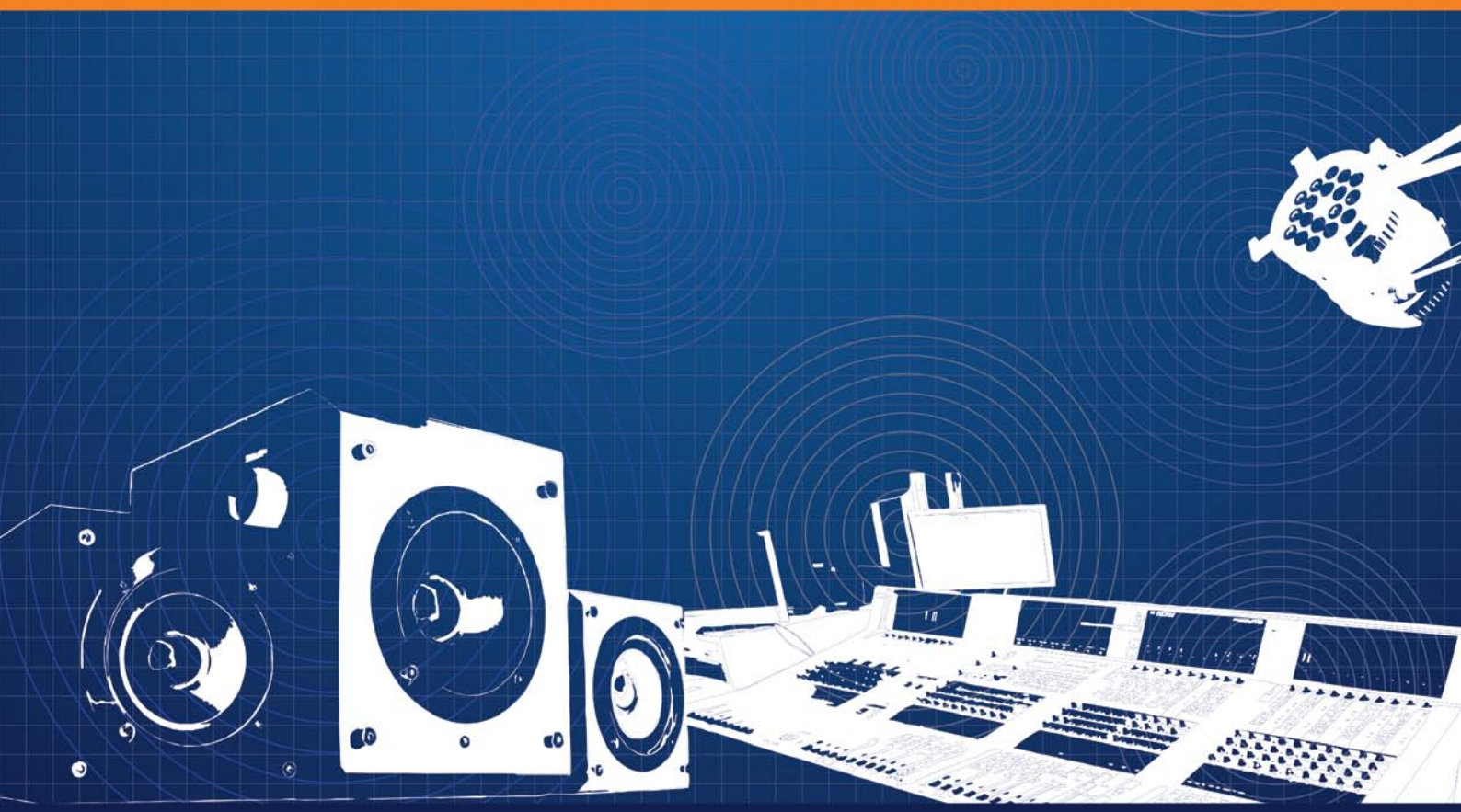
8 CONSIDERAÇÕES FINAIS

Neste tutorial, foram discutidas formas de obtenção de efeitos de áudio analógicos e digitais, assim como de equalizadores de áudio com funções de transferência dos tipos *shelving* e *bump***.** A partir dos diagramas de blocos sugeridos, o projetista pode buscar soluções variadas, considerando-se a gama de circuitos propostos para implementar os diagramas apresentados. Foram apresentadas expressões explícitas para o cálculo de filtros passa-baixa, passa-alta, passa-faixa e passa-tudo. Alguns exemplos práticos foram mostrados, com o objetivo de auxiliar os *hobistas* na implementação dos circuitos desejados.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Hunter, D.; "Guitar Effects Pedals the Practical handbook", Backbeat Book, 2004.
[2] Loy, G ; "Musimathics- Volumes 1 e 2", MIT Press, 2007.
[3] Anderton, C.; "Eletronic Projects for Musicians", Amsco Publications, 1980.
[4] Noceti Filho, S.; "Filtros Seletores de Sinais", Editora UFSC, 2010.
[5] Time –Domain Synthesis of Linear Networks, Kendalll. SU, Prentice-Hall, New Jersey, 1971
[6] K. J. Gundry, "Constant-Q Graphic Equalizers," *Journal of Audio Engineering Society*, vol. 34, no. 9, pp. 1-16, set. 1986
[7] K. J. Gundry, *Adjustable Equalizers Useable in Audio Spectrum*, U.S. Patent 3.921.104, nov. 1975.
[8] www.analog.com
[9] http://www.analog.com/en/embedded-processing-dsp/blackfin/ADSP-BF533/processors/product.html
[10] http://www.analog.com/en/embedded-processing-dsp/software-and-reference-designs/content/visualdsp_software_test_drive/fca.html
[11] ADSP-BF533 EZ-KIT Lite "Evaluation System manual" – Analog Devices.
[12] VisuaDSP++ C/C++ Compiler and Library Manual for Blackfin Processors – Analog Devices.
[13] VisualDSP++ Assembler and Preprocessor Manual – Analog Devices.
[14] Orfanidis, S. J.; "Introduction to Signal processing", Prentice Hall, 2009.
[15] Bloom, P. J.; "High-Quality Digital Audio in the Entertainment Industry: An Overview of Achievements and Challenges", IEEE ASSP.Mag, 2, October 1985.
[16] Oppenheim, A., V.; "Applications of Digital Signal Processing", ", Prentice Hall, 1978.
[17] Tomarakos, J.; Ledger, D.; "Using the Low-cost, High Performance ADSP-21065L Digital Signal Processor for Digital Audio Applications", Analog Devices Applications, April 1998.
[18] Zolzer, U.; "DAFX: Digital Audio Effects" John Wiley & Sons, Ltd – 2002.
[19] Dattorro, J; "Effect Design, part 2: Delay-Line Modulation and Chorus" J. Audio Eng. Soc, October 1997.
[20] Schroeder; M. R., "Natural Sounding Artificial reverberation" J. Audio Eng. Soc, 10; 1962.
[21] Moorer; J. A.; "About This Reverberation Business", 1979.
[22] Zolzer, U; "Digital Audio Signal Processing" John Wiley & Sons, Ltd – 2008.
[23] http://www.linse.ufsc.br/research-public.
[24] http://www.ocenaudio.com.br

FULL PAPERS



Sociedade de Engenharia de Áudio
Artigo de Congresso
Apresentado no 9º Congresso de Engenharia de Áudio
15ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2011, São Paulo, SP

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

Um tutorial sobre a medição *in situ* do coeficiente de absorção com as técnicas PP e PU

Eric Brandão[1], Arcanjo Lenzi[1]
[1]Universidade Federal de Santa Catarina
Depto. de Eng. Mecânica
Florianópolis, SC, Brasil, 88037-001
ericbacustica@gmail.com

RESUMO
Este artigo apresenta um tutorial sobre as técnicas de medição PP (pressão-pressão) e PU (pressão-velocidade de partícula) para aplicações *in situ*, da medição do coeficiente de absorção. Com tais técnicas de medição é possível caracterizar materiais acústicos aplicados em salas de concerto, salas de ensaio, estúdios, etc. As técnicas de medição PU e PP são apresentadas em detalhes, bem como o método iterativo de dedução do coeficiente de absorção. Medições feitas em câmara semi-anecóica são comparadas com medições feitas *in situ*. As técnicas PU e PP são comparadas entre si, de modo a mostrar seus pontos fortes e fracos.

0 INTRODUÇÃO

O coeficiente de absorção de uma amostra de material acústico fornece a razão entre a quantidade de energia sonora absorvida pela quantidade de energia sonora incidente na amostra. O coeficiente de absorção dos materiais aplicados em uma determinada sala é um parâmetro importante na determinação dos parâmetros acústicos como RT60, C80, etc. [1].

O termo *in situ* vem do latim e significa em sítio, o que significa que a medição acústica é tomada no ambiente em que a amostra está instalada, levando portanto em conta suas condições de montagem, e de acúmulo de sujeira e umidade, por exemplo. Tais condições não são levadas em conta em métodos medição laboratoriais, como o do tubo de impedância e o da câmara reverberante [2,3].

Existem vários textos na literatura reportando métodos de medição *in situ* do coeficiente de absorção. Tanto Garai [4] como Mommertz [5], por exemplo, apresentam métodos de medição baseados na separação temporal entre as pressões incidente e refletida. Já Allard [6], e Li e Hodgson [7] se baseiam na medição da função de transferência entre dois microfones, localizados próximos à amostra. Lanoye [8] apresenta a medição do coeficiente de absorção com a sonda PU em câmara semi-anecóica. A sonda PU é um sensor que combina em um único probe um microfone e um sensor de velocidade de partícula.

O objetivo deste artigo é apresentar um tutorial sobre as técnicas de medição *in situ* do coeficiente de absorção. Uma comparação entre as técnicas PP, similar a usada nas referências [6,7], e a técnica PU é também apresentada. A principal diferença, neste artigo, é que ambas as técnicas se valem de uma formulação do campo acústico mais precisa do que as apresentadas nas referências [6,7,8]. Adicionalmente, enquanto estes trabalhos focam em medições em câmara semi-anecóica, o objetivo aqui é demonstrar a usabilidade dos métodos em aplicações *in situ*.

O processo de medição e calibração será apresentado em detalhes, de forma que o leitor do texto seja capaz de reproduzi-lo com facilidade. O algoritmo de dedução do coeficiente de absorção também será apresentado na seção

2, e a seção 3 apresenta alguns resultados experimentais seguidos das conclusões.

1 PROCEDIMENTO DE MEDIÇÃO

O procedimento de medição pode ser dividido em duas etapas: a medição em si e o pós-processamento. A medição consiste em utilizar um sistema que possui um alto-falante, usado para excitar a amostra que se deseja caracterizar, e os sensores PU ou PP.

Um esquema do sistema de medição pode ser visto na Figura 1. A fonte sonora está a uma altura h_s [m] da superfície da amostra sob medição e a uma distância horizontal r [m] das sondas. Tais distâncias devem ser conhecidas e o operador deve medi-las, bem como as alturas z_1 e z_2, indicadas na Figura 1.

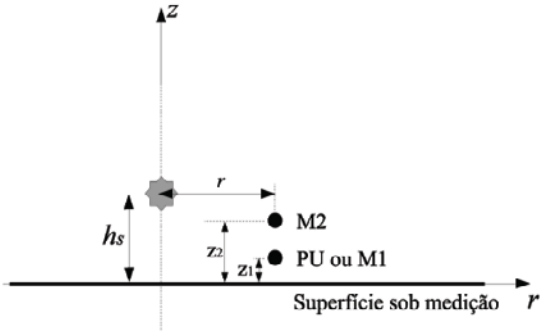


Figura 1 - Esquema do sistema de medição *in situ*.

A fonte sonora consiste num alto-falante montado no interior de uma caixa acústica esférica (chamada aqui de alto-falante esférico), e pode ser visto na Figura 2. A escolha deste tipo de caixa acústica tem a ver com suas características de radiação, próxima às características de monopolo [9,10]. O razão da necessidade da proximidade com o comportamento de um monopolo será discutido na seção 2.

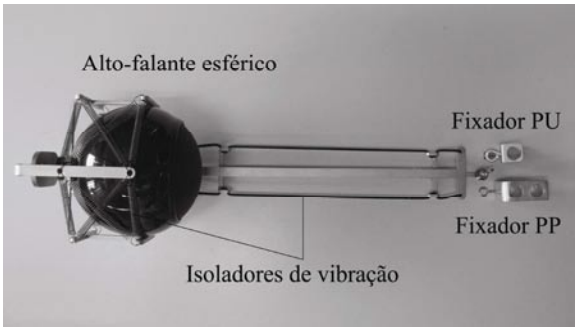


Figura 2 - Sistema de medição: alto-falante esférico e fixadores PU e PP.

Uma varredura exponencial de senos é usada como ruído de excitação. A utilização de tal estímulo (em detrimento do ruído branco) está relacionada à maior quantidade de energia, em baixas frequências, que ele é capaz de fornecer. Durante a medição a fonte sonora tem seu eixo principal orientado para a superfície da amostra, e os sensores são posicionados próximos a superfície desta. Um computador fornece à interface de áudio a varredura de senos. A saída da interface de áudio é conectada à entrada de um amplificador de potência, que alimenta o alto-falante esférico. As saídas dos sensores, por sua vez, alimentam as entradas da placa de áudio. Os sinais captados pelos sensores são digitalizados pela interface de áudio e o computador armazena os dados medidos e faz os cálculos de FFT e pós-processamento.

1.1 Técnica PP

A sonda PP é constituída por dois microfones omnidirecionais (M1 e M2) separados de uma distância vertical $d = z_2 - z_1$ [m], usualmente de 2 [cm]. O microfone M1 mede a pressão sonora p_1 [Pa] e o microfone M2 mede a pressão sonora p_2 [Pa]. Veja a Figura 3, que mostra uma fotografia da sonda PP.

Como os dois microfones podem ser ligeiramente diferentes entre si uma calibração relativa ajuda a reduzir os erros de medição. Durante esta calibração os dois microfones são posicionados sobre uma superfície rígida, com seus diafragmas a cerca de 1 [mm] um do outro. A fonte sonora, posicionada à 30 [cm] dos microfones, é acionada e a função de transferência H_c entre p_1 e p_2 é medida ($H_c = p_1/p_2$). Como os dois microfones ocupam basicamente a mesma posição no espaço eles devem, em teoria, medir a mesma pressão sonora; portanto, se M1 e M2 forem exatamente iguais $H_c = 1$. Seguindo este raciocínio as diferenças entre M1 e M2 se expressam numa $H_c \neq 1$.

Após a calibração relativa M1 e M2 são posicionados acima da superfície da amostra a se caracterizar, como na Figura 3. A função de transferência é novamente medida e corrigida pela calibração relativa, de forma que:

$$H_m(f) = \frac{H_{12}(f)}{H_c(f)}, \quad (1)$$

onde H_{12} é a função de transferência medida próxima a amostra e H_m é a função de transferência corrigida por H_c; H_m será usada no algoritmo de dedução do coeficiente de absorção.

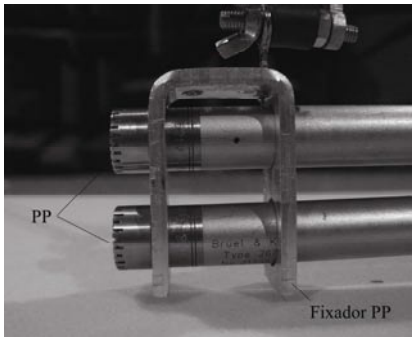


Figura 3 - Sonda PP acima da amostra a ser caracterizada.

1.2 Técnica PU

A sonda PU consiste numa sonda que integra 1 microfone omnidirecional e 1 sensor de velocidade de partícula (figura de 8); seu comportamento físico é descrito nas referências [11,12]. Como a sonda integra os dois sensores em um único ponto a altura da sonda PU em relação a superfície da amostra é z_1 [m].

A calibração da sonda PU também é necessária e o método utilizado aqui é a calibração em campo livre, também utilizado nas referências [8,10]. Este passo

consiste em apontar a fonte sonora e a sonda PU, separados de uma distância r_c, para longe de qualquer superfície refletora; esta condição simula uma condição de campo livre. Neste caso a função de transferência entre pressão (p_{ff}) e velocidade de partícula (u_{ff}) é medida e uma função de correção é criada a partir da impedância de radiação em campo livre de um monopolo, dada por:

$$CF(f) = \frac{ikr_c}{ikr_c+1}\left(\frac{u_{ff}(f)}{p_{ff}(f)}\right), \quad (2)$$

onde $i = \sqrt{-1}$ e $k = 2\pi f/c$ é o número de onda no ar e r_c é a distância horizontal entre fonte sonora e sensor.

Após a medida de calibração a sonda PU é posicionada acima da superfície da amostra a se caracterizar, como na Figura 4. A função de transferência p/u é novamente medida e corrigida por CF, de forma que a impedância característica na posição do sensor seja:

$$Z_m(f) = \frac{p(f)}{u(f)} CF. \quad (3)$$

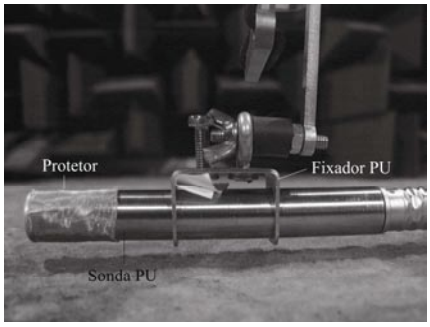


Figura 4 - Sonda PU acima da amostra a ser caracterizada.

Uma média móvel é aplicada após cada medição de forma a suavizar as repostas em frequência [10], e assim minimizar os efeitos de reflexões indesejadas.

2 ALGORITMO DE DEDUÇÃO

Devido a proximidade entre o alto-falante esférico e os sensores as frentes de onda sonora, acima da amostra, não podem ser consideradas planas, mas sim esféricas. Devido a diferença de simetria entre as frentes de onda (esféricas) e a superfície da amostra (plana) o fenômeno de reflexão é bastante complexo [13], e é necessário recorrer a transformada espacial de Fourier para resolvê-lo. Para uma superfície infinita e localmente reativa a equação a seguir, derivada por Di e Gilbert [14], descreve a pressão sonora acima da amostra.

$$p = \frac{e^{-ikR_1}}{R_1} + \frac{e^{-ikR_2}}{R_2} - \frac{2k}{Z}\int_0^\infty e^{-k\frac{k}{Z}q}\,\frac{e^{-ik\sqrt{r^2+(h_s+z-iq)^2}}}{\sqrt{r^2+(h_s+z-iq)^2}}\,\mathrm{d}q, \quad (4)$$

onde $R_1 = \sqrt{r^2+(h_s-z)^2}$, $R_2 = \sqrt{r^2+(h_s+z)^2}$ e Z é a impedância de superfície da amostra que se deseja caracterizar.

Como se pode observar é impossível inverter analiticamente a Equação 4, para se obter Z através das duas pressões sonoras (H_m) ou da pressão e velocidade de partícula (Z_m) medidas acima da amostra. Um algoritmo iterativo deve, então, ser utilizado. No caso da medição com a técnica PP tal algoritmo busca o melhor valor de Z que se ajusta à função de transferência H_m. Similarmente, no caso da medição com a técnica PU tal algoritmo busca o melhor valor de Z que se ajusta à impedância característica Z_m.

O algoritmo iterativo pode ser descrito nos seguintes passos:

1. Uma estimativa inicial da impedância de superfície (Z^0) é calculada por um método de dedução mais simples, a ser descrito no apêndice A;
2. A estimativa inicial Z^0 é inserida na Equação 4 (e ou na equação da velocidade de partícula - anexo B) e uma função de transferência (H_{mc}) ou uma impedância característica (Z_{mc}) é calculada.
3. Para a técnica PP H_{mc} é subtraído de H_m. Para a técnica PU Z_{mc} é subtraído de Z_m. Se o módulo da diferença é suficientemente pequeno, isto significa que a estimativa inicial de Z é a impedância de superfície da amostra e o algoritmo é parado;
4. Se o módulo da diferença é maior que 0.000001 uma segunda estimativa de Z é calculada pelo método da secante [15], e os passos 3 e 4 são repetidos novamente até que o critério de convergência seja satisfeito ou um número máximo de iterações seja atingido.

Uma vez que Z é obtido, o coeficiente de absorção é calculado por:

$$\alpha = 1 - \left|\frac{Z-1}{Z+1}\right|^2. \quad (5)$$

Vale observar que o algoritmo iterativo assume que um monopolo irradia ondas esféricas acima da amostra, e esta é a razão que leva a necessidade que a fonte sonora tenha uma impedância de radiação próxima a de um monopolo.

3 RESULTADOS

Nesta seção são apresentados os resultados de algumas medições feitas com as técnicas PU e PP com fins de comparação.

Num primeiro experimento a amostra medida é uma lã de rocha de dimensões 1.2 x 0.6 [m^2], densidade 64 [Kg/m^3] e espessura $d_1 = 0.025$ [m]. A amostra é posicionada sobre o piso rígido de uma câmara semi-anecóica cuja frequência de corte inferior é 150 [Hz]. O alto-falante esférico é montado a uma altura $h_s = 0.3$ [m] da superfície da amostra. A sonda PU estava a uma altura $z_1 = 0.015$ [m] da superfície da amostra. O microfone M1 estava a uma altura $z_1 = 0.015$ [m] e o microfone M2 a uma altura $z_2 = 0.035$ [m]. Incidência normal é considerada de forma que $r = 0$ em todos os experimentos. Uma fotografia do sistema de medição usado pode ser vista na Figura 5.

Na medição da amostra na câmara semi-anecóica o procedimento foi primeiramente calibrar as sondas PU e PP e depois de posicionar cada sensor próximo a amostra. 30 medições eram tomadas em sequência, sem mexer no aparato de medição.

A Figura 6 mostra o coeficiente de absorção, média das 30 medições, com as técnicas PU (linha cheia) e PP (linha pontilhada). Em geral pode-se observar que as curvas concordam muito bem numa faixa de frequência que vai de 100-6000 [Hz].

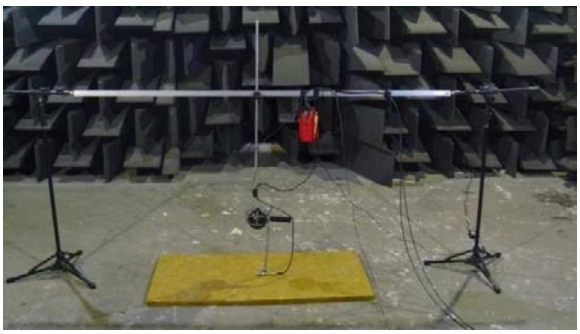


Figura 5 - Sistema de medição na usado na câmara semi-anecoica.

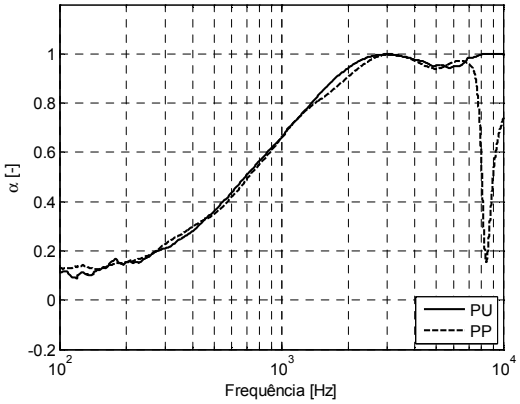


Figura 6 - Coeficiente de absorção da amostra de lã de rocha de densidade 64 $[\mathrm{Kg/m^3}]$ e 25 [mm] de espessura.

Observou-se, no entanto, que a repetibilidade da medição com a técnica PU foi altíssima na faixa de 100-10000 [Hz]. Já a repetibilidade com a técnica PP foi baixa para frequências menores que 400 [Hz]. Isto se deve ao fato que em baixas frequências microfones muito próximos (2 [cm] no caso) irão medir essencialmente a mesma pressão sonora; como a precisão do analisador de sinais é finita, ocorrerá um erro aleatório em cada medição, que contribui para diminuição da repetibilidade. Esta frequência de corte inferior esta de acordo com as recomendações encontradas na referência [16] e na norma para medição em tubo de impedância referência [2], que recomenda uma separação entre os microfones pelo menos 5% do maior comprimento de onda de interesse. Neste caso, para uma frequência de corte inferior de 100 [Hz], a separação entre os microfones deveria ser de cerca de 18 [cm].

No entanto, aumentar a distância entre os microfones tem pelo menos dois custos: 1) O aumento do número de medições e equipamentos necessários, o que acaba por complicar o processo de medição. Como a sonda PU exibe alta reprodutibilidade esta se torna uma alternativa mais viável que a técnica PP; e 2) O aumento da distância entre os microfones implica que um deles estará bastante afastado da superfície da amostra, estando mais sujeito portanto aos efeitos de reflexões espúrias em aplicações *in situ* e à influência do tamanho finito da amostra, como mostram as referências [10,17].

A técnica PP também exibe um comportamento errático para frequências acima de 6 [kHz]; acima desta frequência a distância entre os microfones se torna da ordem de 1/2 comprimento de onda e dificuldades na localização exata dos centros acústicos dos microfones tornam a medição errática. Este limite também esta de acordo com a recomendação da norma para tubo de impedância [2], que recomenda que a separação entre os microfones seja no máximo 45% do menor comprimento de onda de interesse. O aumento da distância entre os microfones, a fim de compensar sua baixa repetibilidade em baixas frequências, também acabaria por diminuir esse limite superior da medição, o que implica que para cobrir uma faixa ampla do espectro pelo menos duas medições são necessárias com a técnica PP.

O segundo experimento consiste em usar as técnicas PU e PP em ambientes mais realistas. Dessa vez uma amostra de lã de rocha de dimensões 1.2 x 0.6 $[\mathrm{m^2}]$, densidade 80 $[\mathrm{Kg/m^3}]$ e espessura $d_1 = 0.05\,[\mathrm{m}]$ foi medida primeiramente na câmara semi-anecóica e posteriormente num escritório com um volume de 45 $[\mathrm{m^3}]$ e numa câmara reverberante de 200 $[\mathrm{m^3}]$. O coeficiente de absorção medido com a técnica PU é mostrado na Figura 7 e o medido com a técnica PP é mostrado na Figura 8.

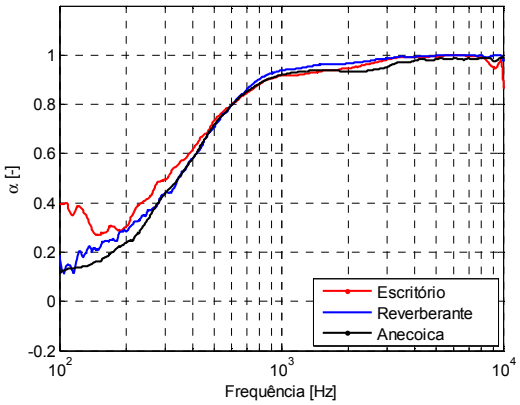


Figura 7 - Medição *in situ* e em câmara semi-anecóica do coeficiente de absorção de uma amostra de lã de rocha de densidade 80 $[\mathrm{Kg/m^3}]$ e 50 [mm] de espessura com a técnica PU.

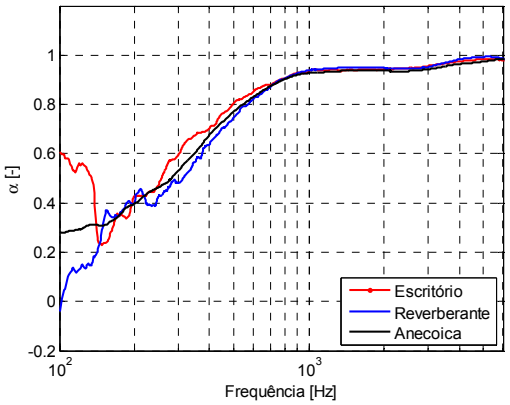


Figura 8 - Medição *in situ* e em câmara semi-anecóica do coeficiente de absorção de uma amostra de lã de rocha de densidade 80 $[\mathrm{Kg/m^3}]$ e 50 [mm] de espessura com a técnica PP.

Em primeiro lugar pode-se notar que, em ambos os casos, a curva do coeficiente de absorção, medido em câmara semi-anecóica, parece mais suave que os coeficientes de absorção medidos *in situ*. Isto se deve à presença das reflexões das paredes laterais e teto, bem como dos modos acústicos que acabam por perturbar a

medição, especialmente nas baixas frequências. A média móvel consegue lidar relativamente bem com o problema, e em ambos os casos, embora haja desvios, eles estão dentro de uma faixa esperada de incerteza, e não são significativos ao ponto de invalidar a medição.

Observa-se também comparando as Figuras 7 e 8 que os desvios do coeficiente de absorção medido *in situ* em relação ao medido em câmara semi-anecoica parecem menores para a medição com a técnica PU. Isto é atribuído ao fato de que o sensor de velocidade de partícula possui um padrão polar de figura de 8 [12], enquanto que os microfones são omni-direcionais. Isto possibilita a orientação conveniente da sonda PU de forma que reflexões indesejadas sejam evitadas.

4 CONCLUSÃO

Neste artigo as técnicas de medição *in situ* do coeficiente de absorção foram apresentadas, sendo estas referidas aqui como a técnica PP, que usa dois microfones, e a técnica PU que usa uma sonda que integra 1 microfone e 1 sensor de velocidade de partícula. Mostrou-se detalhadamente o aparato de medição, o procedimento de medição aplicado a cada técnica, e o esquema de pós-processamento dos dados medidos, que envolve um algoritmo iterativo para o calculo do coeficiente de absorção.

Medições em câmara semi-anecóica e em ambientes realistas foram apresentadas, onde se notou que as técnicas PU e PP possuem boa concordância. A técnica PP sofre, no entanto, com limitações em baixas e altas frequências, devido ora à separação insuficiente entre os microfones (baixas frequências), e ora a dificuldades na localização dos centros acústicos dos mesmos (altas frequências). Neste quesito a técnica PU se mostrou mais robusta, possuindo uma repetibilidade maior.

O desempenho das técnicas em aplicações *in situ* também se mostrou satisfatório o que demonstra a aplicabilidade de ambas as técnicas para a caracterização de materiais em salas já existentes como salas de concerto, estúdios, salas de aula e escritórios.

APÊNDICE A – ESTIMATIVA DE Z^0

O modelo de campo acústico usado na estimativa de Z^0 envolve uma simplificação na reflexão das ondas esféricas na superfície sob medição. Tal modelo considera que as ondas esféricas refletem como se fossem ondas planas.

Neste caso, o coeficiente de reflexão, em função da impedância característica Z_m é:

$$V_p = \frac{Z_m \frac{h_s - z}{R_1}\left(\frac{1}{ikR_1}+1\right)\frac{R_2}{R_1}}{Z_m \frac{h_s + z}{R_2}\left(\frac{1}{ikR_2}+1\right)} e^{-ik(R_{11}-R_{12})}. \quad \text{(A1)}$$

Já o coeficiente de reflexão em função de H_m é:

$$V_p = \frac{\frac{e^{-ikR_{11}}}{R_{11}} - H_m \frac{e^{-ikR_{21}}}{R_{21}}}{H_m \frac{e^{-ikR_{22}}}{R_{22}} - \frac{e^{-ikR_{12}}}{R_{12}}}. \quad \text{(A2)}$$

onde $R_{11} = \sqrt{r^2 + (h_s - z_1)^2}$, $R_{12} = \sqrt{r^2 + (h_s + z_1)^2}$, $R_{21} = \sqrt{r^2 + (h_s - z_2)^2}$ e $R_{22} = \sqrt{r^2 + (h_s + z_2)^2}$.

Uma vez obtido o coeficiente de reflexão a estimativa inicial para Z pode ser feita da seguinte forma:

$$Z^0 = \frac{1+V_p}{1-V_p} \frac{\sqrt{r^2+h_s^2}}{h_s} \frac{ik\sqrt{r^2+h_s^2}}{1+ik\sqrt{r^2+h_s^2}}. \quad \text{(A3)}$$

APÊNDICE B – CÁLCULO DA VELOCIDADE DE PARTÍCULA

A velocidade de particula é derivada a partir da Equação (4) e da equação de Euler: $\rho \frac{\partial u}{\partial t} = -\nabla p$, e é dada por:

$$u_z = \frac{e^{-ikR_1}}{R_1}\left[\frac{1}{ikR_1}+1\right]\left(\frac{h_s - z}{R_1}\right) - \frac{e^{-ikR_2}}{R_2}\left[\frac{1}{ikR_2}+1\right]\left(\frac{h_s + z}{R_1}\right) + \frac{2k}{Z}\int_0^\infty e^{-k\frac{k}{Z}q} \frac{e^{-ikR_q}}{R_q}\left[\frac{1}{\mathrm{ikR_q}}+1\right]\left(\frac{\mathrm{h_s+z-iq}}{\mathrm{R_q}}\right)\mathrm{d}q. \quad \text{(B1)}$$

onde $R_q = \sqrt{r^2 + (h_s + z - iq)^2}$.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] H. Kuttruff, *Room acoustics*.: Spon Press, 2009.

[2] E. DIN, "10534-2: Acoustics-Determination of sound absorption coefficient and impedance in impedance tubes-Part 2: Transfer-function method," ANSI, 2001.

[3] ISO. 354, Measurement of sound absorption in a reverberation room, 1985.

[4] M. Garai, "Measurement of the sound-absorption coefficient in situ: the reflection method using periodic pseudo-random sequences of maximum length," *Applied Acoustics*, vol. (39), pp. 119-139, 1993.

[5] E. Mommertz, "Angle-dependent in-situ measurements of reflection coefficients using a subtraction technique," *Applied Acoustics*, vol. 46, pp. 251-264, 1995.

[6] J.F. Allard and B. Sieben, "Measurements of acoustic impedance in a free field with two microphones and a spectrum analyzer," *The Journal of the Acoustical Society of America*, vol. 77 (4), pp. 1617-1618, 1985.

[7] J.F. Li and M. Hodgson, "Use of pseudo-random sequences and a single microphone to measure surface impedance at oblique incidence," *The Journal of the Acoustical Society of America*, vol. 102 (4), pp. 2200-2210, 1997.

[8] R. Lanoye, G. Vermeir, W. Lauriks, R. Kruse, and V. Mellert, "Measuring the free field acoustic impedance and absorption coefficient of sound absorbing materials with a combined particle velocity-pressure sensor," *The Journal of the Acoustical Society of America*, vol. 119 (5), pp. 2826-2831, 2006.

[9] F. Jacobsen and V. Jaud, "A note on the calibration of pressure-velocity sound intensity probes," *The Journal of the Acoustical Society of America*, vol. 120, p. 830, 2006.

[10] K. Hirosawa et al., "Comparison of three measurement techniques for the normal absorption coefficient of sound absorbing materials in the free field," *The Journal of the Acoustical Society of America*, vol. 126 (6), pp. 3020-3027, 2009.

[11] H.E. Bree, "An overview of microflown technologies," *Acta acustica united with Acustica*, vol. 89, pp. 163-172, 2003.

[12] H.E. de Bree, "The Microflown E-Book (Chapter 6)," Microflown Technologies, 2007.

[13] L.M. Brekhovskikh and O.A. Godin, *Acoustics of layered media II: point sources and bounded beams*.: Springer Verlag, 1992.

[14] X. Di and K.E. Gilbert, "An exact Laplace transform formulation for a point source above a ground surface," *The Journal of the Acoustical Society of America*, vol. 93 (2), pp. 714-720, 1993.

[15] J.E. Dennis and R.B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*.: Society for Industrial Mathematics, 1996.

[16] J.F. Allard, Y. Champoux, and J. Nicolas, "Pressure variation above a layer of absorbing material and impedance measurement at oblique incidence and low frequencies," *The Journal of the Acoustical Society of America*, vol. 86, p. 766, 1989.

[17] E. Brandão, E.R. Fulco, and A. Lenzi, "Estimative of the errors caused by the sample size effect on the in situ measurement of the absorption coefficient," 2010, p. 1-10.

Sociedade de Engenharia de Áudio

Artigo de Congresso

Apresentado no 9º Congresso de Engenharia de Áudio
15ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2011, 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.

Avaliação de medições de coeficiente de espalhamento de amostras de difusores sonoros com elementos semicilíndricos e cilíndricos concebidos a partir de design serial: influência da natureza dos elementos e da qualidade de ordenação

José Augusto Mannis[1], Christian dos Santos[2], Stelamaris Rolla Bertoli[2], Regiane Pupo[3], Marcelo Fernandes Oliveira[4], Jorge Vicente Lopes da Silva[4]

[1]Departamento de Música, Unicamp
Campinas, SP, 13083-970, Brazil

[2]Laboratorio de Conforto Ambiental e Física,
Faculdade de Engenharia Civil, Arquitetura e Urbanismo - Universidade Estadual de Campinas
Av. Albert Einstein, 951 - Caixa Postal: 6021 - CEP: 13083-852 - Campinas – SP

[3]Laboratório de Automação e Prototipagem para Arquitetura e Construção,
Faculdade de Engenharia Civil, Arquitetura e Urbanismo - Universidade Estadual de Campinas
Av. Albert Einstein, 951 - Caixa Postal: 6021 - CEP: 13083-852 - Campinas – SP

[4]Divisão de Tecnologias Tridimensionais (DT3D),
Centro de Tecnologia da Informação Renato Archer – CTI
Rodovia Dom Pedro I (SP - 65) Km 143,6 Bairro: Amarais
Campinas - São Paulo Brasil CEP: 13069-901

jamannis@uol.com.br , christian-st@uol.com.br , rolla@fec.unicamp.br, regipupo@terra.com.br, marcelo.oliveira@cti.gov.br, jorge.silva@cti.gov.br

RESUMO

Este trabalho busca avaliar o espalhamento produzido por difusores sonoros com elementos semicilíndricos ordenados serialmente (SCSS) como também comparar a eficiência dos mesmos em relação à seqüência idênticas com elementos cilíndricos, e ainda verificar o impacto da variação da ordenação dos elementos no desempenho de um difusor. Foram realizados modelos em escala 1:5 com métodos de prototipagem rápida (3D *printing*) empregando tecnologia de Sinterização Seletiva a Laser (SLS). Os modelos foram submetidos a medições em Câmara Reverberante em escala 1:5, obedecendo o método descrito na norma internacional ISO 17497:2004. O coeficiente de espalhamento acústico é uma grandeza associada às características geométricas de uma superfície, notadamente a rugosidade. Os elementos difusores acústicos, utilizados para readequação da energia sonora em um ambiente, utilizam destes coeficientes de espalhamento para representar a rugosidade e o grau de redirecionamento da energia que incide nestes elementos. Este trabalho apresentará os resultados

obtidos do ensaio de coeficiente de espalhamento de difusores com elementos cilíndricos e semicilíndricos dispostos em diferentes seqüências a fim de observar os resultados dos ensaios considerando as variações resultantes das seqüências. Observa-se que os difusores com elementos cilíndricos repousados sobre uma superfície plana apresentaram um coeficiente de espalhamento maior, devido a sua maior altura e maior incidência de cavidades na superfície resultante, mas ao mesmo tempo um coeficiente de absorção maior em bandas intermediárias dos registros médio e grave. Dentre as superfícies observadas esperava-se um coeficiente de espalhamento maior para superfícies com maior contraste entre as alturas adjacentes dos seus componentes, bem como uma certa semelhança entre os resultados a partir de elementos simplesmente permutados, portanto pertencentes a uma mesma superfície rugosa. Porém foi observado que os maiores valores de espalhamento se deram para sequencias de elementos com variação média entre as saliências (ou seja, variação de proporção média entre os raios dos cilindros e semicilindros), mas apresentando características claras de repetição ou elevada semelhança de *patterns* de agrupamento de elementos sequenciados, apontando para uma influência no desempenho de espalhamento da maneira (qualidade) como os elementos são ordenados, com efeito significativo em relação às características quantitativas de rugosidade e variação média de saliências da superfície.

0 INTRODUÇÃO

Superfícies seriais com elementos semicilíndricos (SCSS)

As superfícies seriais com elementos semicilíndricos foram desenvolvidas durante pesquisas na Unicamp [1] e patenteadas por esta instituição em 2008. Trabalhando sobre o contorno geométrico de uma superfície, o principio básico das superfícies seriais é obter o máximo de variação com um mínimo de elementos, o que no campo artístico seria uma atitude própria a processo criativos de linhagens estéticas mais voltadas à estruturação do material empregado para o discurso expressivo, como por exemplo o serialismo em música, desenvolvido a partir do Séc. XX, tendo inspirado esta pesquisa de superfícies seriais.

Figura 1 – SCSS: Sequenciamento de elementos semicilíndricos de diâmetros variados : Museu da Imagem e do Som (MIS) de São Paulo

No principio serial, a partir de um reduzido numero de elementos, busca-se ordenações resultando num máximo de variações possíveis. Fazer o *máximo* com o *mínimo*. No caso das superfícies Seriais com elementos semicilíndricos (SCSS) o elemento básico a ser ordenado é o diâmetro do cilindro gerador de cada elemento.

1 TÉCNICA DE PROTOTIPAGEM

Prototipagem rápida

A prototipagem rápida (PR) é uma expressão que define uma série de tecnologias de construção de peças, fatia-a-fatia [2]. Uma característica distintiva do processo de construção fatia-a-fatia é que podem ser construídas peças de alta complexidade, não possíveis de serem construídas por qualquer outro método de fabricação. Uma outra característica importante é que essas peças complexas podem ser construídas rapidamente daí a origem da expressão prototipagem rápida. A prototipagem rápida é também conhecida como *additive fabrication*, *solid free form fabrication*, *3D printing, layer by layer fabrication*. No entender dos autores, a expressão "fabricação fatia-a fatia" é a melhor definição para a série de tecnologias em questão, pois faz distinção clara da natureza do processo de fabricação evitando controvérsias relativas às demais tecnologias de fabricação existentes e há mais tempo estabelecidas. Quando as peças utilizadas têm uma aplicação como protótipo funcional – como no presente caso – elas são melhor definidas como protótipos funcionais [3]. Na PR, um desenho digital (um modelo CAD) é transformado num arquivo STL (*Stereolithography* – aproximação da superfície da peça usando malhas de triângulo). Esse desenho é fatiado e, em seguida, estas fatias são reproduzidas fisicamente por um processo físico-químico, controlado por computador, em um material que pode estar na forma de pó, folha, líquido ou fio, conforme a tecnologia PR empregada. A tecnologia empregada também vai determinar o tipo de suporte que sustenta a peça. O suporte pode ser feito com o mesmo material com que a peça vai ser construída ou com um material diferente. Também o suporte pode ser estruturado ou não pelo mesmo processo físico-químico de construção da peça. À medida que uma fatia física vai sendo construída ela vai sendo empilhada, alinhada e aderida simultaneamente à anterior até que a peça fique completa. Ao final do processo, a peça é liberada do suporte e, dependendo da tecnologia, é submetida ou não a um pós-processo de reforço mecânico.

Tecnologia de Sinterização Seletiva a Laser (*Selective Laser Sintering - SLS*)

O processo físico-químico sobre o qual se baseia a tecnologia SLS é a sinterização, que é um fenômeno de aglutinação ao nível das superfícies das partículas sólidas provocado pela elevação da temperatura. Esta aglutinação superficial é o que diferencia a sinterização da fundição. Neste último todo o volume das partículas é fundido, requerendo uma temperatura superior à requerida pela sinterização. As peças resultantes da sinterização são intrinsecamente porosas e apresentam superfícies rugosas. Para o caso da tecnologia SLS, o calor necessário para a sinterização é gerado pela interação de um feixe de laser de dióxido de carbono (comprimento de onda = 10.3 μm) com as partículas de pó. A seletividade é obtida pela ação de um scanner que direciona um feixe de laser obedecendo acompanhando as geometrias da correspondentes fatias digitais.

A Figura 2 ilustra como a tecnologia SLS funciona. Fundamentalmente são dois os processos envolvidos – a formação das camadas e a sinterização seletiva que se segue para construção das fatias. Esses processos ocorrem ao nível de uma plataforma que possui cinco cavidades sob as quais estão conectadas cinco câmaras dispostas em linha reta. As três câmaras centrais possuem um fundo móvel enquanto que as duas nas extremidades possuem fundos fixos. A câmara central é a câmara de construção, onde ocorre a sinterização seletiva a laser e as duas adjacentes são as câmaras de alimentação de pó. As câmaras nas extremidades da plataforma (fundo fixo) são as câmaras de coleta do pó. A mobilidade dos fundos das três câmaras centrais, aliada ao descolamento translacional do rolo rotativo, ao longo de toda a plataforma, são os fatores que permitem a formação das camadas a serem sinterizadas.

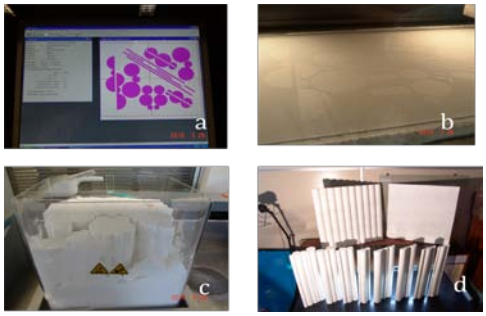


Figura 2 - Fotos mostrando as etapas de construção dos elementos sonoros deste trabalho com a tecnologia SLS: (a) Fatia digital da peça; (b) construção física da fatia digital; (c) volume de construção (conjunto peça + pó) sendo retirado da máquina dentro de um invólucro transparente; (d) Peças após completa remoção do pó.

A seqüência de etapas na formação das camadas Figura 2 é a seguinte: O fundo móvel de uma câmara de alimentação se eleva ao ponto de expor uma determinada quantidade de pó em posição um pouco acima do nível da plataforma. Simultaneamente o fundo da câmara de construção abaixa formando um recesso com altura igual à espessura da camada a ser formada. Nesta situação, o rolo coleta o pó exposto e o arrasta ao longo de toda a plataforma preenchendo o recesso da câmara de construção e levando o excesso de pó para a câmara coletora. Após o tempo decorrido para sinterização da camada de pó, o fundo da segunda câmara de deposição se eleva, expondo o pó acima do nível da plataforma. Concomitantemente, o movimento de descida do fundo da câmara de construção se repete, abrindo espaço para preenchimento como uma nova camada de pó que, ao ser seletivamente sinterizada, produz a fatia seguinte da peça. Os movimentos sincrônicos do rolo, fundos das câmaras e do feixe do laser se repetem até que a peça seja finalizada. Terminada a peça, o volume de construção (conjunto pó/peça) é coletado para dentro de recipiente (Figura 2c) que, por sua vez, é levado para uma estação de limpeza onde o pó é removido liberando as peças nele imersas.

2 MEDIÇÃO EM CÂMARA REVERBERANTE

Assim como as superfícies que absorvem a energia acústica incidente, as superfícies denominadas reflexivas possuem coeficientes que representam a reflexão da energia incidente [4]. Estes coeficientes são empregados na predição e projeto acústico de salas e espaços próprios para *performance* artística, reprodução ou gravação sonora. A necessidade de se obter um coeficiente representativo das irregularidades de uma superfície surgiu não apenas com o modelamento computacional, mas por observações de melhoria no parâmetro de lateralização de salas de concerto que adotavam difusores para redistribuir a energia sonora [5]. Um destes coeficientes é conhecido como coeficiente de espalhamento [6]. O coeficiente de espalhamento representa a capacidade de uma superfície de redirecionar aleatoriamente a energia acústica incidente e pode ser determinado por um método descrito na norma internacional ISO 17497:2004, originada a partir dos estudos realizados por Vorländer e Mommertz [6].

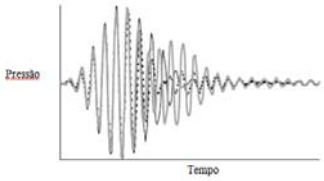


Figura 3 - Pulsos refletidos para diferentes orientações de amostra. [6][1]

O princípio básico do coeficiente de espalhamento é obter a quantidade de energia refletida fora da zona especular. A zona especular é aquela que representa uma reflexão com o mesmo ângulo, em relação a normal ao plano de reflexão, do ângulo da energia incidente.
Comparando os impulsos resultantes de reflexões para uma amostra com diferentes orientações, segundo Vorländer e Mommertz [6], a parte inicial dos impulsos possui uma grande correlação, o que não ocorre com a parte tardia destes impulsos, como visto na Figura 3.
A resposta impulsiva resultante pode ser considerada como a resposta impulsiva da energia especular.

Figura 4 - Diagrama esquemático das componentes da energia espalhada sobre uma superfície rugosa. [6][2]

A energia especular refletida é dada pela equação 2.

$$E_{spec} = (1-\alpha)(1-s) = (1-a) \qquad (2)$$

onde: α = coeficiente de absorção.
a = coeficiente de absorção especular.
s = coeficiente de espalhamento.

A energia total incidente pode ser representada pela equação 3.

$$E_{total} = (1-\alpha) \qquad (3)$$

Lembrando que o coeficiente de espalhamento é a energia refletida fora da região especular, o coeficiente de espalhamento pode ser determinado pela equação 4.

$$s = 1 - \frac{E_{spec}}{E_{total}} = 1 - \frac{(1-a)}{(1-\alpha)} = \frac{a-\alpha}{1-\alpha} \qquad (4)$$

[1] p.187-199
[2] p. 187-199

O coeficiente de espalhamento pode ser determinado em campo livre, mas a técnica mais utilizada e descrita na norma internacional ISO 17497:2004 [7] é a que utiliza campo reverberante, podendo ser realizada em escala ou em tamanho real.

O método descrito pela ISO 17497:2004 se utiliza da medição de dois coeficientes conhecidos: o coeficiente de absorção da superfície e o coeficiente de absorção especular da superfície. O coeficiente de absorção especular é o obtido pela determinação do coeficiente de absorção segundo ISO 3382-1:2009 [8] na condição de se obter a resposta impulsiva especular. Esta, conforme visto anteriormente, pode ser obtida pela soma de várias respostas em várias orientações da amostra ou da determinação do coeficiente de absorção com a amostra em movimento circular. Para realizar este movimento recomenda-se o uso de uma mesa giratória para posicionamento ou realização de giro contínuo.

3 AMOSTRAS DE ENSAIO

A sequência de diâmetros para as amostras a serem medidas foi estabelecida buscando construir relações sempre variadas entre os mesmos elementos, de forma que a cada vez que apareçam estejam ladeados diferentemente. Assim definiu-se a sequência de diâmetros: (6,0 / 3,0 / 7,0 / 2,0 / 5,0 / 4,0 / 3,0 / 6,0 / 5,0 / 7,0 / 2,0 / 4,0 / 6,0cm) contendo quatro grupos de tamanhos comparáveis: 16 / 11 / 14 / 19 cm.

• AMOSTRA 0: Placa plana de 60x60cm contendo somente o material de revestimento (SLS) dos elementos, para medir o coeficiente de absorção do material empregado na prototipagem;

• AMOSTRA 1: 10 elementos semicirculares iguais alinhados, tendo cada um 6cm de diâmetro, constituindo portanto um difusor com elementos semicirculares regulares. Esta amostra teve por objetivo verificar a influência da regularidade dos elementos no desempenho dos difusores semicilíndricos. Portanto não foram realizados modelos com elementos cilíndricos para esta amostra, mesmo porque, por terem todos os cilindros o mesmo diâmetro, o perfil resultante seria muito similar ao da amostra com elementos semicilíndricos, com exceção das bordas;

• AMOSTRAS 2 a 5: Semicilíndricas: quatro grupos de elementos semicirculares alinhados.

• AMOSTRAS 2 a 5: Cilíndricas: quatro grupos de elementos circulares com diâmetros idênticos aos semicirculares, para investigar a resposta do difusor constituído por elementos cilíndricos inteiros.

Todos os materiais acima, quando reunidos integralmente com seu grupo ocupam o espaço de 60x60cm do disco de rotação da Câmara Reverberante.

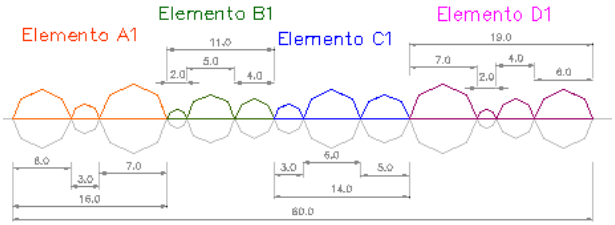


Figura 5 – Agrupamentos de elementos adotados A, B, C e D em corte transversal.

O objetivo da divisão em grupos foi poder permutar os elementos no momento das medições para verificar o impacto da ordenação no desempenho do difusor. As amostras de elementos para medições compreenderam:

	DIÂMETROS	
	1:5	1:1
Período	6,0 cm	30 cm
A	6,0 / 3,0 / 7,0 cm	30 / 15 / 35 cm
B	2,0 / 5,0 / 4,0 cm	10 / 25 / 20 cm
C	3,0 / 6,0 / 5,0 cm	15 / 30 / 25 cm
D	7,0 / 2,0 / 4,0 / 6,0 cm	35 / 10 / 20 / 30 cm

Tabela 1 – Os quatro grupos e os diâmetros de seus elementos seqüenciais em escala 1:5 e equivalente em 1:1

Como a rugosidade das superfícies (suas irregularidades, saliências e reentrâncias) tem impacto na medição do coeficiente de espalhamento, foi estabelecido um critério para avaliação da relação de variação média entre diâmetros adjacentes:

$$\overline{\Delta d} = \frac{\sum_{i=1}^{n-1} |d_{i+1} - d_i|}{n-1} \qquad (5)$$

sendo

d_i = diâmetro do i-ésimo elemento da sequência
n = número total de elementos em uma amostra

Foram definidas quatro sequências, uma primeira pensada em termos de proporções e ritmo de diâmetros como na ordenação de uma série musical, com critérios de equilíbrio e elegância, e as demais simples permutações a partir da mesma. Quando um elemento (A,B,C,D) se encontra invertido, a representação se dá por seu negativo, por exemplo: (A) = 6,0 / 3,0 / 7,0cm ; (-A) = 7,0 / 3,0 / 6,0cm

- AMOSTRA 1: Sem variação $\overline{\Delta d} = 0{,}00cm$
- AMOSTRA 2: A B -C D $\overline{\Delta d} = 2{,}83cm$
- AMOSTRA 3: B A D C $\overline{\Delta d} = 2{,}42cm$
- AMOSTRA 4: A C D B $\overline{\Delta d} = 2{,}83cm$
- AMOSTRA 5: -C A B D $\overline{\Delta d} = 2{,}93cm$

4 MATERIAIS E METODOS DA PROTOTIPAGEM

Foram utilizados o equipamento SLS modelo *Sinterestation HiQ*, *3D System*) e o pó *Duraform PA 12* (*3D System*) não-virgem. Os desenhos das peças – cilindros, semi-cilindros e placas – foram modificados em relação aos desenhos originais para se adequarem às limitações dimensionais da câmara de construção cujas dimensões são: largura: 381mm; comprimento: 330 mm; altura: 457 mm). Elas foram cortadas e encaixes foram adicionados às suas extremidades para posterior montagem do protótipo físico. O volume total das peças foi equivalente aproximadamente ao volume total de três câmaras de construção. O arranjo das peças na câmara de construção (para o caso de preenchimento apenas com as peças do presente projeto). Os valores dos parâmetros de construção são dados na Tabela – 2.

Parâmetros de Processo	valores
Temperatura do leito de construção	177 ºC
Temperatura do cilindro de construção	130 ºC
Temperatura do fundo da câmara de construção	150 ºC
Temperatura dos alimentadores	127 ºC
Potência do *laser*	12,0 watts
Potência de *outline*	6,5 watts

Espessura da camada	0, 10 mm
Pó: Duraform PA 12 (3D System)	não virgem

Tabela 2 - Valores do parâmetros de processo (equipamento *HiQ Sinterstation – 3D System*).

5 MATERIAIS E MÉTODOS DA PROTOTIPAGEM

Os ensaios de determinação do coeficiente de espalhamento foram realizados no Laboratório de Conforto Ambiental da Faculdade de Engenharia Civil Arquitetura e Urbanismo da Unicamp. Um aparato de ensaio para caracterização do coeficiente de espalhamento constituído de uma câmara reverberante em escala 1:5 e equipamentos para aquisição de sinais [9], como visto na Figura 6.

Figura 6 – Câmara reverberante em escala 1:5 construída no Laboratório de Conforto Ambiental da Faculdade de Engenharia Civil Arquitetura e Urbanismo da Unicamp. No detalhe a mesa giratória abaixo da câmara.

Os ensaios foram realizados para quatro posições de microfone e duas posições de fonte num total de oito arranjos entre microfone e fonte sonora. O método utilizado foi de varredura tonal *(sweep)* linear, para uma ênfase nas altas freqüências. O tamanho da varredura foi de 2,73 segundos. A mesa giratória foi utilizada para realização de medições contínuas para a velocidade de 192 segundos para completar uma rotação, o que resulta em uma varredura para cada 5º de giro da mesa. As medições foram realizadas para uma mesa criada com uma placa constituída do mesmo material dos difusores. Cada amostra foi alocada sobre a placa compondo uma área linear de 60x60cm. Este arranjo foi realizado para os elementos semicilíndricos e para os elementos cilíndricos. O método de ensaio da norma ISO 17497:2004 [7] calcula o coeficiente de absorção sonora das amostras para auxiliar na determinação do coeficiente de espalhamento. Foi calculado o coeficiente de absorção sonora para verificar a absorção do material utilizado.

6 RESULTADOS DAS MEDIÇÕES

Para todas as amostras ensaiadas foram realizados cálculos do coeficiente de absorção sonora, tanto para verificar o valor do coeficiente do material empregado na manufatura das amostras, como para verificar se o coeficiente de absorção não excede o valor de 0,4, conforme exigido pela norma ISO 17497:2004.

Os coeficientes de espalhamento resultantes das medições são apresentados a seguir. Pelos motivos expostos acima na seção 3 *(Amostras de ensaio)* a primeira amostra compreende somente medições com elementos semicilindros periódicos, portanto não foram construídos nem foram medidos modelos cilíndricos regulares.

AMOSTRA 1 – Elementos semicilindricos

Figura 7 - Amostra 1: corte transversal. Elementos semicilíndricos regulares, sem variação de diâmetro.

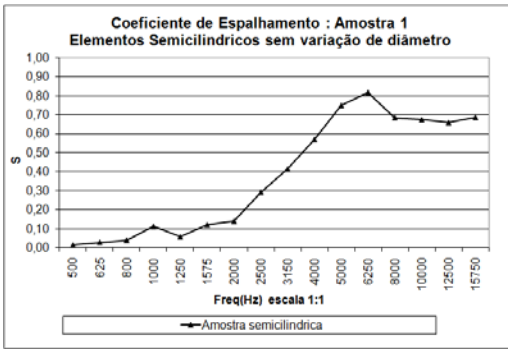


Figura 8 - Amostra 1: coeficientes de espalhamento em função da frequência.

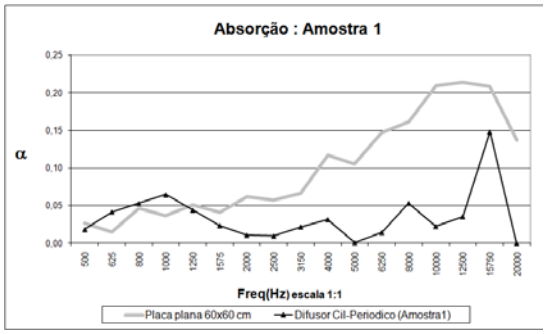


Figura 9 - Amostra 1: coeficientes de absorção em função da freqüência, e comparação com a absorção de um placa plana de 60x60cm com o mesmo material de revestimento.

AMOSTRAS 2 – Elementos semicilíndricos e cilíndricos

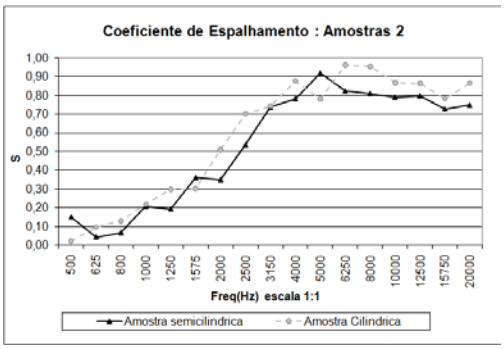


Figura 10 - Amostra 2: coeficientes de espalhamento

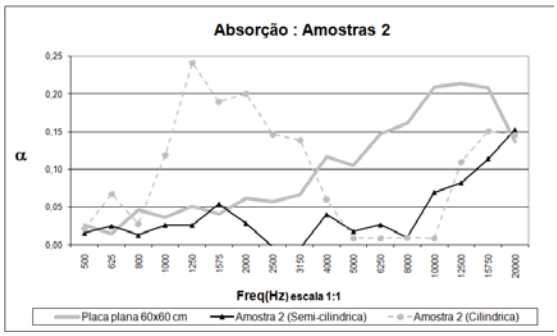


Figura 11 - Amostra 2: absorção - elementos semicilíndricos e cilíndricos x placa plana de 60x60cm.

AMOSTRAS 3 – Elementos semicilíndricos e cilíndricos

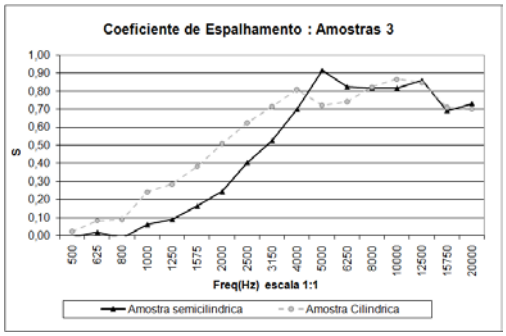


Figura 12 - Amostra 3: coeficientes de espalhamento

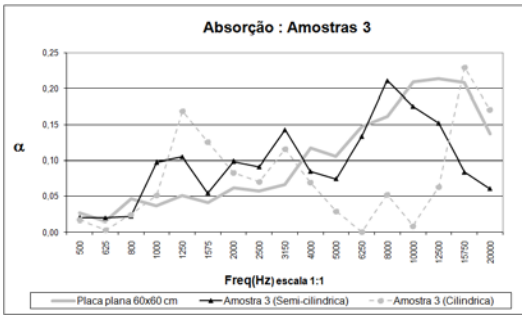


Figura 13 - Amostra 3: absorção - elementos semicilíndricos e cilíndricos x com placa plana de 60x60cm.

AMOSTRAS 4 – Elementos semicilíndricos e cilíndricos

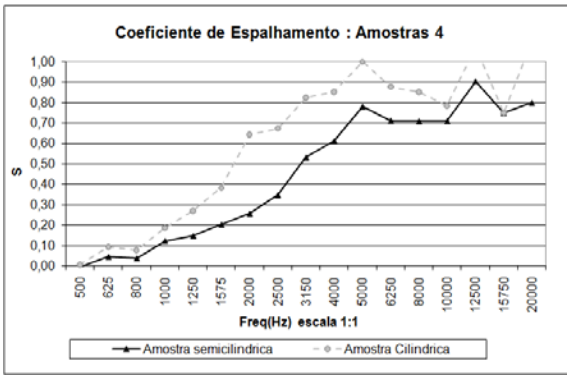


Figura 14 - Amostra 4: coeficientes de espalhamento

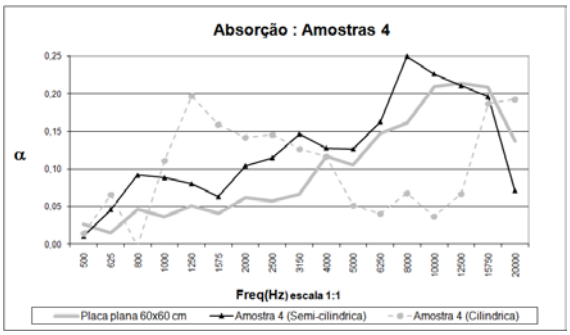


Figura 15 - Amostra 4: absorção - elementos semicilíndricos e cilíndricos x placa plana de 60x60cm.

AMOSTRAS 5 – Elementos semicilíndricos e cilíndricos

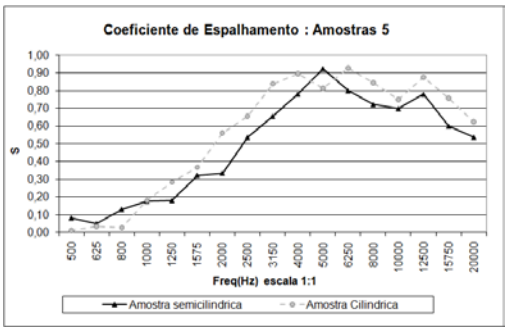


Figura 16 - Amostra 5: coeficientes de espalhamento

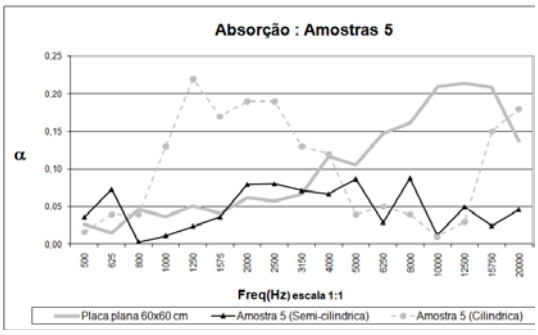


Figura 17 - Amostra 5: absorção - elementos semicilíndricos e cilíndricos x placa plana de 60x60cm.

GRAFICO COMPARATIVO

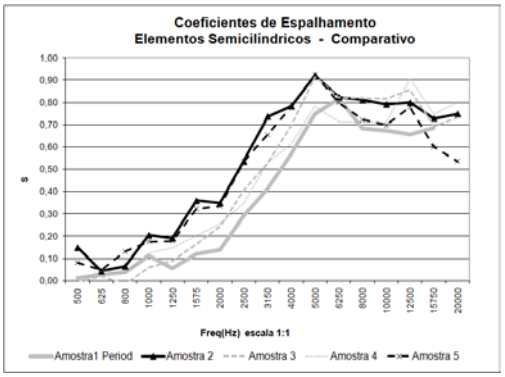


Figura 18 – Gráfico comparativo entre os coeficientes de espalhamento de superfícies difusoras com diferentes arranjos dos mesmos elementos constitutivos.

7 COMENTÁRIOS

Coeficientes de absorção das amostras

Os resultados das medições de coeficiente de absorção apresentaram valores abaixo de 0,4 qualificando o material utilizado para a manufatura dos difusores segundo a norma ISO 17497:2004. Nos cálculos para absorção foram feitas as devidas compensações para a variação de área das superfícies de cada elemento.

Os coeficientes de absorção medidos nas amostras com elementos semicilíndricos seguiram aproximadamente os coeficientes de absorção obtidos para a placa plana de 60x60cm revestida do mesmo material. Já os coeficientes de absorção sonora dos difusores cilíndricos se destacaram por descrever uma característica de ressoador, com coeficientes de absorção acima daqueles obtidos para a placa plana com o mesmo material empregado na manufatura dos difusores e igualmente acima das amostras com elementos semicilíndricos, nas bandas de 200Hz a 500Hz, com uma provável interferência destrutiva, e uma absorção muito mais baixa que as demais amostras de 800 a 2500Hz, portanto com uma provável interferência construtiva. As bandas de 800 a 2500Hz correspondem a comprimentos de onda cujos valores $\lambda/4$ (11 a 3cm) coincidem com os diâmetros dos elementos (2 a 7cm).

Observa-se que os difusores com elementos cilíndricos possuem alturas maiores e diâmetros variáveis havendo a presença de cavidades tanto entre elementos de diâmetros contrastantes, como abaixo da união dos elementos.

Figura 19 - Elementos cilíndricos formando cavidades.

Coeficientes de espalhamento das amostras

Os resultados obtidos indicam que os difusores semicilíndricos e cilíndricos analisados distribuem acima de 500Hz e 400Hz (respectivamente, em escala 1:5) e 2500Hz e 2000Hz (respectivamente, em escala 1:1) 50% da energia refletida fora do ângulo de reflexão especular, chegando a atingir valores superiores a 90%. Observando a Figura 18, em escala 1:1, notam-se resultados acima de 20% da energia incidente a partir de 1000Hz, freqüência com comprimento de onda próximo ao diâmetro do maior elemento (35cm), evoluindo até 5000Hz, comprimento de onda cinco vezes menor (7cm).

Como se esperava, as superfícies com diâmetros variados tem maior eficiência de espalhamento que com diâmetros regulares. As amostras cilíndricas, por sua vez, apresentam um coeficiente de espalhamento maior que as amostras semicilíndricas. Este fenômeno ocorre devido à altura das peças e ao fato das reentrâncias presentes nas amostras compostas por elementos cilíndricos serem maiores que as das amostras semicilíndricas, resultando numa superfície de maior rugosidade.

A organização de elementos cilíndricos poderia também estar funcionando como um conjunto de elementos estruturados espacialmente, constituindo um contexto semelhante ao estudado por Vorlaender [10][3], sendo aqui caracterizado por um alinhamento sequencial de cilindros repousados numa superfície plana. Isso seria coerente com fenômenos de interferência devidos a espaços entre elementos subsequentes, considerando a distância entre elementos e as alturas dos mesmos.

O melhor resultado obtido de espalhamento é o da Amostra 2, possuindo ordenação tal que os elementos somente se repetem depois de esgotada a exposição da totalidade do conjunto dos mesmos, um principio semelhante à organização estrutural serial musical.

Entre os quatro grupos semelhantes (1,2,3 e 4) houve variação de 20% na performance de espalhamento.

Influência do padrão ou da forma de organização dos elementos

Em principio, considerando a irregularidade da superfície proporcional ao espalhamento, os maiores valores de variação média de diâmetros *(Δd)* deveriam corresponder a coeficientes de espalhamento mais elevados. Porém é observado que os maiores valores são os obtidos para as Amostras 2 e 5. Apesar de apresentarem variação média de diâmetros significativa, deveriam ter padrões parecidos aos resultados obtidos para as amostras 4 e 2, justamente por apresentarem a mesma variação média.

Amostras apresentando uma rugosidade mais semelhante (AMOSTRA 2: A B -C D $\overline{\Delta d} = 2{,}83cm$ e AMOSTRA 4: A C D B $\overline{\Delta d} = 2{,}83cm$) deram resultados mais distantes do que outras Amostras com $\overline{\Delta d}$ diferentes, indicando que somente a variação média da rugosidade não bastaria para definir totalmente a eficiência de espalhamento de uma superfície.

8 CONCLUSÕES

Pudemos confirmar aqui uma hipótese da pesquisa com superfícies seriais aplicada a difusores sonoros: segundo os dados obtidos, a absorção das superfícies com elementos semicilíndricos apresentou apenas a absorção do material de revestimento, ou seja, o desenho geométrico da superfície e as interferências das reflexões não acrescentaram especialmente absorções, sobretudo nos casos das Amostras 2 e 5.

Considere-se ainda que as superfícies com elementos semicilindricos e cilíndricos, por suas saliências e reentrâncias, têm uma superfície revestida maior do que a placa plana, por ex., uma superfície com elementos semicilindricos regulares (diâmetro 6cm; S = 5655cm^2) é 57% maior do que a placa plana (3600cm^2)) e, portanto, deveriam, em principio, apresentar naturalmente uma absorção maior.

Constatamos também que, aos casos em que houve os melhores resultados de espalhamento de superfícies com elementos semicilindricos, correspondem os melhores (e menores) resultados de absorção.

Se entre as amostras analisadas a única diferença era somente o arranjo entre os elementos, a diferença de espalhamento observada nos faz concluir que rugosidades semelhantes podem ter resultados distintos. O coeficiente de espalhamento depende também de qualidades derivadas da ordenação dos elementos da superfície e não apenas da sua rugosidade média. Em resumo, não é somente a quantidade de rugosidade que importa, mas também a maneira como a rugosidade evolui no espaço e, ainda, se faz uso de alguma sistematização geradora.

O valor obtido de 20% de variação de desempenho de espalhamento somente pela permutação dos elementos é um

[3] (p. 311)

valor significativo. Experimentos ulteriores poderão aprofundar este estudo, buscando correlações entre os resultados de espalhamento e parâmetros dos elementos como dimensões, qualidades de ordenação, proporções, numero de repetições, formas de superfície.

Poderiam ser verificados em outros experimentos:

(1) Influência da composição do seqüenciamento (no processo de elaboração de amostras, o seqüenciamento de elementos na Amostra 2 foi concebido com o cuidado próprio de uma organização musical, enquanto que as demais foram permutações de seus grupos de elementos);

(2) Amostras com os elementos ordenados de maneira similar teriam resultados aproximados? (As Amostras 2 (A B -C D) e 5 (-C A B D) apresentam os elementos com a mesma orientação, ambas com elemento C invertido (-C) e as Amostras 3 (B A D C) e 4 (A C D B) todos elementos sem nenhuma inversão: A B C D).

(3) A presença de articulação marcada claramente repetida na sequência dos diâmetros estaria influenciando no aumento da eficiência de espalhamento?

Quanto ao terceiro item, a presença de repetições na organização do sequenciamento poderia ser benéfica para o desempenho do difusor, assim como é de fato nos Difusores de Schroeder, nos quais para que tenham maior espalhamento é necessário que o sequenciamento de cavidades seja repetido varias vezes, sendo a distribuição equilibrada de energia pelos lóbulos de difração proporcional à periodicidade da superfície [11][4].

Uma forte coerência organizacional[5] poderia ter influência benéfica no desempenho de espalhamento de um difusor. Esta proposta tem como base a melhoria efetivamente constatada em um difusor de Schroeder quando este apresenta diversas repetições de seu sequenciamento[6] e, para melhorar seu desempenho, se aplica uma sequência binária de dispersão pseudo-randômica (ou então a sequência de Barker (igualmente binária)) controlando em que elementos do difusor haverá ou não a inversão de um período inteiro de cavidades [12][13][14], o que resulta numa resposta mais homogênea do difusor.

Trabalhos ulteriores investigarão o efetivo impacto da qualidade da ordenação dos elementos na eficiência de um difusor SCSS, confrontando os resultados com o comportamento descrito e novas hipóteses levantadas. Uma das possibilidades seria produzir a alteração aleatória do seqüenciamento dos elementos da amostra e comparar os resultados com alterações produzidas através de complexos algoritmos para investigar a formação de diferentes rugosidades das amostras e quando se evidencia a formação ou não de algum padrão de seqüenciamento ou efeitos de repetição, simetria ou permutação influenciando a rugosidade.

BIBLIOGRAFIA

[1] MANNIS, J. A. *Difusores sonoros projetados a partir de processo serial*: adequação acústica de pequenas salas à performance e audição musical. 2008. 424 p. Tese (Doutorado). Universidade Estadual de Campinas (UNICAMP), 2008.

[2] VOLPATO N. (editor) *Prototipagem rápida: tecnologias e aplicações,* Blücher, 2007, 1ed., ISBN 85-212-0388-8, São Paulo, 244 p.

[3] OLIVEIRA, M. F., *Aplicações de prototipagem rápida em projetos de pesquisa*. Campinas, 2008, 130p. Tese de Mestrado – Faculdade de Engenharia Mecânica - Universidade Estadual de Campinas

[4] COX, T. J., D'ANTONIO, P. *Acoustic absorbers and diffusers*: theory, design and application. Spon Press 2004. ISBN 0-415-29649-8.

[5] SCHROEDER, M. R. Binaural dissimilarity and optimum ceilings for concert halls: More lateral sound diffusion. In: *The Journal of Acoustical Society of America*, v.65, n.4 Abr. 1979.

[6] VORLÄNDER, M., MOMMERTZ, E., Definition and measurement of random-incidence scattering coefficients In: *Applied Acoustics*. 60, 187 199 (2000)

[7] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION: 17497-1, *Acoustics - Sound-scattering properties of surfaces – Part 1*: Measurement of the random-incidence scattering coefficient in a reverberation room. 2004.

[8] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION: 3382, *Acoustics-Measurement of room acoustic parameters – Part 1*: Performance spaces. 2009.

[9] SANTOS, C., BERTOLI, S. R.; Construção de câmara reverberante em escala para realização de ensaios acústicos. *Congreso Iberoamericano de Acústica*. Proceedings - FIA 2008. Nov. 2008.

[10] VORLÄNDER, M. *Auralization: fundamentals of acoustics, modeling, simulation, algorithms and acoustic virtual reality*. 1st ed. Berlin: Springer, 2008. 335 p.

[11] COX, T. J.; D'ANTONIO, P. Schroeder diffusers: a review. In: *Building Acoustics*, v.10, n.1, 2003. p. 1-32.

[12] ANGUS, James A. S. Using modulated phase reflection gratings to achieve specific diffusion characteristics. In: *Convention Audio Engineering Society*, 99.,1995, New York. Preprint 4117, 1995.

[13] ANGUS, James A. S. Wideband two dimensional diffusers using orthogonal modulated sequences. In: *Convention Audio Engineering Society*, 103., 1997, New York. Preprint 4640, 1997.

[14] ANGUS, James A. S. Using grating modulation to achieve wideband large area diffusers. In: *Applied Acoustics*, v.60, n.2, p.143-165, June, 2000.

[4] (p.8)

[5] Buscar um máximo de variação com um mínimo de elementos, aplicando simetrias e mantendo equilíbrio na distribuição de elementos e na atribuição de parâmetros.

[6] um grande número de períodos das cavidades dos Difusores de Schroeder fará com que os lóbulos se tornem muito estreitos, o que leva a um espalhamento não equilibrado devido ao grande número de nulos. A aplicação de sequências binárias busca homogeneizar o nível entre os lóbulos.

Sociedade de Engenharia de Áudio
Artigo de Congresso

Apresentado no 9º Congresso de Engenharia de Áudio
15ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2011, São Paulo, SP

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

Um Estudo Comparativo de Equalizadores de Áudio Gráficos Analógicos

André Luís Dalcastagnê[1] e Sidnei Noceti Filho[2]

[1] Departamento Acadêmico de Eletrônica, Instituto Federal de Santa Catarina
Campus Florianópolis, Florianópolis, SC, 88020-300, Brasil

[2] LINSE - Laboratório de Circuitos e Processamento de Sinais
Departamento de Engenharia Elétrica, Universidade Federal de Santa Catarina
Florianópolis, SC, 88040-900, Brasil

andreld@ifsc.edu.br, sidnei@linse.ufsc.br

RESUMO

Este trabalho apresenta uma análise comparativa de três estruturas comumente usadas na implementação de equalizadores de áudio gráficos analógicos: equalizador série, equalizador paralelo e o aqui chamado equalizador Gundry. Primeiramente, estuda-se o efeito que desvios estatísticos nos valores dos elementos causam nas magnitudes das respostas em freqüência desses equalizadores. Além disso, avalia-se a capacidade dos equalizadores estudados de atenuar ou amplificar faixas de freqüência adjacentes. Os resultados obtidos mostram a superioridade do equalizador Gundry sobre as estruturas série e paralela, tanto em termos de sensibilidade à variação dos valores dos elementos quanto no desempenho ao atenuar ou amplificar freqüências adjacentes.

0 INTRODUÇÃO

Os equalizadores de áudio surgiram na década de 1930 com o advento do cinema falado. O pioneiro da área foi John Volkman, que buscou equalizar a resposta de uma sala de cinema com péssima qualidade de reprodução. Com o passar dos anos, os equalizadores de áudio foram sendo aprimorados e o seu uso difundido em diversas aplicações, tais como: correção da resposta em freqüência de ambientes, ajuste do som ao gosto do ouvinte, em próteses auditivas e em circuitos RIAA, usados tanto na gravação quanto na reprodução de discos de vinil.

Apesar de os equalizadores de áudio também estarem disponíveis atualmente na forma digital, os equalizadores analógicos, foco deste trabalho, continuam a ter importância, com algumas vantagens

sobre os digitais. Por exemplo, equalizadores analógicos dispensam o uso de conversores analógico-digital e digital-analógico e podem apresentar menor custo caso não se disponha de um processador de sinais digitais. Além disso, o equalizador analógico é mais acessível ao usuário leigo ou que pratica eletrônica por *hobby*, que teria uma maior dificuldade tanto para implementar quanto para compreender o projeto de um equalizador digital.

Diversos fatores contribuem para que um circuito eletrônico apresente desempenho diferente do desejado, tais como tolerâncias dos componentes; uso de valores comerciais no lugar dos valores de projeto; condições ambientais; não idealidades dos componentes, tais como elementos parasitas, fatores de qualidade finitos dos elementos reativos e produtos ganho-banda (GBs) finitos dos amplificadores operacionais. Sendo assim, o projetista possui duas alternativas se desejar obter um circuito eletrônico com comportamento próximo do ideal. A primeira delas é usar componentes de baixa tolerância e com não idealidades desprezáveis, conseqüentemente componentes de maior custo. A segunda opção é usar circuitos com baixa sensibilidade às mudanças nos valores dos seus elementos. Em tais circuitos, são impostas menos restrições às variações dos componentes e das outras não idealidades, o que geralmente resulta em um menor custo de fabricação. Por isso, a sensibilidade tem sido um dos mais importantes critérios usados para comparar diferentes realizações de uma mesma função.

O objetivo deste trabalho é apresentar um estudo comparativo de três tipos de equalizadores gráficos analógicos comumente usados na prática: equalizador série, equalizador paralelo e o equalizador proposto em [1, 2], aqui chamado equalizador Gundry. São analisados dois aspectos principais acerca das estruturas: i) a sensibilidade dos equalizadores à variação dos valores dos seus elementos; ii) o comportamento das estruturas ao se exigir ganhos ou atenuações em faixas de freqüência adjacentes.

O restante deste artigo está organizado da forma descrita a seguir. A Seção 1 apresenta definições relacionadas com a função sensibilidade adotadas neste trabalho. A Seção 2 detalha os equalizadores de áudio aqui estudados. Os resultados obtidos são mostrados na Seção 3. Finalmente, a Seção 4 apresenta as conclusões deste trabalho.

1 SENSIBILIDADE

Como mencionado anteriormente, a função sensibilidade possibilita avaliar como variam certas características de uma rede quando um ou mais dos seus parâmetros são alterados. Logo, ela permite prever os desvios estatísticos e/ou determinísticos das funções de rede, tais como os desvios na magnitude da resposta em freqüência de um equalizador.

Considere uma função y dependente da variável x. De acordo com [3], a sensibilidade de y em relação a x é definida por

$$S_x^y \triangleq \frac{\partial y/y}{\partial x/x} = \frac{x}{y}\frac{\partial y}{\partial x} = \frac{\partial \ln y}{\partial \ln x}\,. \tag{1}$$

Nesse caso, a função sensibilidade relaciona a variação percentual de y com a variação percentual de x, a chamada sensibilidade normalizada.

A variação de um elemento qualquer X de uma rede é freqüentemente expressa pela variabilidade V_{X}, definida em função dos valores nominal x_0 e modificado x_1 do componente, como segue

$$V_{\mathrm{X}} = \frac{\Delta x}{x_0} = \frac{x_1 - x_0}{x_0}\,. \tag{2}$$

A variabilidade pode ser uma variável determinística ou estocástica. Um caso típico de variação determinística se dá quando os valores de projeto dos componentes são substituídos por valores comerciais. Apesar disso, a variabilidade é geralmente uma variável estocástica, com função densidade de probabilidade caracterizada pelo fabricante que, na maioria dos casos, apresenta uma distribuição próxima de uma normal (gaussiana), sendo caracterizada pela média $\mu(V_{\mathrm{X}})$ e pela variância $\sigma^2(V_{\mathrm{X}})$ [4]. Um caso típico de variabilidade estocástica é a tolerância de fabricação de componentes discretos. Considere, por exemplo, a fabricação de um lote de resistores de valor nominal R_0. O valor de cada resistor pode ser expresso por $R = R_0(1 + V_{\mathrm{R}})$, com a variabilidade V_{R} apresentando $\mu(V_{\mathrm{R}}) = 0$ e $\sigma^2(V_{\mathrm{R}}) = 1{,}11 \times 10^{-5}$. Nesse caso, $3\sigma \cong 0{,}01$ e o fabricante pode especificar que 99,74 % desses resistores estão na faixa de tolerância de ± 1 % [4].

Finalmente, pode-se definir como se relaciona o desvio na magnitude da resposta em freqüência de uma rede com as funções sensibilidade e com as variabilidades dos seus elementos. Considere que as variabilidades dos N elementos de um equalizador com resposta em freqüência $T(\mathrm{j}\omega)$ são variáveis estocásticas independentes. Como deduzido em [3], o desvio estatístico em decibéis [correspondente a $3\sigma(\Delta|T(\mathrm{j}\omega)|_{\mathrm{dB}})$, ou seja, 99,74 % de probabilidade de ocorrência] causado na magnitude da resposta em freqüência do equalizador é dado por

$$\Delta = 8{,}685\left\{\sum_{i=1}^{N}\left[\Re e\left(S_{\mathrm{X}_i}^{T(\mathrm{j}\omega)} 3\sigma(V_{\mathrm{X}_i})\right)\right]^2\right\}^{1/2}. \tag{3}$$

Logo, os desvios, em decibéis, da magnitude da resposta em freqüência de um equalizador, e de qualquer outro circuito eletrônico, são dependentes das N funções sensibilidade da rede e das variabilidades dos N componentes dessa rede. Por isso, os desvios podem ser reduzidos utilizando-se redes

de baixa sensibilidade e/ou componentes com baixa tolerância. Além disso, a expressão (3) mostra que quanto maior o número de elementos de um equalizador, maior é o desvio estatístico da sua resposta.

2 EQUALIZADORES ESTUDADOS

2.1 Freqüências Centrais

Um equalizador gráfico é composto por n seções, cada uma delas apresentando uma freqüência central f_0, de valor fixo, cujo ganho pode ser ajustado em uma faixa de, por exemplo, -12 a $+12$ dB, de acordo com a necessidade do usuário. Muitos equalizadores gráficos utilizam o padrão ISO de freqüências centrais, em que o espectro é dividido em freqüências relacionadas por um fator multiplicativo k. Para um equalizador de 31 bandas, foco deste trabalho, também chamado de 1/3 de oitava, em cada oitava deve haver três freqüências centrais. Assim, se entre uma determinada freqüência f_0 e uma freqüência uma oitava acima ($2f_0$) devem existir três freqüências centrais, pode-se deduzir que

$$k^3 f_0 = 2f_0 \Rightarrow k = 2^{1/3} \cong 1{,}2599\,. \tag{4}$$

A Tabela 1 apresenta o conjunto de freqüências centrais (f_0) para um equalizador de 31 bandas, assumindo 1 kHz como freqüência de referência (f_r) e $k = 2^{1/3}$.

Além da freqüência central, cada seção de um equalizador gráfico deve possuir um determinado fator de qualidade Q. A fim de deduzi-lo, considere a Figura 1, que demonstra parte do espectro de freqüências de um equalizador gráfico genérico com função de transferência $T(s)$. A freqüência central f_0 está relacionada com as freqüências de corte f_i e f_s (freqüências onde o ganho decai 3 dB) por

$$f_0^2 = f_\mathrm{i} f_\mathrm{s}\,. \tag{5}$$

Fazendo $f_\mathrm{i} = f_\mathrm{x}$ e $f_\mathrm{s} = kf_\mathrm{x}$, obtém-se

$$f_0^2 = f_\mathrm{x} k f_\mathrm{x} = k f_\mathrm{x}^2 \tag{6}$$

de onde se determina que

$$f_0 = \sqrt{k} f_\mathrm{x}\,. \tag{7}$$

Além disso, a largura de banda, em hertz, é

$$B = kf_\mathrm{x} - f_\mathrm{x} = (k-1)f_\mathrm{x}\,. \tag{8}$$

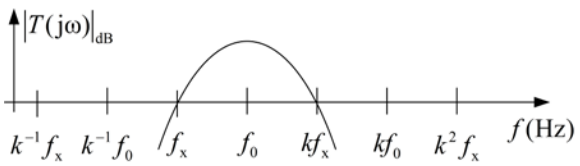


Figura 1: Espectro de freqüências de um equalizador gráfico genérico.

Tabela 1: Freqüências centrais de um equalizador gráfico de 31 bandas, adotando $k = 2^{1/3}$ e $f_\mathrm{r} = 1$ kHz como freqüência de referência

Seção	Freq. central f_0	Seção	Freq. central f_0
1	$k^{-17} f_\mathrm{r} = 19{,}68$ Hz	17	$k^{-1} f_\mathrm{r} = 793{,}7$ Hz
2	$k^{-16} f_\mathrm{r} = 24{,}80$ Hz	18	$f_\mathrm{r} = 1$ kHz
3	$k^{-15} f_\mathrm{r} = 31{,}25$ Hz	19	$kf_\mathrm{r} = 1{,}259$ kHz
4	$k^{-14} f_\mathrm{r} = 39{,}37$ Hz	20	$k^2 f_\mathrm{r} = 1{,}587$ kHz
5	$k^{-13} f_\mathrm{r} = 49{,}60$ Hz	21	$k^3 f_\mathrm{r} = 2$ kHz
6	$k^{-12} f_\mathrm{r} = 62{,}50$ Hz	22	$k^4 f_\mathrm{r} = 2{,}519$ kHz
7	$k^{-11} f_\mathrm{r} = 78{,}74$ Hz	23	$k^5 f_\mathrm{r} = 3{,}174$ kHz
8	$k^{-10} f_\mathrm{r} = 99{,}21$ Hz	24	$k^6 f_\mathrm{r} = 4$ kHz
9	$k^{-9} f_\mathrm{r} = 125$ Hz	25	$k^7 f_\mathrm{r} = 5{,}039$ kHz
10	$k^{-8} f_\mathrm{r} = 157{,}4$ Hz	26	$k^8 f_\mathrm{r} = 6{,}349$ kHz
11	$k^{-7} f_\mathrm{r} = 198{,}4$ Hz	27	$k^9 f_\mathrm{r} = 8$ kHz
12	$k^{-6} f_\mathrm{r} = 250$ Hz	28	$k^{10} f_\mathrm{r} = 10{,}07$ kHz
13	$k^{-5} f_\mathrm{r} = 314{,}9$ Hz	29	$k^{11} f_\mathrm{r} = 12{,}69$ kHz
14	$k^{-4} f_\mathrm{r} = 396{,}8$ Hz	30	$k^{12} f_\mathrm{r} = 16$ kHz
15	$k^{-3} f_\mathrm{r} = 500$ Hz	31	$k^{13} f_\mathrm{r} = 20{,}15$ kHz
16	$k^{-2} f_\mathrm{r} = 629{,}9$ Hz		

Por definição, o fator de qualidade é

$$Q = f_0/B\,. \tag{9}$$

Substituindo (7) e (8) em (9), pode-se deduzir

$$Q = \frac{\sqrt{k}}{k-1}\,. \tag{10}$$

Logo, o fator de qualidade de cada seção do equalizador de 1/3 de oitava ($k = 2^{1/3}$) é $Q \cong 4{,}318$.

2.2 Função de Transferência

A função de transferência de cada seção de um equalizador gráfico é do tipo *bump*, definida como

$$T(s) = \frac{s^2 + (\omega_0/Q_\mathrm{z})s + \omega_0^2}{s^2 + (\omega_0/Q_\mathrm{p})s + \omega_0^2} \tag{11}$$

onde $\omega_0 = 2\pi f_0$ é a freqüência central da seção do equalizador, em rad/s, Q_z é o fator de qualidade dos zeros e Q_p é o fator de qualidade dos pólos. Como a freqüência central é fixa, controlam-se apenas os fatores Q_z e Q_p. Assim, há três possíveis situações: i) se $Q_\mathrm{z} > Q_\mathrm{p}$, obtém-se uma atenuação (*cut*) em torno da freqüência f_0; ii) se $Q_\mathrm{p} > Q_\mathrm{z}$, obtém-se uma amplificação (*boost*) em torno da freqüência f_0; iii) se $Q_\mathrm{z} = Q_\mathrm{p}$, obtém-se uma resposta plana.

Uma forma simples de obter uma função de transferência do tipo *bump* é utilizar um inversor, um amplificador de ganho K e um filtro passa-faixa com função de transferência

$$T_\mathrm{PF}(s) = \frac{(\omega_0/Q)s}{s^2 + (\omega_0/Q)s + \omega_0^2} \tag{12}$$

onde $\omega_0 = 2\pi f_0$ é a freqüência central, em rad/s, e Q é o fator de qualidade dos pólos. Para obter a

função de transferência (11) através de (12), deve-se conectar tais estruturas da forma mostrada na Figura 2, obtendo-se assim a seguinte relação:

$$T(s) = 1 - T_{\mathrm{PF}}(s) + KT_{\mathrm{PF}}(s)\,. \qquad (13)$$

Substituindo (12) em (13) e fazendo as devidas manipulações matemáticas, obtém-se a função de transferência do equalizador, dada por

$$T(s) = \frac{s^2 + K(\omega_0/Q)s + \omega_0^2}{s^2 + (\omega_0/Q)s + \omega_0^2}\,. \qquad (14)$$

Fazendo $Q = Q_{\mathrm{p}}$ e $K/Q = 1/Q_{\mathrm{z}}$ em (14), confirma-se que tal função de transferência é equivalente a função *bump* desejada, dada por (11).

A Figura 3 ilustra a resposta de uma seção da Figura 2, com $f_0 = 1\,\mathrm{kHz}$ e $Q = 4{,}318$, tanto para uma amplificação de 12 dB ($K = 3{,}9811$) quanto para uma atenuação de 12 dB ($K = 0{,}2511$). Pode-se observar que a resposta é assimétrica: para níveis iguais de amplificação e atenuação na freqüência f_0, as curvas apresentam formas diferentes.

2.3 Rede Passa-Faixa

Neste trabalho, a função de transferência dada por (12) é implementada pela rede passa-faixa mostrada na Figura 4. Essa é uma estrutura *multiple feedback* (MFB), inversora, cuja análise pode ser encontrada em [3]. De forma simplificada, o projeto dessa rede consiste em estipular um valor comercial C para os capacitores e utilizar as seguintes expressões para o cálculo dos valores dos resistores:

$$R_{1\mathrm{S}} = \frac{Q}{2\pi C f_0} \qquad (15)$$

$$R_{1\mathrm{P}} = \frac{1}{2\pi C f_0 \left(2Q - \frac{1}{Q}\right)} \qquad (16)$$

$$R_2 = 2R_{1\mathrm{S}}\,. \qquad (17)$$

Para um equalizador de 31 bandas série ou paralelo, é necessário projetar um filtro passa-faixa com fator de qualidade $Q = 4{,}318$ para cada uma das 31 freqüências centrais f_0 mostradas na Tabela 1.

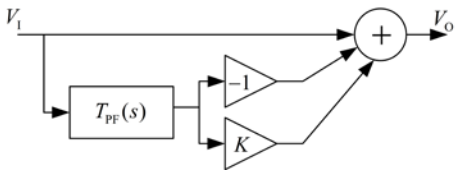


Figura 2: Diagrama de blocos da função de transferência *bump*.

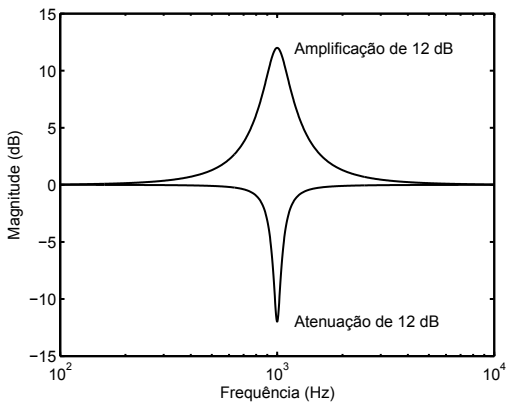


Figura 3: Exemplo de resposta em freqüência de uma função *bump*, com $f_0 = 1\,\mathrm{kHz}$ e $Q = 4{,}318$.

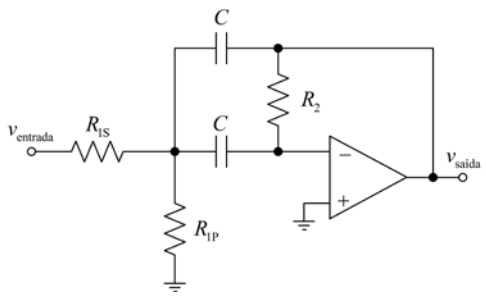


Figura 4: Filtro passa-faixa MFB.

2.4 Equalizadores série e paralelo

De maneira geral, um equalizador série ou paralelo de n bandas é composto por n seções iguais às da Figura 2. Os diagramas de blocos dessas duas topologias são mostrados na Figura 5.

De acordo com a Figura 5(a), a função de transferência do equalizador série é

$$T(s) = T_1(s)T_2(s)\cdots T_n(s)\,. \qquad (18)$$

Assim, basta conectar em série blocos iguais aos da Figura 2 para realizar um equalizador de n bandas, como mostra a Figura 6.

Para o equalizador paralelo, pode-se deduzir através da Figura 5(b) que a sua função de transferência é dada por

$$T(s) = 1 - n + \sum_{i=1}^{n} T_i(s)\,. \qquad (19)$$

Se cada função de transferência $T_i(s)$ for implementada da forma mostrada na Figura 2, cuja função de transferência é dada por (13), deduz-se que

$$T(s) = 1 - n + \sum_{i=1}^{n} 1 - T_{\mathrm{PF}i}(s) + K_i T_{\mathrm{PF}i}(s)\,. \qquad (20)$$

Desenvolvendo (20), chega-se a

$$T(s) = 1 + \sum_{i=1}^{n} \left[K_i - 1\right] T_{\mathrm{PF}i}(s)\,. \qquad (21)$$

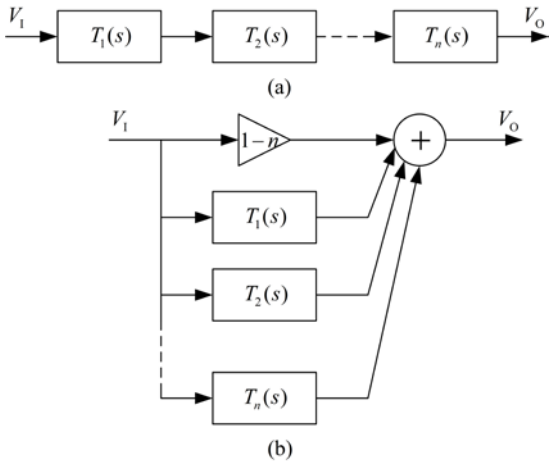


Figura 5: Diagramas de blocos de equalizadores. (a) Equalizador série. (b) Equalizador paralelo.

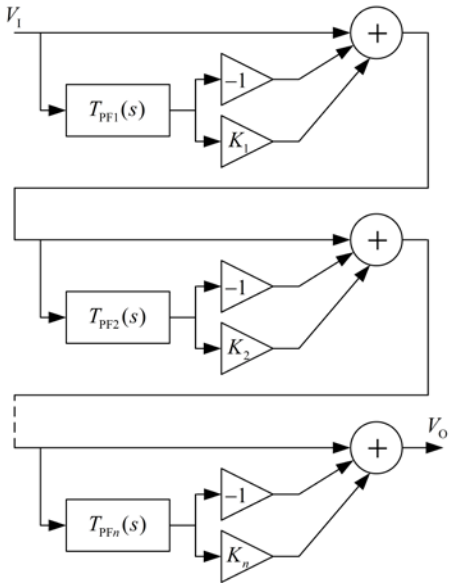


Figura 6: Diagrama de blocos do equalizador série.

Tal função pode ser representada pelo diagrama de blocos da Figura 7. Em comparação com a estrutura série, o equalizador paralelo apresenta menor complexidade, uma vez que todas as n seções compartilham o mesmo somador e o mesmo inversor.

2.5 Equalizador Gundry

A terceira estrutura estudada é o aqui chamado equalizador Gundry, cujo diagrama de blocos é apresentado na Figura 8. Com a chave S na posição amplificação, a função de transferência é

$$T_{\mathrm{G}}(s) = 1 + KT_{\mathrm{PF}}(s)\,. \tag{22}$$

Já quando a chave S é deslocada para a posição atenuação, a função de transferência passa a ser

$$T_{\mathrm{A}}(s) = \frac{1}{1 + KT_{\mathrm{PF}}(s)}\,. \tag{23}$$

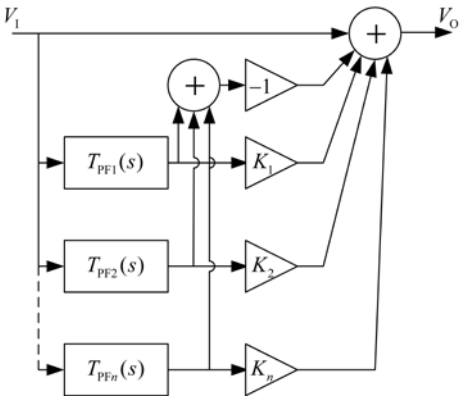


Figura 7: Diagrama de blocos do equalizador paralelo.

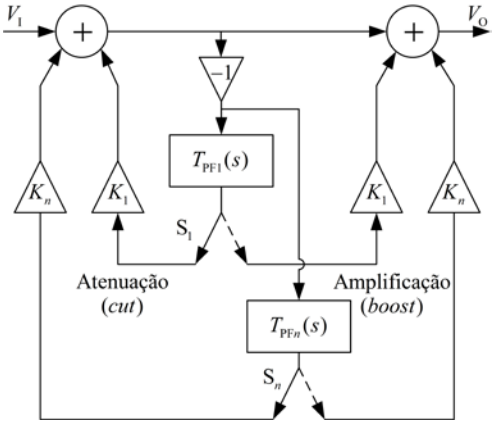


Figura 8: Diagrama de blocos do equalizador Gundry.

Substituindo (12) em (22) e (23), define-se a função de transferência do equalizador para as duas possíveis posições da chave S:

$$T_{\mathrm{G}}(s) = \frac{s^2 + (1+K)\,Bs + \omega_0^2}{s^2 + Bs + \omega_0^2}\,. \tag{24}$$

$$T_{\mathrm{A}}(s) = \frac{s^2 + Bs + \omega_0^2}{s^2 + (1+K)\,Bs + \omega_0^2}\,. \tag{25}$$

onde $B = \omega_0/Q$ é a largura de banda da seção.

O equalizador Gundry apresenta uma vantagem sobre as estruturas série e paralela. Nessas últimas, para se obter uma amplificação/atenuação na freqüência central f_0, deve-se alterar o parâmetro K da função de transferência (14), ou seja, altera-se o fator de qualidade do zero para ambos os casos. Como resultado, o circuito apresenta a resposta assimétrica da Figura 3. No caso do equalizador Gundry, para se obter uma amplificação, altera-se o parâmetro K da função de transferência (24). Já para uma atenuação, altera-se o parâmetro K da função de transferência (25). Pelo fato de serem funções inversas, para o mesmo valor de

K, (24) e (25) apresentam respostas em freqüência de mesma magnitude em decibéis, porém com sinal trocado. Logo, o equalizador Gundry possui uma resposta simétrica, como a mostrada na Figura 9 para $f_0 = 1\,\text{kHz}$, $Q = 4{,}318$ e $K = 2{,}9811$ (amplificação/atenuação de 12 dB).

Uma maneira de implementar cada seção do diagrama da Figura 8 é usar dois somadores inversores e um potenciômetro com *tap* central aterrado (*grounded center tap*), estrutura mostrada na Figura 10. Para implementar equalizadores de maior ordem, basta acrescentar blocos $T_{\text{PF}}(s)$ em paralelo, com todas as entradas ligadas ao mesmo inversor (bloco -1) e com cada saída ligada a um potenciômetro com *tap* central aterrado, cujos terminais devem ser conectados aos somadores.

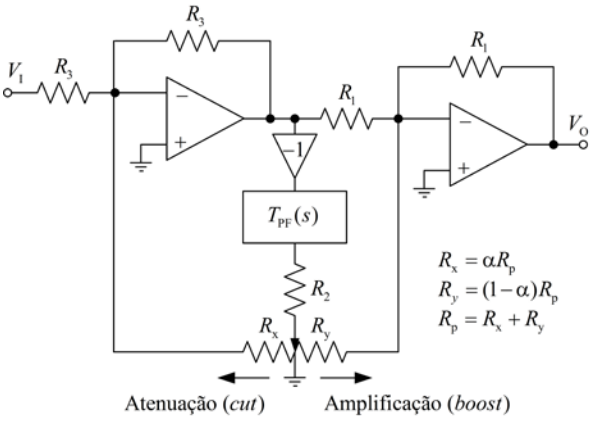


Figura 10: Equalizador Gundry de 1 banda.

3 RESULTADOS OBTIDOS

Esta seção apresenta as análises realizadas com os equalizadores gráficos estudados. As simulações são realizadas com o *software* SG2 (disponível para *download* em http://www.linse.ufsc.br/∼sidnei). Tal programa é usado para obter as curvas nominais de magnitude da resposta em freqüência dos equalizadores e também os desvios estatísticos dessas magnitudes [calculados através de (3)].

Inicialmente, deseja-se estudar os desvios, em decibéis, da magnitude da resposta em freqüência de cada equalizador. Para isso, consideram-se componentes passivos com tolerância de $\pm 1\,\%$ e amplificadores operacionais com GB de 4,5 MHz e resistência de saída de $700\,\Omega$. Para fins de comparação, define-se a variável desvio máximo ΔM, que representa o máximo desvio estatístico da resposta do equalizador em relação à nominal. Como existem inúmeras possibilidades de ganho em cada freqüência f_0, estipulou-se que todos os ganhos são de 0 dB (*flat*). Essa é a condição que resulta no menor desvio máximo, mas é a que permite determinar com mais clareza esse parâmetro. Para outros valores de ganho, simulações realizadas mostram que os resultados são análogos aos apresentados, apenas acentuando ainda mais as diferenças entre os desvios dos equalizadores. Por simplicidade e sem perda de generalidade, são simulados equalizadores de uma, três, seis, nove e doze bandas.

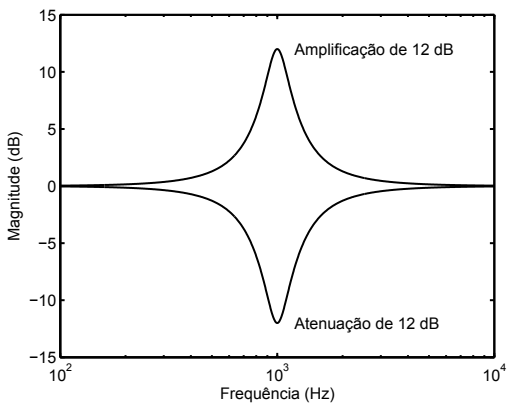


Figura 9: Exemplo de resposta em freqüência de uma seção ilustrada na Figura 8.

Considere um equalizador de uma banda com $f_0 = 1\,\text{kHz}$ (seção 18 da Tabela 1). A Figura 11 apresenta as respostas dos três equalizadores (curvas centrais), juntamente com os limites superior e inferior que cada uma pode atingir. Nota-se que os equalizadores série e paralelo apresentam um desvio em relação à resposta ideal menor do que o obtido pelo equalizador Gundry para a maior parte das freqüências. Entretanto, na freqüência $f_0 = 1\,\text{kHz}$, eles apresentam desvios maiores do que o alcançado pelo equalizador Gundry ($\Delta\text{M}_1 = \Delta\text{M}_2 = 0{,}2128\,\text{dB}$ e $\Delta\text{M}_3 = 0{,}1737\,\text{dB}$).

A Figura 12 apresenta respostas similares às da Figura 11, para equalizadores de três bandas (seções 17 a 19 da Tabela 1). O desvio máximo do equalizador série aumentou razoavelmente, de 0,2128 para 0,279 dB. Para o equalizador paralelo, o aumento no desvio foi bem menor, de 0,2128 para 0,2183 dB. Já para o equalizador Gundry, o desvio máximo permanece no valor de 0,1737 dB.

A Figura 13 apresenta as respostas de equalizadores de seis bandas (seções 16 a 21 da Tabela 1). Novamente, o desvio máximo do equalizador série aumentou substancialmente, de 0,279 para 0,352 dB, ao passo que o equalizador paralelo apresentou um leve aumento no desvio, de 0,2183 para 0,2194 dB. Já o desvio máximo do equalizador Gundry permaneceu inalterado.

As Figuras 14 e 15 apresentam respostas em freqüência similares às anteriores, para equalizadores de nove (seções 16 a 21 e 26 a 28 e da Tabela 1) e doze bandas (seções 16 a 21 e 26 a 31 da Tabela 1), respectivamente. Assim como nos casos anteriores, o equalizador série tende a apresentar um acréscimo substancial no desvio máximo e o equaliza-

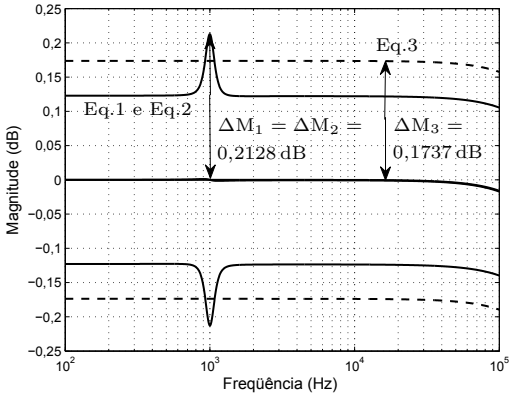


Figura 11: Respostas ideais e desvios estatísticos dos equalizadores série (1), paralelo (2) e Gundry (3) de uma banda, todos com ganho 0 dB.

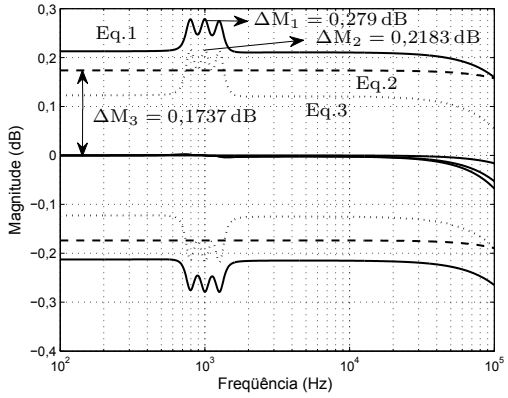


Figura 12: Respostas ideais e desvios estatísticos dos equalizadores série (1), paralelo (2) e Gundry (3) de três bandas, com todos os ganhos em 0 dB.

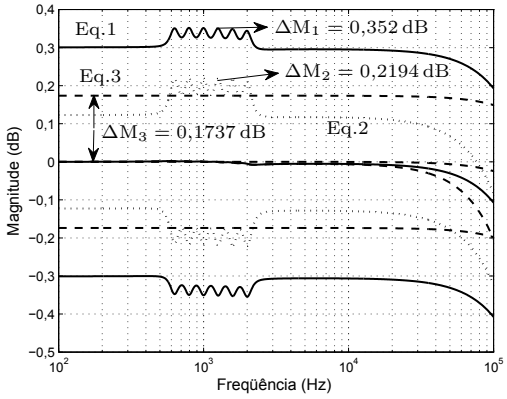


Figura 13: Respostas ideais e desvios estatísticos dos equalizadores série (1), paralelo (2) e Gundry (3) de seis bandas, com todos os ganhos em 0 dB.

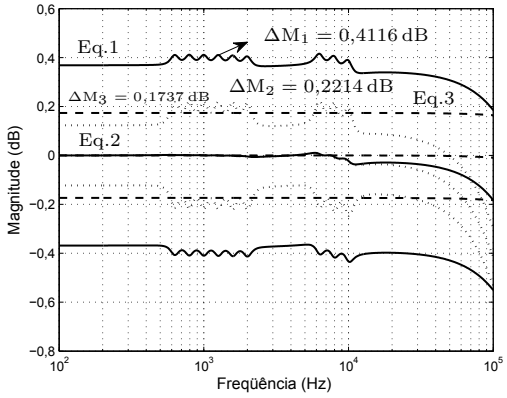


Figura 14: Respostas ideais e desvios estatísticos dos equalizadores série (1), paralelo (2) e Gundry (3) de nove bandas, todos com ganho de 0 dB.

dor paralelo, apenas um leve aumento. No caso do equalizador Gundry, o desvio máximo permanece no mesmo valor de 0,1737 dB.

Outro estudo aqui realizado refere-se à capacidade de o equalizador atenuar ou amplificar freqüências adjacentes. Por simplicidade e sem perda de generalidade, são considerados neste estudo apenas equalizadores de três bandas, com freqüências centrais iguais a 793,7 Hz, 1 kHz e 1,259 kHz (seções 17 a 19 da Tabela 1).

Primeiramente, considere a situação em que se deseja uma amplificação de 12 dB nas freqüências centrais das três seções. A Figura 16 apresenta tal situação. A resposta do equalizador série é muito ruim, apresentando ganhos superiores a 20 dB em uma faixa aproximada de 800 a 1200 Hz. Os equalizadores paralelo e Gundry apresentam respostas idênticas para esse caso. Tal fato ocorre porque para amplificação, as funções de transferência dessas estruturas são iguais [(14) e (24)]. Essas respostas não apresentam exatamente 12 dB nas freqüências f_0, mas já são muito mais próximas da desejada em relação à resposta do equalizador série.

Considere agora a situação inversa, em que se deseja uma atenuação de 12 dB nas freqüências centrais das três seções, situação ilustrada na Figura 17. A resposta do equalizador paralelo é muito ruim, com atenuação superior a 25 dB em 1 kHz. A resposta do equalizador série é um pouco melhor, mas apresenta uma grande ondulação na faixa entre 793,7 Hz e 1,259 kHz. Já a resposta do equalizador Gundry possui uma ondulação bem menor nessa mesma faixa de freqüência, em torno de 1 dB. Portanto, novamente o equalizador Gundry é o único que apresenta uma resposta aceitável. Pode-se notar que tal resposta é igual à obtida na situação anterior (Figura 16), mas com sinal negativo.

Por fim, considere o caso em que se deseja uma amplificação de 12 dB nas freqüências 793,7 Hz e 1,259 kHz e uma atenuação de 12 dB em 1 kHz, condição mostrada na Figura 18. Como no caso anterior, a única estrutura com comportamento aceitá-

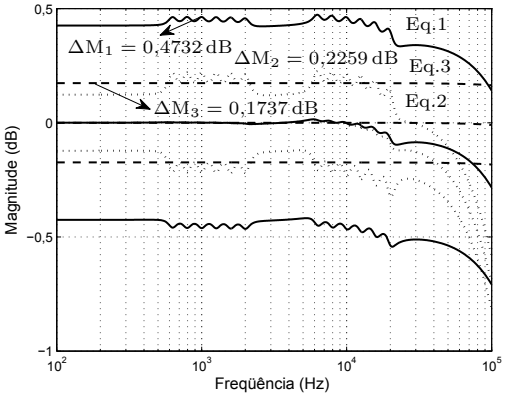


Figura 15: Respostas ideais e desvios estatísticos dos equalizadores série (1), paralelo (2) e Gundry (3) de doze bandas, todos com ganho de 0 dB.

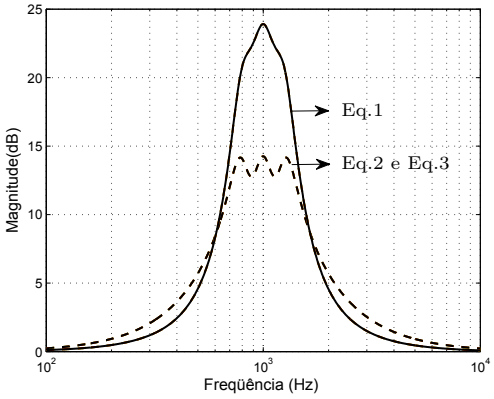


Figura 16: Respostas dos equalizadores série (1), paralelo (2) e Gundry (3) de três bandas, com amplificação de 12 dB em 793,7 Hz, 1 kHz e 1,259 kHz.

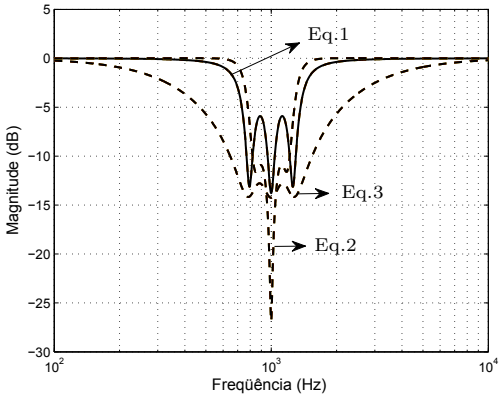


Figura 17: Respostas dos equalizadores série (1), paralelo (2) e Gundry (3) de três bandas, com atenuação de 12 dB em 793,7 Hz, 1 kHz e 1,259 kHz.

vel é o equalizador Gundry. As demais nem chegam a causar atenuação em alguma faixa de freqüência.

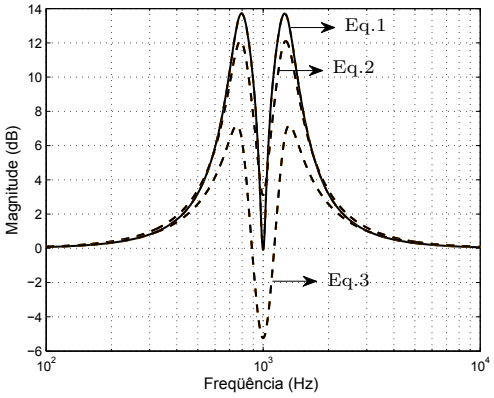


Figura 18: Respostas dos equalizadores série (1), paralelo (2) e Gundry (3) de três bandas, com amplificação de 12 dB em 793,7 Hz e 1,259 kHz e atenuação de 12 dB em 1 kHz.

4 CONCLUSÕES

Este trabalho apresentou um estudo comparativo de três estruturas comumente empregadas na implementação de equalizadores de áudio gráficos analógicos: equalizador série, equalizador paralelo e o aqui chamado equalizador Gundry [1, 2]. Primeiramente, foram analisados os desvios estatísticos causados por variações nos valores dos componentes dos equalizadores. Na seqüência, foram estudados os comportamentos das estruturas quando se deseja obter amplificações ou atenuações em faixas de freqüência adjacentes. No primeiro estudo, o equalizador Gundry demonstrou ser o menos sensível à tolerância dos componentes, apresentando um desvio máximo constante, independente do número de seções do equalizador. Na segunda análise, apenas o equalizador Gundry apresentou um resultado satisfatório para diferentes configurações de ganho ou atenuação em freqüências adjacentes, demonstrando que essa estrutura é a única das aqui estudadas capaz de implementar com desempenho satisfatório um equalizador gráfico comercial, como o equalizador de 1/3 de oitava (31 bandas).

REFERÊNCIAS BIBLIOGRÁFICAS

[1] K. J. Gundry, *U.K. Patent 1.452.920*, November 1973.

[2] K. J. Gundry, *Adjustable Equalizers Useable in Audio Spectrum*, U.S. Patent 3.921.104, November 1975.

[3] S. Noceti Filho, *Filtros Seletores de Sinais*, Editora da UFSC, Florianópolis, 3 edition, 2010.

[4] A. Papoulis and S. P. Unnikrishna, *Probability, Random Variables and Stochastic Processes*, Mcgraw Hill, 4 edition, 2002.

Sociedade de Engenharia de Áudio
Artigo de Congresso

Apresentado no 9º Congresso de Engenharia de Áudio
15ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2011, 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.

Aplicação de Controle Baseado em Passividade a Fontes de Alimentação para Amplificadores de Potência

Thiago R. de Oliveira[1] e Pedro F. Donoso Garcia[2]

[1] Instituto Federal de Minas Gerais,
Congonhas, Minas Gerais, Brasil.

[2] Universidade Federal de Minas Gerais, Departamento de Engenharia Eletrônica
Belo Horizonte, Minas Gerais, Brasil

thiago.oliveira@ifmg.edu.br, pedro@cpdee.ufmg.br

RESUMO

Apesar dos avanços tecnológicos verificados nas etapas de amplificação, as fontes de alimentação utilizadas em sistemas de áudio sofreram poucas modificações. Atualmente, as fontes não-reguladas ainda são largamente utilizadas, no entanto, a presença de componentes volumosos e grandes bancos capacitivos tem se tornado um entrave ao aumento da densidade energética dos equipamentos de amplificação dificultando também a redução de custos. Neste trabalho a utilização de fontes chaveadas para áudio, controladas por uma técnica não-linear baseada em passividade é analisada, tendo como foco a redução dos componentes passivos e a manutenção de baixos índices de *ripple* nos barramentos de alimentação.

0 INTRODUÇÃO

Ao longo das últimas décadas, as tecnologias envolvidas na cadeia de reprodução de sinais de áudio sofreram um avanço significativo. Em conseqüência dos estudos realizados em eletrônica aplicada, desenvolvidos a partir de meados dos anos 90, os amplificadores classe D aliam hoje alta eficiência e qualidade sonora compatível com tecnologias lineares, isso obtido por meio de topologias de inversores que produzem baixa distorção harmônica [1] e técnicas de controle que melhoram o seu desempenho ao longo de toda faixa de operação [2]. Contudo, apesar da evolução da tecnologia das etapas de amplificação de potência, a grande maioria das fontes de alimentação utilizadas em sistemas de amplificação ainda é baseada em topologias não-reguladas, as quais necessitam de componentes magnéticos volumosos e grandes bancos de capacitores para manter a tensão de seu barramento dentro de patamares aceitá-

veis. Além de possuírem um custo relativamente elevado, as fontes convencionais limitam o desempenho de amplificadores, em função da inexistência de regulação de linha e baixa regulação de carga [3]. Nota-se, então, que a manutenção dessa topologia apresenta um obstáculo ao aumento da integração de sistemas de áudio e um entrave à redução de custos. Assim sendo, a aplicação de fontes chaveadas a sistemas de amplificação permitiria mitigar os problemas referentes à regulação, além de possibilitar um aumento da densidade energética do equipamento. Em [4] é mostrado que um aumento ainda maior do nível de compactação pode ser obtido, com fontes chaveadas, a partir do emprego de técnicas de controle mais avançadas, como o controle robusto. Isso abre um ramo de investigação interessante, atrelando o desempenho de conversores em sistemas de amplificação à técnica de controle e não tanto à sua estrutura física. Neste trabalho, essa linha de investigação será estendida, sendo analisado o desempenho de um conversor estático ponte-completa, funcionando como fonte de alimentação para um amplificador de áudio, para a aplicação de uma técnica de controle baseada em passividade, a qual tem sido empregada com bons resultados em conversores funcionando como pré-reguladores de fator de potência [5].

1 ORIGEM DO RIPPLE DE TENSÃO

As oscilações de tensão presentes sobre os barramentos de alimentação de uma fonte podem ser geradas por uma série de fatores. Em fontes convencionais, por exemplo, o intervalo de transferência de energia entre a rede elétrica e o banco capacitivo provoca o aparecimento de um *ripple* de 120Hz na tensão de saída. Entretanto, quando se analisa o comportamento de fontes de alimentação aplicadas a sistemas de amplificação, observa-se o surgimento de perturbações na tensão dos barramentos devido à operação dos amplificadores. Em [6] é comentado que a corrente demandada por um amplificador possui uma forte característica alternada, esta ao interagir com a impedância de saída da fonte provoca oscilações. O conhecimento da magnitude das componentes alternadas presentes na corrente de fonte, para uma determinada classe de amplificadores, permite que se estipule limites para a impedância de saída da fonte de alimentação, de modo a se obter um nível de *ripple* de tensão desejado.

Tomando como exemplo um amplificador linear classe AB, a corrente de fonte pode ser determinada a partir da análise de seu comportamento. A Figura 1 mostra as etapas de funcionamento do estágio de potência do amplificador classe AB, ao sintetizar um sinal senoidal.

Observa-se que cada barramento de alimentação fornece ao amplificador um semiciclo da corrente de carga. A potência instantânea fornecida pela fonte pode ser definida como:

$$P_S(t) = V_{CC}(I_{S1} - I_{S2}) \tag{1}$$

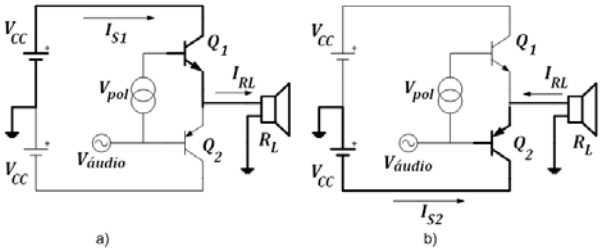


Figura 1: **Fluxo da corrente do amplificador para a carga. a) semiciclo positivo da senóide de saída. b) semiciclo negativo da senóide de saída**

Desse modo, a forma de onda da potência fornecida se apresenta como o ilustrado na Figura 2.

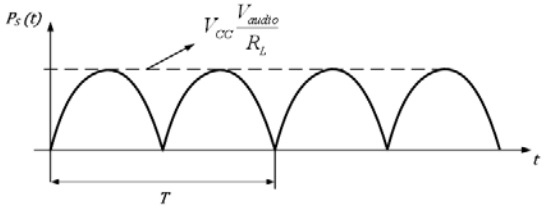


Figura 2: **Potência instantânea na fonte**

Onde:

V_{audio} - Valor de pico do sinal de áudio reproduzido pelo amplificador

Assumindo que a tensão nas barramentos de alimentação é regulada, a forma de onda de corrente de fonte será equivalente ao apresentado na Figura 2. Fazendo a decomposição em série de Fourier da corrente, encontra-se:

$$i_S(t) = \frac{V_{audio}}{R_L}\left[\frac{2}{\pi} - \frac{4}{\pi}\sum_n \frac{1}{n^2-1}cos(n\omega_a t)\right] \quad n = 2, 4, 6, 8... \tag{2}$$

Onde:

ω_a - Frequência angular do sinal de áudio

A equação 2 mostra que além da componente contínua, estão presentes na corrente de fonte componentes alternadas de frequência igual aos múltiplos pares da frequência do sinal de áudio sintetizado. A amplitude das componentes decai de acordo com a ordem harmônica, como mostra a Figura 3.

Observa-se que a a partir de *n = 10*, a magnitude das componentes é inferior a 1% da amplitude máxima da corrente de fonte, o que indica que essas componentes irão possuir um efeito sobre o *ripple* de tensão muito pequeno. Assumindo uma banda audível de 20kHz, pode-se constatar que a maior componente alternada presente na corrente de fonte, cuja magnitude pode afetar significativamente o *ripple* de tensão ocorrerá até 200kHz.

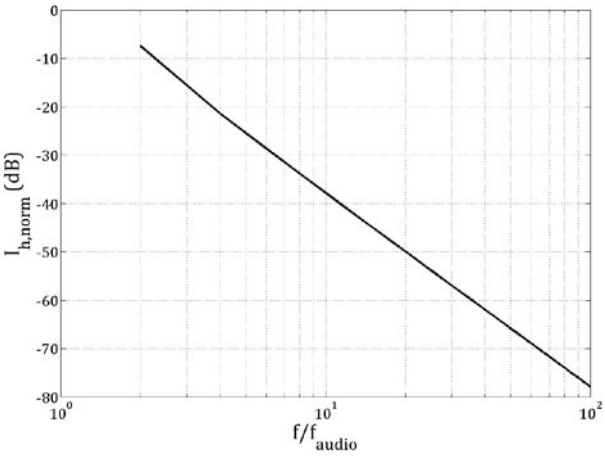


Figura 3: **Amplitude das componentes harmônicas em função da ordem harmônica, normalizadas pelo valor de pico máximo da corrente de fonte.**

Considerando que a magnitude do *ripple* possa ser estimada pela componente alternada com maior amplitude, define-se:

$$\Delta E = I_{h2_{MAX}} Z_{out} = \frac{4}{3\pi} \frac{V_{audio}}{R_L} Z_{out} \tag{3}$$

Como $\Delta E = kV_{CC}$, pode-se definir o *ripple* da tensão do barramento c.c. como sendo:

$$k = \frac{4}{3\pi} \frac{V_{audio}}{V_{CC}} \frac{Z_{out}}{R_L} \tag{4}$$

A equação 4 apresenta a relação entre o *ripple* de tensão (k) e a impedância de saída da fonte (Z_{out}), podendo ser muito útil para se estimar o desempenho de fontes de alimentação.

Em fontes convencionais, como comentado anteriormente, existem pelo menos dois mecanismos que afetam as oscilações de tensão sobre o barramento c.c. Considerando apenas o efeito da perturbação de carga e assumindo que durante os intervalos de carregamento do banco capacitivo, o fornecimento de energia ao amplificador é feito apenas pelos capacitores de saída da fonte, pode-se definir a seguinte relação:

$$V(\omega) = \frac{1}{j\omega C} I_S(\omega) \tag{5}$$

Onde:

C - Valor da capacitância do banco por barramento

A partir da equação 5, pode-se calcular o valor da impedância de saída de uma fonte convencional, e então determinar o máximo *ripple* definido por cada banco capacitivo. Para se obter, por exemplo, um *ripple* de 10%, com tensão de saída do amplificador igual a 80% do seu valor máximo ($\frac{V_O}{V_{CC}} = 0,8$) e carga do amplificador de 2Ω, seria necessário o emprego de um banco capacitivo de 6800μF em cada barramento de saída.

2 CONVERSOR UTILIZADO

Para analisar o emprego da técnica de controle baseada em passividade utilizou-se um conversor estático ponte-completa, com a configuração apresentada na Figura 4.

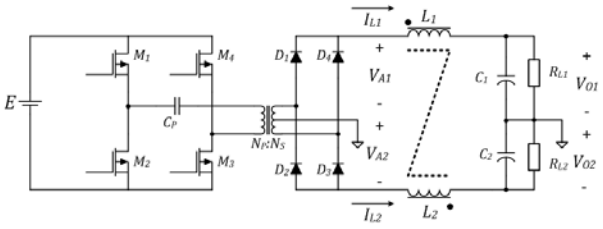


Figura 4: **Estrutura do conversor ponte-completa utilizado**

Esse conversor apresenta dois barramentos de alimentação com o intuito de se poder utilizar amplificadores em configuração *Single-ended*. O acoplamento dos indutores foi introduzido para se melhorar a simetria entre as barramentos, minimizando problemas causados pelo desequilíbrio de cargas provocado pelo amplificador de áudio [7]. O tipo de acionamento utilizado em conversores desse tipo permite apenas que a tensão entre os dois barramentos de saída seja controlada, dessa forma, é interessante fazer o modelamento do conversor de modo a se obter um modelo *buck* equivalente, o qual simplifica a aplicação de técnicas de controle.

Considerando o conversor da Figura 4, as equações de estado dos dois barramentos podem ser escritas segundo a Equação 6.

$$\begin{cases} uV_{A1} = L_1 \frac{dI_{L1}}{dt} + V_{O1} + L_M \frac{dI_{L2}}{dt} \\ I_{L1} = C_1 \frac{dV_{O1}}{dt} + \frac{V_{O1}}{R_1} \end{cases}$$
$$\begin{cases} uV_{A2} = L_2 \frac{dI_{L2}}{dt} + V_{O2} + L_M \frac{dI_{L1}}{dt} \\ I_{L2} = C_2 \frac{dV_{O2}}{dt} + \frac{V_{O2}}{R_2} \end{cases} \tag{6}$$

Onde:

L_M - Indutância mútua entre os dois indutores do filtro

Para se fazer a definição do modelo *buck* equivalente, deve-se implementar as seguintes mudanças de variáveis:

$$\begin{aligned} V_A &= V_{A1} + V_{A2}; \\ V_O &= V_{O1} + V_{O2}; \\ I_L &= \tfrac{I_{L1}+I_{L2}}{2}; \end{aligned} \tag{7}$$

Somando os dois sistemas da Equação 6 e considerando as mudanças de variáveis da Equação 7 e assumindo ainda que a corrente de carga pode ser decomposta em uma parcela contínua e uma perturbação alternada (I_R), encontra-se as equações de estado do conversor equivalente:

$$\begin{cases} uV_A = 2(L + L_M)\frac{dI_L}{dt} + V_O \\ I_L = \frac{C}{2}\frac{dV_O}{dt} + \frac{V_O}{2R} + I_R \end{cases} \tag{8}$$

Onde: $L = L_1 = L_2$;
$C = C_1 = C_2$;
$R = R_1 = R_2$;

Considerando a técnica de projeto de uma fonte chaveada ponte-completa, definida em [8], definiu-se os componentes do conversor como sendo: $L = 330\mu H$, $C = 18\mu F$ e $R = 6,28\Omega$. O conversor foi projetado para as seguintes condições: $V_{CC} = \pm 35V$, $R_{Lmin} = 2\Omega$, $f_{switch} = 44kHz$ e razão cíclica máxima de 50%. Com isso, obtém se uma potência de saída da fonte máxima de aproximadamente 400W e corrente de fonte média, em cada barramento, de 5,6A.

3 PASSIVIDADE

Em [9] é apresentada uma metodologia de modelagem de sistemas dinâmicos focada nos conceitos de energia, *i.e.*, nos mecanismos de armazenamento e dissipação de energia dos sistemas, baseada na abordagem de Euler-Lagrange (*EL*). A partir do modelo de *EL*, é proposta uma técnica de controle conhecida como controle baseado em passividade, cuja finalidade é definir uma relação onde o sistema armazene menos energia do que lhe é fornecida pela fonte primária, e assim garantir estabilidade assintótica em um ponto de equilíbrio. Associada a conversores estáticos, a técnica de controle baseada em passividade possibilita uma melhor rejeição à perturbações de fonte [10], em comparação a técnicas lineares, e oferece um melhor comportamento em relação à condição de rastreamento de trajetória, tornando essa técnica interessante na estabilização de pré-reguladores de fator de potência, como discutido em [5].

Existem porém, algumas restrições sobre as condições que levam um sistema passivo, *i.e.*, que armazena menos energia do que lhe é fornecida, a ser assintoticamente estável. No entanto, [9] mostra que essas restrições podem ser satisfeitas por meio da aplicação de um controle em malha fechada.

3.1 Modelo de Euler-Lagrange

O modelo de Euler-Lagrange para o conversor *buck* equivalente é descrito por [5]. Ele é deduzido a partir da solução da equação de Euler-Lagrange, apresentada em (9).

$$\frac{d}{dt}\left(\frac{\partial \ell}{\partial \dot{q}}\right) - \frac{\partial \ell}{\partial q} = -\frac{\partial \mathcal{N}}{\partial \dot{q}} + \mathcal{F}_q \tag{9}$$

Onde:
q - Vetor carga elétrica.
$\dot{q}$ - Vetor corrente.
$\mathcal{N}$ - Função de dissipação de Rayleigh.
$\mathcal{F}_q$ - Fontes de energia externas.
ℓ - Lagrangiano do sistema.

O termo Lagrangiano representa a diferença entre a co-energia magnética do circuito ($\mathcal{M}(\dot{q})$) e a energia de campo elétrico do circuito ($\mathcal{W}(q)$), ou seja:

$$\ell = \mathcal{M}(\dot{q}) - \mathcal{W}(q) \tag{10}$$

Considerando um conversor estático, o vetor de cargas elétricas pode ser constituído das cargas armazenadas nos capacitores (q_C) e nos indutores (q_L). De acordo com [11], um conversor *buck* apresenta os seguintes parâmetros do modelo de *EL*:

$$\mathcal{M}(\dot{q_L}) = \frac{1}{2}L_f \dot{q_L}^2 \tag{11}$$

$$\mathcal{W}(q_C) = \frac{1}{2C_f}q_C^2 \tag{12}$$

$$\mathcal{N}(\dot{q}) = \frac{1}{2}\frac{(\dot{q}_L - \dot{q}_C)^2}{G_f} \tag{13}$$

$$\mathcal{F}_{q_L} = \mu E \tag{14}$$

$$\mathcal{F}_{q_C} = 0 \tag{15}$$

Onde: $G_f = \frac{1}{R_f}$.
$R_f = 2R$;
$C_f = \frac{C}{2}$;
$L_f = 2(L + L_M)$.

Resolvendo a equação de Euler-Lagrange, encontra-se:

$$\begin{cases} L_f \ddot{q_L} = -\frac{(\dot{q_L} - \dot{q_C})}{G_f} + \mu E \\ \frac{q_C}{C_f} = \frac{(\dot{q_L} - \dot{q_C})}{G_f} \end{cases} \tag{16}$$

Realizando na Equação 16 uma mudança de variáveis, onde: $z_1 = i_L = \dot{q_L}$ e $z_2 = v_O = q_C/C_f$, encontra-se o modelo não perturbado para o conversor *buck*:

$$\begin{cases} \dot{z_1} = -\frac{1}{L_f}z_2 + \mu\frac{E}{L_f} \\ \dot{z_2} = \frac{1}{C_f}z_1 - \frac{G_f}{C_f}z_2 \end{cases} \tag{17}$$

Observa-se que o modelo encontrado é semelhante ao definido em (8) desconsiderando a perturbação de carga. O modelo pode ser escrito na forma matricial, como se segue:

$$D_B\dot{z} + (J_B + R_B)z = \mu E_B \tag{18}$$

Onde:

$$z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \quad E_B = \begin{bmatrix} E \\ 0 \end{bmatrix} \quad D_B = \begin{bmatrix} L_f & 0 \\ 0 & C_f \end{bmatrix}$$
$$J_B = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad R_B = \begin{bmatrix} 0 & 0 \\ 0 & G_f \end{bmatrix} \tag{19}$$

A função de armazenamento, que mede a quantidade de energia armazenada pelo conversor, pode ser descrita como:

$$\mathcal{V}(z) = \frac{1}{2}z^T D_B z \tag{20}$$

3.2 Controle baseado em passividade

Baseada na função de armazenamento do sistema, dada pela Equação 20, pode-se propor uma função de armazenamento desejada para o sistema em malha fechada, esta, por sua vez, designada pela Equação 21.

$$\mathcal{V}_d = \frac{1}{2}\tilde{z}^T D_B \tilde{z} \tag{21}$$

Por definição têm-se:

$$\tilde{z} \triangleq z - z_d \tag{22}$$

$$\tilde{z} = \begin{bmatrix} \tilde{z_1} & \tilde{z_2} \end{bmatrix}^T \tag{23}$$

$$z_d = \begin{bmatrix} z_{d1} & z_{d2} \end{bmatrix}^T \tag{24}$$

Onde:
z - Valor real dos estados.
z_d - Valor desejado dos estados.
$\tilde{z}$ - Erro entre o valor desejado dos estados e o seu valor real.

A dinâmica desejada para o erro médio (Ψ), associada com a função de armazenamento definida na Equação 21, é descrita pela Equação 25.

$$D_B\dot{\tilde{z}} + (J_B + R_{Bd})\tilde{z} = \Psi \tag{25}$$

Na equação 25, R_{Bd} é a matriz de amortecimento modificada de forma a se obter uma função de dissipação desejada. Com isso a função de dissipação associada ao erro médio do sistema pode ser escrita como:

$$\mathcal{N} = \frac{1}{2}\tilde{z}^T R_{Bd}\tilde{z} = \frac{1}{2}\tilde{z}^T (R_B + R_{nB})\tilde{z} \tag{26}$$

R_{nB} representa um termo de amortecimento virtual a ser inserido no sistema pelo circuito de controle, o qual pode ser definido genericamente como:
$R_{nB} = \begin{bmatrix} R_1 & 0 \\ 0 & G_2 \end{bmatrix}$. Sendo $R_1 > 0$ e $G_2 > 0$.

Dessa forma, a dinâmica do controlador pode ser determinada a partir das Equações 18 e 25, obtendo-se por fim a equação 27.

$$\Psi = \mu E_B - [D_B\dot{z_d} + (J_B + R_B)z_d - R_{nB}\tilde{z}] \tag{27}$$

Fazendo $\Psi = 0$, de modo a se obter um erro médio nulo, encontra-se:

$$\mu E_B = D_B\dot{z_d} + (J_B + R_B)z_d - R_{nB}\tilde{z} \tag{28}$$

A Equação 28 pode ser expandida para o sistema apresentado na Equação 29.

$$\begin{cases} \mu E = L_f\dot{z_{d1}} + z_{d2} - R_1(z_1 - z_{d1}) \\ 0 = C_f\dot{z_{d2}} - z_{d1} + G_f z_{d2} - G_2(z_2 - z_{d2}) \end{cases} \tag{29}$$

Essa estrutura de controlador permite a implementação de diversas técnicas de controle. Dentre elas serão apresentadas as técnicas de controle direto da tensão de saída e o controle indireto modificado, proposto em [7].

3.2.1 Controle direto

Deseja-se que a tensão de saída do conversor assuma um valor desejado, constante, $z_{2d} = V_{od}$. Fazendo $R_1 = 0$ as equações do controlador podem ser derivadas da Equação 29 de forma que:

$$\begin{cases} z_{d1} = G_f V_{od} - G_2(z_2 - V_{od}) \\ \mu = \frac{1}{E}(L_f\dot{z_{d1}} + V_{od}) \end{cases} \tag{30}$$

De acordo com [11], a aplicação direta do controlador não garante que o sistema apresentará erro em estado estacionário nulo. Isso se deve, principalmente, a elementos parasitas não modelados no circuito do conversor, de forma que não há um equilíbrio exato entre fornecimento e dissipação de energia, para o ponto de equilíbrio determinado pelo sistema de controle. Assim sendo, para se corrigir tal problema, recomenda-se a inserção de um termo integral às equações do controlador, termo este definido como:

$$Int = -k_i \int_0^t [z_2(s) - V_{od}]ds, \quad k_i > 0 \tag{31}$$

Neste trabalho o termo integral foi inserido na equação de definição de z_{d1}.

3.2.2 Controle indireto modificado

Nesta topologia de controle, a tensão de saída será definida indiretamente pela corrente média no indutor do filtro. Fazendo $z_{d1} = I_d + \Delta i_L$, onde $I_d = GV_{od}$ e $\Delta i_L = (k_p + \frac{k_i}{s})(V_{od} - z_2)$, têm-se:

$$\begin{cases} \dot{z_{d2}} = \frac{1}{C_f}(z_{d1} - G_f z_{d2}) \\ \mu = \frac{1}{E}[L_f\dot{z_{d1}} + z_{d2} - R_1(z_1 - z_{d1})] \end{cases} \tag{32}$$

3.2.3 Desempenho dos controladores

Considerando o conversor definido na seção 2, foi realizada a simulação do desempenho da sua impedância de saída. Assumiu-se para essa simulação um capacitor de filtro de 18μF, $k_p = G_2 = 1$, $R_1 = 25\Omega$ e $k_i = 100$. A Figura 5 mostra curva de resposta em frequência para a impedância de saída do conversor com controle baseado em passividade.

Tendo como base a equação 4, pode-se definir o valor do *ripple* máximo por meio da impedância de saída em malha fechada. Pela Figura 5, observa-se que o controle direto oferece uma oscilação de tensão máxima na faixa de 20%, para toda faixa de áudio. Já o controle indireto modificado, apesar de apresentar um *ripple* máximo maior (38%) permite uma redução significativa do *ripple* em baixa frequência (<3%), onde se sabe que os amplificadores de áudio irão apresentar maior demanda de energia. Uma melhoria do desempenho pode ser obtida aumentando-se a capacitância do filtro de saída do conversor ponte-completa, fazendo

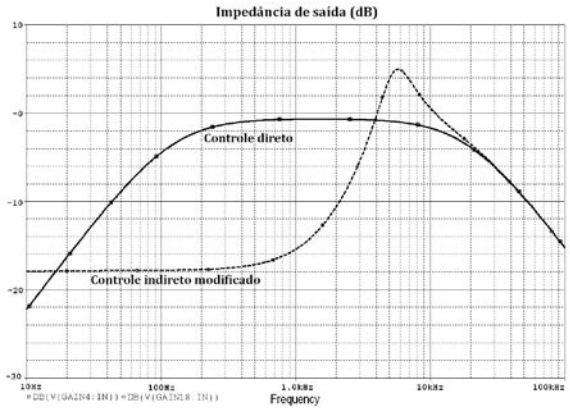


Figura 5: **Comparação da resposta da impedância de saída do conversor *buck* equivalente para controle direto da tensão e controle indireto modificado.**

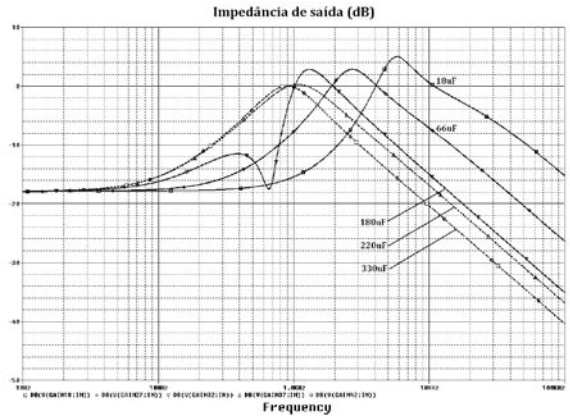


Figura 6: **Variação da capacitância do filtro do conversor e sua influência na impedância de saída, para uma técnica de controle indireto modificada**

isso obteve-se as curvas de resposta em frequência da impedância de saída mostradas na Figura 6.

Conforme a Figura 6, o aumento da capacitância do filtro provoca um deslocamento do máximo valor da impedância para região de baixa frequência e reduz a magnitude da impedância. Observa-se que para capacitâncias de 66μF o *ripple* máximo será de 30% e ocorrerá para sinais de áudio de 1,25kHz, já para capacitâncias de 220μF e 330μF, o valor máximo do *ripple* será 21% e ocorrerá para sinais de aproximadamente 500Hz.

3.3 Simulação do conversor para carga reativa

Ao se analisar a Figura 6 nota-se que o emprego de um capacitor de filtro maior pode fazer com que o desempenho do conversor melhore. No entanto, o modelo não contempla o comportamento de conversores ao alimentarem amplificadores com carga reativa (alto-falante). Para verificar o desempenho nessa situação, o conversor projetado foi simulado, tendo como carga um amplificador que aciona um arranjo de quatro alto-falantes em paralelo, representados pelo modelo elétrico da bobina de um alto-falante comercial de código 12CO1P. O modelo da bobina segue a abordagem de Thiele-small [12], sendo os parâmetros mais relevantes apresentados na Tabela 1.

Tabela 1: Parâmetros para o modelo elétrico do alto-falante.

Parâmetro	Símbolo	Valor
Fator de força	Bl	12,8Tm
Resistência da bobina	R_E	$6,3\Omega$
Indutância da bobina	L_E	2,287mH
Massa móvel	M_D	45,8g
Compliância mecânica	C_D	$250,4\mu m/N$
Resistência mecânica da suspensão	R_D	3,01kg/s

A Figura 7 mostra a simulação do conversor, para o ponto de maior impedância de saída, considerando um capacitor de filtro igual a 18μF. A Figura 8, por sua vez, apresenta a simulação para um conversor com capacitância de 330μF. As informações acerca da magnitude do *ripple* resultante são apresentadas na Tabela 2.

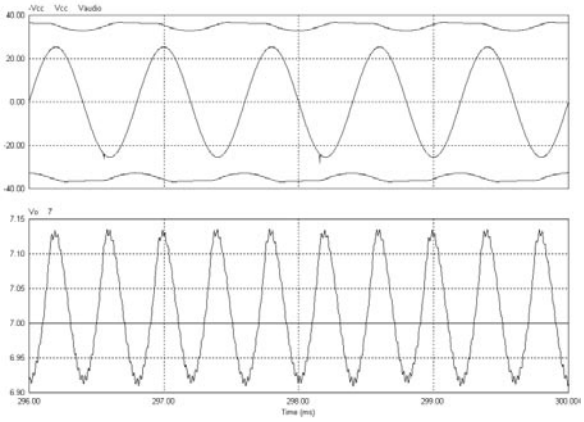


Figura 7: **Formas de onda de tensão nos barramentos da fonte e na saída do amplificador (Superior). Tensão entre barramentos (Inferior). Sinal de áudio de 1,25kHz, carga reativa e controle baseado em passividade. C = 66μF.**

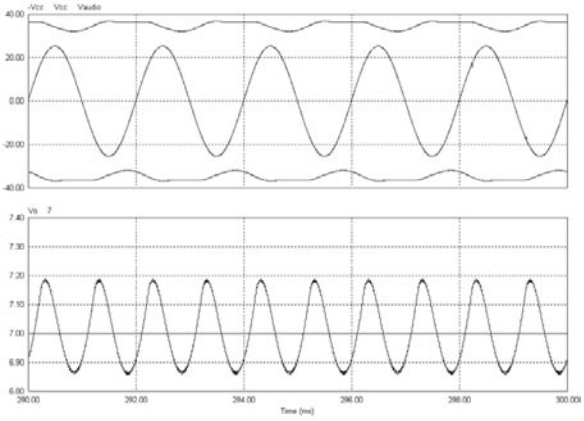


Figura 8: **Formas de onda de tensão nos barramentos da fonte e na saída do amplificador (Superior). Tensão entre barramentos (Inferior). Sinal de áudio de 500Hz, carga reativa e controle baseado em passividade. C = 330μF.**

Tabela 2: *Ripple* de tensão na saída do conversor com carga reativa.

***Ripple* no barramento c.c.**	**Modelo**	**Simulação**
C = 66μF	15,6%	11,7%
C = 330μF	16,0%	13,5%

Observa-se, pela Tabela 2 que o *ripple* obtido é inferior ao estipulado pelo modelo, indicando que além da variação da impedância com a frequência, o defasamento entre a corrente e tensão de saída também surte efeito na definição das oscilações de tensão. Nota-se, também, que apesar do modelo indicar que a utilização de um conversor com capacitância de 330μF implicaria em uma redução significativa do *ripple* em relação às demais condições avaliadas, ao se acionar um sistema com carga reativa, isso não se concretiza. As simulações mostram que o conversor com capacitância de 66μF é capaz de produzir um desempenho melhor, isso porque, o seu ponto de máxima impedância ocorre em uma região onde o módulo da impedância do arranjo de alto-falantes apresenta um valor superior à 2Ω. Para a outra condição considerada, o ponto de máxima impedância ocorre em uma frequência mais baixa, onde a impedância do arranjo é menor do que 2Ω.

4 RESULTADOS EXPERIMENTAIS

Nesta seção o conversor projetado para um banco capacitivo de 66μF foi construído. O diagrama de controle foi implementado por meio de amplificadores operacionais, dando um caráter totalmente analógico ao conversor. Os testes foram realizados com cargas de 5 e 8Ω resistivas e amplitude da tensão de saída do amplificador de $50V_{pp}$. Inicialmente, fez-se uma varredura de frequência no sinal de áudio sintetizado pelo amplificador. A Figura 9, mostra o resultado da varredura.

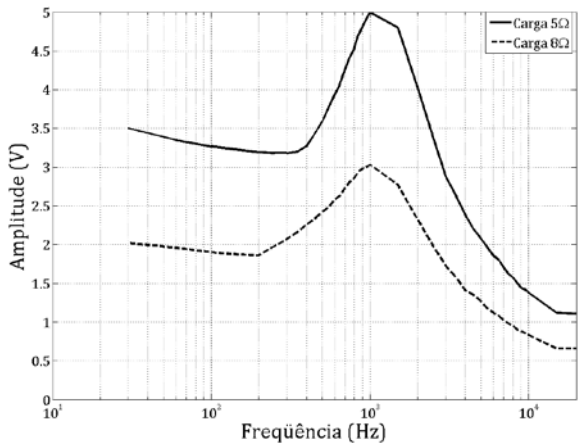


Figura 9: **Resposta em frequência do valor de pico-a-pico do *ripple* de tensão na barramento positiva da fonte de alimentação.**

Ao se analisar a Figura 9, observa-se que o perfil do *ripple* de tensão condiz com o formato da curva de impedância prevista no modelo, coincidindo também o ponto de máxima oscilação. Nota-se que o comportamento do *ripple* de tensão para baixas freqüências, ao contrário do deduzido no modelo teórico, não apresenta um patamar fixo. Isso se deve à influência da perturbação de linha no barramento c.c. do primário do transformador isolador e à presença de oscilações provocadas por imperfeições no acoplamento entre os indutores do filtro. Essas imperfeições não contempladas pelo modelo, somadas à presença das demais componentes alternadas além de *n=2*, fazem com que o *ripple* observado na prática seja superior ao deduzido através das curvas da Figura 6, sendo que no conversor implementado obteve-se um *ripple* máximo de 14,3% ($R_L = 5\Omega$), enquanto o modelo estimava um valor de aproximadamente 8%.

A Figura 10 mostra a forma de onda de tensão na saída do amplificador e nos barramentos de alimentação, para um sinal de 1,25kHz, referente ao ponto de máximo *ripple*. Observa-se que, apesar do aumento do *ripple* máximo em relação ao modelo teórico, a magnitude da oscilação não provoca ceifamento da forma de onda do sinal de áudio.

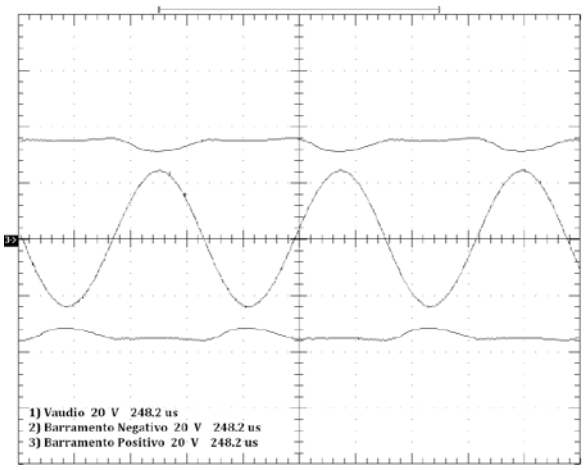


Figura 10: **Formas de onda de tensão de saída do amplificador e barramentos da fonte de alimentação, para $50V_{pp}$ de tensão de saída do amplificador, 5Ω de carga e $f_{udio} = 1,25kHz$**

Substituindo as cargas resistivas por um alto-falante do mesmo modelo descrito na seção anterior, realizou-se nova varredura. O resultado dessa varredura foi comparado com o obtido para uma carga de 8Ω, sendo mostrado na Figura 11.

Conforme observado na Figura 11, para frequências inferiores a 1kHz, a magnitude do *ripple* de tensão não se altera significativamente com o aumento da frequência, sendo inferior a 5,5%. Isso indica que o controle implementado permite uma menor dependência do desempenho da fonte em relação à variação de carga, para sinais de baixa frequência. Além disso, percebe-se que em todas as condições avaliadas na figura, o desempenho de um conversor ao alimentar um sistema com carga reativa é superior ao verificado para carga resistiva, como exposto pela teoria.

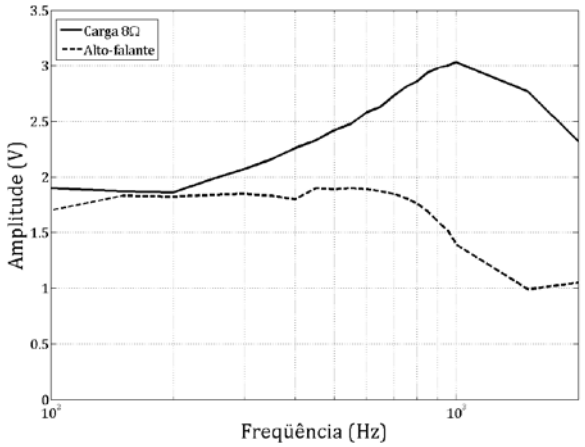


Figura 11: **Resposta em frequência do valor de pico-a-pico do *ripple* de tensão no barramento positivo da fonte de alimentação para um alto-falante.**

5 CONCLUSÕES

Neste trabalho verificou-se o desempenho de um conversor estático ponte-completa funcionando como fonte de alimentação para amplificadores de áudio, controlado por uma técnica baseada nos conceitos de passividade de sistemas dinâmicos. Deduziu-se que as oscilações de tensão nos barramentos de alimentação são originadas por harmônicos do sinal de áudio injetadas na corrente de fonte.

A análise teórica mostrou que um conversor controlado pela técnica proposta pode oferecer magnitudes de *ripple* inferiores a 20%, para capacitâncias de filtro de aproximadamente 330μF e considerando carga resistiva de 2Ω. No entanto, a redução do valor de capacitância não implica necessariamente em uma piora do desempenho. Como mostrado, para a situação de sistemas com cargas reativas, o perfil da impedância do arranjo de alto-falantes pode produzir uma situação onde um conversor com capacitância menor poderá apresentar melhor desempenho.

Os resultados experimentais mostraram que a presença de elementos não modelados provocam a elevação do *ripple* de tensão em relação à teoria, no entanto, essa elevação não afeta significativamente a forma de onda do sinal de áudio com amplitudes inferiores a 70% do valor máximo. Além disso, ao alimentar cargas reativas, o controle permite uma menor dependência da tensão de alimentação com a variação da carga.

AGRADECIMENTOS

Os autores gostariam de agradecer à FAPEMIG pelo suporte e apoio financeiro ao trabalho de pesquisa.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] K. Nielsen, "Parallel phase shifted carrier pulse width modulation (pscpwm) - a novel approach to switching power amplifier design," *102nd Audio Engineering Society Convention*, 1997, Munique - Alemanha.

[2] K. Nielsen, "Mecc - a novel control method for high end switching audio power amplification," *105th Audio Engineering Society Convention*, 1998, São Francisco - EUA.

[3] E. Mendenhall, "Power supply regulation in audio power amplifiers," *113th Audio Engineering Society Convention*, 2002, Conference paper 5694. Los Angeles - EUA.

[4] T. Ogawa, Y. Miyazawa, K. Higuchi, T. Kajikawa, A. Shimizu, and O. Yoshizawa, "Compact audio power supply using approximate 2dof robust digital control," *SICE annual conference*, pp. 2142–2145, 2008, Tokyo - Japão.

[5] E. A. Oliveira, L. M. F. Morais, I. S. Seleme Jr., P. F. Donoso-Garcia, P. C. Cortizo, and M. A. S. Mendes, "Power factor correction via passivity-based adaptive controller using a buck converter operating in continuous mode," *IEEE Workshop on Control and Modeling for Power Electronics (COMPEL 08)*, pp. 1 – 6, 2008, Zurique - Suíça.

[6] L. Risbo and T. Morch, "Performance of an all digital power amplification system," *104th Audio Engineering Society Convention*, 1998, Amsterdã - Holanda.

[7] T. R. Oliveira, "Estudo e implementação de uma fonte chaveada empregando controle baseado em passividade, para alimentação de amplificadores de potência de áudio," M.S. thesis, PPGEE/UFMG, 2011, Belo Horizonte, Brasil.

[8] I. Barbi, *Eletrônica de potência: Projeto de fontes chaveadas*, Edição do autor, 2ª edição edition, 2007, Florianópolis - UFSC.

[9] R. Ortega, A. Loría, P. J. Nicklasson, and H. Sira-Ramírez, *Passivity-based control of euler-lagrange systems: Mechanical, electrical and electromechanical applications*, Springer-Verlag, 1ª edition, 1998, Londres.

[10] G. Escobar, R. Ortega, H. Sira-Ramirez, J.P. Vilain, and I. Zein, "An experimental comparison of several non linear controllers for power converters," *Proceedings of IEEE International Symposium on Industrial Electronics. ISIE'97.*, 1997, Guimarães - Portugal.

[11] E. A. Oliveira, "Aplicação do controle baseado em passividade em conversores estáticos operando como pré-reguladores de fator de potência," M.S. thesis, PPGEE/UFMG, 2008, Belo Horizonte, Brasil.

[12] M. R. Andersen, *Compensation of nonlinearities in transducers*, Ph.D. thesis, Universidade Técnica da Dinamarca, 2005, Dinamarca.

Sociedade de Engenharia de Áudio
Artigo de Congresso

Apresentado no 9º Congresso de Engenharia de Áudio
15ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2011, 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.

Hardware Básico para Controle via MIDI

Miguel B. Ratton[1]

[1] Informus Music Center
Caixa Postal 13.507, Centro, Curitiba, PR, Brasil

suporte@music-center.com.br

RESUMO

Este artigo apresenta o projeto de um núcleo básico de hardware para controle via MIDI. O hardware em questão é um circuito eletrônico baseado em microcontrolador, dotado de portas de MIDI e de USB e capaz de gerenciar vários sinais digitais e analógicos vindos de dispositivos de entrada (chaves e potenciômetros). Graças à sua arquitetura modular, utilizando o mesmo hardware central é possível implementar diferentes tipos de equipamentos controladores MIDI, para diversas aplicações diferentes, tais como teclados musicais, controladores de software de gravação, dentre outros.

0 INTRODUÇÃO

MIDI (Musical Instrument Digital Interface) é um protocolo de comunicação digital para a transferência de informações entre instrumento musicais e computadores [1]. Logo que foi criado, em 1983, revolucionou a música de maneira irreversível, pois trouxe recursos que facilitaram os processos de composição e execução musical.

A aplicação mais comum que se faz do protocolo MIDI é o controle de um instrumento a partir de outro. Para isto, basta conectar a saída de MIDI (Out) de um teclado à entrada de MIDI (In) do outro e configurar corretamente as condições de transmissão e recepção (canais, etc.) em cada um deles.

Basicamente, o protocolo MIDI funciona da seguinte maneira: quando se toca uma tecla (nota) no instrumento, ele transmite uma mensagem codificada digitalmente que informa qual foi a tecla pressionada (*key number*) e com que intensidade (*key velocity*) ela foi pressionada. Essa mensagem é enviada através do cabo de MIDI e, ao ser recebida pelo outro instrumento, este então executará, com seu próprio timbre, a nota musical correspondente.

A transferência das mensagens de MIDI é feita de forma serial assíncrona, a uma taxa de 31.250 bps, havendo um bit de início (*start bit*) e outro de fim (*stop bit*) em cada byte transferido. As mensagens para acionamento de notas e controle de parâmetros em geral contêm dois ou três bytes, mas há alguns tipos de mensagens que podem conter apenas um byte (ex: mensagens de sincronização do tipo start, stop, clock, etc.) ou mesmo dezenas de bytes (ex: mensagens exclusivas do tipo SysEx).

Embora a concepção inicial do protocolo MIDI tenha sido orientada para o uso com sintetizadores e instrumentos musicais em geral, o sistema foi idealizado de tal forma que permitiu – e ainda permite – ser expandido para que possa ser usado também com diversos outros tipos de equipamentos, direta ou indiretamente associados à produção, execução e reprodução de música. Portanto, além das ações estritamente musicais, tais como a execução e o controle de notas, existem também muitas outras informações que podem ser transferidas através de mensagens de MIDI.

Com o uso intensivo de softwares de gravação de áudio e de sintetizadores virtuais (*software synthesizers*), é cada vez maior a necessidade de dispositivos "físicos" para o controle desses aplicativos, não só para a execução das notas musicais propriamente ditas, mas também para que se possa efetuar de forma prática o ajuste dos diversos parâmetros sonoros (volume, pan, efeitos, etc.). Isto tem estimulado o surgimento de uma variedade de teclados controladores. Apesar da sua aparência, esses teclados não produzem sons, mas são capazes de enviar os comandos digitais (via MIDI) necessários para controlar instrumentos MIDI. Alguns desses teclados controladores são dotados de controles deslizantes e/ou rotativos, que podem ser configurados para enviar mensagens de MIDI apropriadas para o ajuste dos botões virtuais dos softwares (veja Fig.1).

Figura 1 – Exemplo de teclado controlador MIDI

Além dos teclados controladores, há também a categoria de equipamentos MIDI projetados especificamente para controlar softwares de gravação, chamados genericamente de "superfícies de controle". Nos últimos anos, têm surgido inúmeros modelos deste tipo de equipamento para controle remoto de softwares [2]. Eles possuem um conjunto razoável de botões deslizantes (*faders*) e rotativos (*knobs*), para ajustar os parâmetros dos softwares (volume, pan, etc), além de teclas para o acionamento remoto das funções de transporte (*play*, *rec*, *stop*, etc). Esses equipamentos podem ser conectados ao computador através de uma interface de MIDI ou, o que é mais comum nos dias de hoje, diretamente através de uma porta USB.

Existem vários tipos de superfícies de controle, sendo que as mais econômicas possuem somente botões rotativos (*knobs*) simples, outras possuem *faders* e *knobs*. As mais sofisticadas, no entanto, possuem *faders* motorizados que se ajustam automaticamente conforme a situação real que acontece no software; esses equipamentos possuem também *knobs* do tipo sem-fim (*rotary encoders*) e visor com informações adicionais da operação (veja Fig.2).

Figura 2 – Exemplo de superfície de controle

O objeto deste trabalho é um circuito eletrônico baseado em microcontrolador, concebido de maneira que possa ser usado como núcleo fundamental para diversas aplicações diferentes de controladores MIDI, tais como teclado controlador, superfície de controle, pedaleira para órgão e diversos outros tipos de funções.

1 ARQUITETURA DO HARDWARE

A premissa fundamental para o projeto foi a concepção de um núcleo básico que pudesse ser acoplado a diferentes módulos (outras placas) para a implementação de diversas funções de controle MIDI.

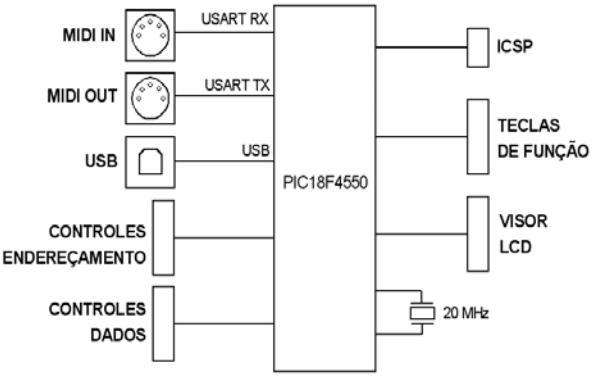


Figura 3 – Diagrama em blocos da placa principal

Este núcleo básico consiste de uma placa principal (Fig.3), onde estão o microcontrolador, os componentes e conectores para comunicação MIDI (Out, In), conector USB, o circuito de regulação da tensão de alimentação, um visor LCD e mais uma pequena placa contendo doze teclas de função. As placas dos módulos funcionais para as aplicações específicas podem ser conectadas à placa principal através de cabos multivias para endereçamento e transferência de dados.

1.1 Descrição dos circuitos

O componente principal do circuito é o microcontrolador PIC18F4550 [3], produzido pela Microchip. Além do fato desta linha de microcontroladores ser muito utilizada no Brasil, este modelo foi escolhido por oferecer as seguintes características:

- 32 pinos de entrada/saída digital
- até 13 canais A/D de 10 bits
- módulo interno de comunicação serial (USART)
- módulo interno de comunicação serial via USB 2.0
- arquitetura otimizada para compilador de linguagem C
- programação "in circuit" (ICSP)
- versão de 40 pinos PDIP, fácil de ser montada
- autoprogramável, via software
- baixo custo

O recurso de interfaceamento USB é importante porque permite conectar o hardware diretamente ao computador, dispensando a necessidade de uma interface de MIDI entre o controlador e o computador.

O circuito da placa principal é relativamente simples. Além do microcontrolador, há apenas os componentes para regulagem de tensão e o circuito para o interfaceamento MIDI. O baixo custo e a facilidade de obtenção dos componentes foram princípios básicos que delinearam a concepção do circuito, de forma a possibilitar a sua montagem sem dificuldades em nosso país.

O protótipo que foi construído já contém todos os componentes previstos para a placa principal. A única alteração efetuada após este protótipo foi o redesenho das placas de circuito impresso principal e das teclas de função, que tiveram suas dimensões levemente reduzidas.

A Fig. 4 mostra a placa principal, a placa de teclas de função e o visor LCD. Os componentes destacados são os seguintes:

1. Conectores de MIDI
2. Conector USB
3. Conector de alimentação (9Vcc)
4. Conector para programação (ICSP)
5. Conector de dados
6. Conector de endereçamento

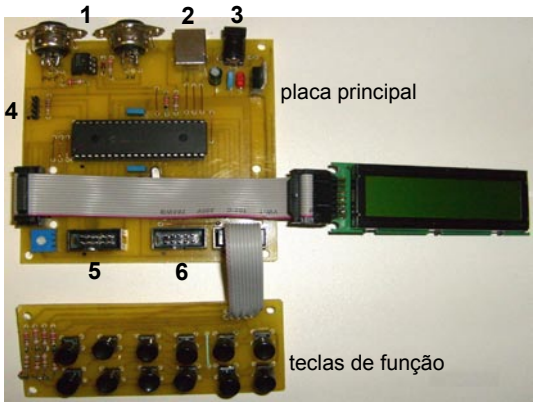


Figura 4 – Montagem do circuito do módulo básico

O custo do hardware básico (placa principal, teclas de função e visor LCD) está estimado em cerca de R$ 100. Para atender a este requisito de baixo custo, foi adotado inicialmente um visor LCD alfanumérico simples, com duas linhas de dezesseis caracteres. No entanto, este pode ser facilmente substituído por um visor gráfico para as aplicações onde seja necessário mostrar ao usuário informações com visual mais detalhado.

Há doze teclas de função, implementadas na forma de um teclado matricial. Na concepção inicial do projeto, quatro destas teclas servem para acessar as páginas de configuração do hardware, que aparecem no visor, e não transmitem mensagens de MIDI ao serem pressionadas, mas as outras teclas de função transmitem mensagens de MIDI. Na aplicação como superfície de controle, por exemplo, estas outras teclas são usadas para realizar as operações de transporte – play, stop, pause, etc., e também permitem a navegação do cursor na tela do software. Obviamente, as funções de todas as teclas podem ser alteradas modificando-se o firmware. Além disto, as funções das teclas que transmitem MIDI podem ser configuradas no aplicativo controlado, já que a maioria dos softwares de gravação (DAW) e sintetizadores virtuais permite mapear [4] as funções dos comandos de MIDI recebidos do hardware externo.

É importante destacar que na concepção do circuito também houve uma preocupação quanto às dimensões das placas de circuito impresso para que estas ficassem dentro de uma padronização, de maneira a facilitar a sua confecção e também a montagem final nos diversos tipos de aplicação de controlador MIDI.

1.2 Aplicação como superfície de controle

A primeira configuração idealizada para o projeto é uma superfície de controle bastante simples, para ser usada com softwares de gravação e sintetizadores virtuais. No entanto, o mesmo hardware pode ser facilmente adaptado para funcionar como um dispositivo de controle auxiliar de consoles digitais. Esta configuração, apresentada na Fig.5, é dotada de nove controles deslizantes (*faders*), oito botões rotativos (*knobs*) e ainda dezesseis teclas de pressão. Isto possibilita controlar diretamente os parâmetros de volume, pan, mute e solo de um software de gravação. O software interno permite o controle além de oito canais por meio de um esquema de *layers*, semelhante ao adotado nas consoles de mixagem digitais.

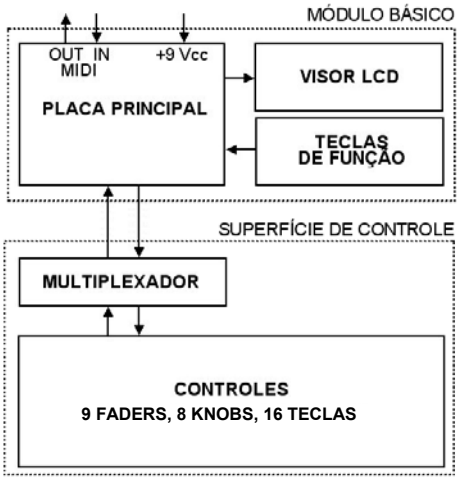


Figura 5 – Configuração para superfície de controle

O custo estimado para a produção em baixa escala de um controlador deste tipo é de cerca de R$ 350 (o custo real vai depender mais da montagem, design, etc.).

A Fig.6 mostra o protótipo da implementação de superfície de controle, com 9 faders e 8 knobs. Este protótipo mostrou-se bastante satisfatório para controlar os softwares de gravação Cakewalk Sonar e Cockos Reaper, além de diversos plug-ins de processamento de áudio e de síntese de sons.

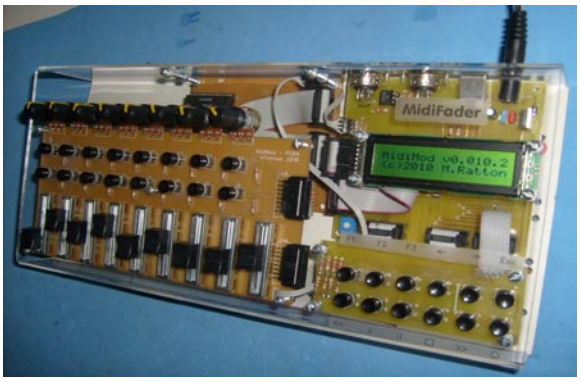


Figura 6 – Protótipo de superfície de controle

2 SOFTWARE INTERNO

Na versão inicial, o software foi escrito em C usando o compilador CCS [5], que oferece rotinas para recursos específicos dos microcontroladores da Microchip. Para aumentar o desempenho em aplicações mais críticas, poder-se-á substituir algumas das rotinas em C por assembly.

A estrutura do software implementado até o momento é bastante rudimentar, sendo o processo baseado em *pooling* (veja Fig.7). Embora este não seja o processo mais "elegante" para controle em tempo real, funcionou satisfatoriamente.

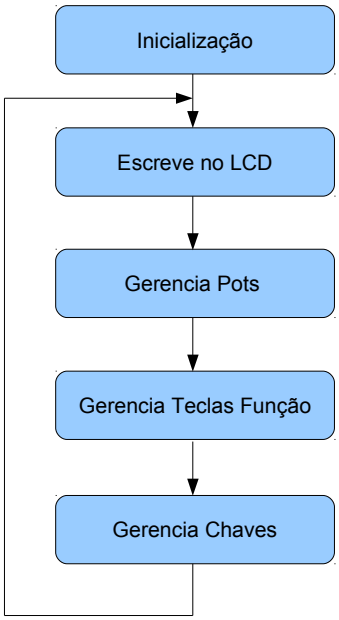


Figura 7 – Estrutura do software interno

O ciclo total de execução do software interno é de aproximadamente de 86 ms, sendo que a maior parte deste tempo (cerca de 37 ms) é dedicada ao gerenciamento dos potenciômetros (veja Fig.8), por causa do tempo necessário para aquisição dos seus valores analógicos.

Em relação ao tempo de espera para a estabilização da posição das teclas para a leitura de seus estados foi adotado o valor de 2 ms, que se mostrou adequado e compatível com estudos já realizados sobre o assunto [6].

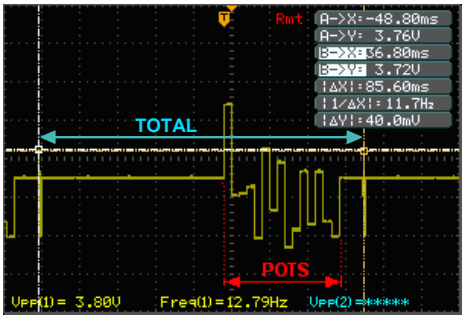


Figura 8 – Ciclo de execução do software interno, destacando o tempo total (85,6 ms) e o tempo necessário (36,8 ms) para a leitura dos valores dos potenciômetros

Para um processamento mais adequado, com menos risco de atrasos indesejáveis nas rotinas de cada processo, está prevista uma nova versão do software baseada em RTOS (*real-time operating system*). Esta melhora em desempenho será fundamental para a implementação de teclados musicais com dúzias de teclas que ofereçam sensibilidade ao toque (*key velocity*), onde a varredura do teclado é um fator crítico [7].

No protótipo da superfície de controle, as mensagens de MIDI adotadas para cada controle são as mesmas usadas nos teclados Edirol PCR [8], bastante populares. Os valores dos potenciômetros (knobs e faders) são medidos e tratados em 8 bits e depois são reduzidos para 7 bits para serem transmitidos via MIDI, na forma de mensagens de *control change* (MSB) com valores de 0 a 127. O acionamento das teclas do painel também é representado por mensagens de *control change*, com valores fixos (0 ou 127, dependendo do caso).

Foram usadas as seguintes codificações de controles:

Controle	Control Change	Canal de MIDI	Valor
fader Vol 1	17 (0x11)	1	0 – 127
fader Vol 2	17 (0x11)	2	0 – 127
fader Vol 3	17 (0x11)	3	0 – 127
fader Vol 4	17 (0x11)	4	0 – 127
fader Vol 5	17 (0x11)	5	0 – 127
fader Vol 6	17 (0x11)	6	0 – 127
fader Vol 7	17 (0x11)	7	0 – 127
fader Vol 8	17 (0x11)	8	0 – 127
fader Master	18 (0x12)	1	0 – 127
knob Pan 1	16 (0x10)	1	0 – 127
knob Pan 2	16 (0x10)	2	0 – 127
knob Pan 3	16 (0x10)	3	0 – 127
knob Pan 4	16 (0x10)	4	0 – 127
knob Pan 5	16 (0x10)	5	0 – 127
knob Pan 6	16 (0x10)	6	0 – 127
knob Pan 7	16 (0x10)	7	0 – 127
knob Pan 8	16 (0x10)	8	0 – 127
tecla Mute 1	80 (0x50)	1	127
tecla Mute 2	80 (0x50)	2	127
tecla Mute 3	80 (0x50)	3	127
tecla Mute 4	80 (0x50)	4	127
tecla Mute 5	80 (0x50)	5	127
tecla Mute 6	80 (0x50)	6	127
tecla Mute 7	80 (0x50)	7	127
tecla Mute 8	80 (0x50)	8	127
tecla Solo 1	81 (0x51)	1	127
tecla Solo 2	81 (0x51)	2	127
tecla Solo 3	81 (0x51)	3	127
tecla Solo 4	81 (0x51)	4	127
tecla Solo 5	81 (0x51)	5	127
tecla Solo 6	81 (0x51)	6	127
tecla Solo 7	81 (0x51)	7	127
tecla Solo 8	81 (0x51)	8	127

Controle	Control Change	Canal de MIDI	Valor
tecla Rew	82 (0x52)	8	127;0
tecla Rec	82 (0x52)	10	127;0
tecla FF	82 (0x52)	11	127;0
tecla Pause	82 (0x52)	12	127;0
tecla Stop	82 (0x52)	13	127;0
tecla Play	82 (0x52)	14	127;0

Para permitir uma realimentação visual ao usuário, o valor do controle que está sendo manipulado é mostrado no visor de LCD, em tempo real.

Dentre as funções previstas para serem incorporadas ao software do núcleo básico estão a recepção de dados MIDI e a atualização do firmware via MIDI ou USB.

Já no caso específico da superfície de controle, está prevista a implementação de alguns outros recursos, dentre eles a possibilidade do usuário configurar a mensagem de MIDI a ser transmitida por cada fader, botão e tecla, e armazenar esta configuração como "User".

3 CONCLUSÃO E PERSPECTIVAS

Com a tendência cada vez maior na implementação de instrumentos musicais e equipamentos de áudio na forma de software, faz-se necessário uma interface física para que se possa controlar esses softwares, que na sua grande maioria adota o protocolo MIDI como forma de acesso aos seus parâmetros. Dependendo do tipo de operação no software, a interface física com o usuário deve oferecer um conjunto de controles adequados, seja um teclado musical, botões rotativos, faders deslizantes ou outros tipos de controles.

A ideia central que motivou este desenvolvimento foi a disponibilização de um núcleo fundamental de hardware que possa ser usado como base para a implementação de diferentes tipos de controladores. Assim, a partir de uma placa básica, de custo bastante baixo, será possível construir produtos diferentes. Isto reduz o tempo e o custo de desenvolvimento dos produtos.

Alguns dos dispositivos controladores MIDI que poderão ser implementados a partir do hardware básico apresentado neste artigo são os seguintes:

- controlador de mixagem com faders convencionais
- controlador de mixagem com faders motorizados
- controlador de mixagem sensível ao toque
- teclado musical
- pedaleira musical (para órgão)
- controlador para DJ
- controlador percussivo com acelerômetro
- controlador de iluminação

Mesmo com o software interno ainda rudimentar, o primeiro protótipo criado em cima deste hardware básico – uma superfície de controle com 9 faders, 8 knobs e 23 teclas – confirmou os objetivos do projeto em termos de funcionalidade e custo.

Com o aprimoramento do software, o desempenho do hardware ficará ainda melhor, facilitando a implementação de aplicações mais críticas, como teclados musicais com sensibilidade ao toque, por exemplo.

O crescimento da economia brasileira tem permitido um número cada vez maior de consumidores também na área de áudio e música, o que abre espaço para alternativas nacionais em alguns segmentos. Graças à simplicidade e ao baixo custo dos componentes, além do fato de que parte do desenvolvimento já está realizada, o presente projeto pode ser atrativo para indústrias nacionais que queiram entrar em novos mercados.

Por outro lado, para estimular a implementação de mais aplicações para este hardware básico, uma possibilidade seria criar algum tipo de licenciamento que permita o desenvolvimento de forma colaborativa pela Internet, por exemplo.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] International MIDI Association. *MIDI Specification 1.0,* 1983.
[2] Ratton, Miguel B. *MIDI Total – Fundamentos e Aplicações.* Editora Música & Tecnologia, 2005.
[3] Microchip Technology Inc. *PIC18F4550 Data Sheet*, 2007.
[4] Francis, Geoffrey. *Reaper User Guide*. Cockos, 2010.
[5] Custom Computer Services Inc. *CCS C Compiler v.4*, 2007.
[6] Ganssle, Jack. *A Guide to Debouncing*. The Ganssle Group, 2004.
[7] Chamberlin, Hal. *Musical Applications of Microprocessors*. Hayden Books, 1985.
[8] Roland Corp. *PCR-300/500/800 Owner's Manual*, 2006.

Sociedade de Engenharia de Áudio
Artigo de Congresso
Apresentado no 9º Congresso de Engenharia de Áudio
15ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2011, 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 de banco de filtros digitais por janelas Kaiser para identificação de notas e acordes de violão usando imagens binarizadas por limiar de Otsu

M. M. Cordeiro Jr[1], H.M. de Oliveira[1]
[1]Universidade Federal de Pernambuco, Departamento de Eletrônica e Sistemas.
Recife, Pernambuco, 50.740-530, Brasil.
mauriciojunior19@hotmail.com, hmo@ufpe.br

RESUMO
Neste trabalho é proposto um novo procedimento para identificação de notas e acordes em registros de violão. O método apresentado utiliza técnicas de processamento de sinais de áudio e de imagem e pode ser dividido em cinco etapas, sendo elas: decomposição em componentes de energia utilizando banco de filtros digitais por janelas Kaiser; conversão matriz-imagem seguida de uma binarização por limiar de Otsu; ajuste a partir de técnicas de conectividade entre os pixels; sobreposição temporal e composição dos resultados. Testes de funcionamento usando registros de violão (gravados via sintetizador) foram realizados, obtendo taxas de acerto superiores a 99%.

0 INTRODUÇÃO

O estudo na área de análise e reconhecimento de padrões musicais vem se intensificando nas últimas décadas, devido ao avanço tecnológico em processamento de sinais. Dispositivos DSP (*Digital Signal Processor*) apresentam cada vez maior capacidade de execução de cálculos computacionais complexos, efetuando grande número de instruções em curtos intervalos de tempo. Considera-se natural, portanto, desenvolver *softwares* que emulem o comportamento do homem, capaz de interpretar estímulos sonoros. Estudos envolvendo processamento de sinais de áudio investigam a possibilidade de descrever sons gerados a partir de instrumentos musicais utilizando modelos matemáticos, desta forma, tornando possível analisá-los e identificá-los.

Na literatura existe uma vasta quantidade de pesquisas realizadas com este propósito, cada qual, utilizando técnicas distintas, algumas das quais são enunciadas: em [1], o autor propõe uma estratégia de classificação baseada em redes neuronais *feed-forward* de múltiplas camadas, treinadas com representações frequenciais com *transformadas com Q-constante* (CQT); A técnica desenvolvida em [2] também é baseada no treinamento de uma rede que extrai um vetor característico contendo as energias da nota (fundamental e seus primeiros harmônicos), sendo a rede constituída por duas camadas – a camada de segmentação e a camada de reconhecimento; em [3], com objetivo similar, aborda-se o uso da transformada rápida de *Fourier* para localização dos picos de frequência do sinal; em [4] os autores deste artigo enunciaram outra técnica utilizada para reconhecimento de notas musicais isoladas no tempo, utilizando um diagrama de estados para segmentação das mesmas e a transformada rápida de *Fourier* para o reconhecimento. Em 2006, K. Lee [5] propôs um método de reconhecimento de acordes o qual denominou de EPCP (*Enhanced Pitch Class Profile*). Em [6,7,8], os autores propuseram outro método com base em cadeias de Markov ocultas. Em [9], Harte e Sandler usaram um *cromagrama* de 36 bits para encontrar o valor de ajuste de entrada de áudio, usando a distribuição das posições dos picos. Em [10], Fujishima também descreveu um método para reconhecimento de acordes, baseado no mesmo principio seguido por Harte e Sandler.

O reconhecimento de acordes pode ser usado como uma ferramenta na elaboração de aplicativos para transcrição automática de música [11].

1 VISÃO GLOBAL DO SISTEMA

Neste trabalho é proposta uma nova abordagem para a identificação de notas e acordes em registros de violão gerados a partir da gravação em um sintetizador digital, *software*, ou por meio do próprio instrumento. O método apresentado combina técnicas de processamento de sinais de áudio e de imagem e pode ser dividido em cinco etapas: na primeira, o arquivo de áudio é processado em um banco de filtros digitais por janelas *Kaiser* [12], compostos por 84 filtros passa-faixas cujas frequências centrais são as próprias frequências das notas musicais em escala temperada (a passagem pelo banco de filtros decompõe o sinal de áudio em componentes de energia e os armazenam em uma matriz); na segunda etapa, a matriz gerada é convertida em uma imagem em tons de cinza com 256 níveis, sendo então binarizada por limiar de *Otsu* [13]; na terceira, são realizados ajustes na imagem binária utilizando técnicas de conectividade entre os pixels [13], a fim de remover todas as oscilações presentes e manter apenas os intervalos nos quais as notas/acordes permanecem fixas; na quarta, analisa-se sua sobreposição temporal, com o objetivo de distinguir entre nota isolada, notas sobrepostas ou um acorde; finalmente, na última etapa, é feita uma análise na saída expressa na etapa anterior a fim de compor em tela os acordes ou notas reconhecidas pelo sistema. Vale mencionar que o sistema foi projetado para escala temperada.

1.1 Objetivo

Desenvolver um *software* em plataforma *Matlab*[1] capaz de ler um arquivo de áudio gravado no formato *wav* o qual contém notas e/ou acordes musicais (com ou sem sobreposição temporal) e identificá-las. A proposta é introduzir uma técnica que seja eficaz tanto no tempo quanto na frequência, sem utilização de nenhuma rede neuronal, e também sem trabalhar diretamente com a transformada de *Fourier*, pois esta limita a análise ao domínio frequencial, perdendo informações temporais.

A ideia consiste em projetar um banco de filtros digitais passa-faixas que opere em toda faixa de frequência a qual o sistema estará projetado para reconhecer. Cada filtro é um "filtro passa-faixa" com largura de banda de passagem proporcional à escala temperada e com frequência central na "nota alvo".

O banco de filtros se baseia no mesmo princípio da transformada de *Wavelets* [14], que, ao contrário da *transformada de Fourier* [15], faz uma análise com resolução adequada no domínio tempo–frequência do sinal analisado.

2 METODOLOGIA

O sistema para reconhecimento de notas e acordes musicais consiste em uma técnica desenvolvida utilizando um banco de filtros digitais projetados usando janelas de *Kaiser* capaz de identificar uma nota musical pura (DÓ-RÉ-MI-FÁ-SOL-LÁ-SI), assim como seus acidentes musicais (sustenidos ou bemóis), sendo um total de 12 notas em qualquer escala que se repetem à medida que se altera a escala. Subir uma escala implica duplicar a freqüência das notas. Também esta mesma técnica permite a identificação de notas musicais compostas (acordes musicais), que de acordo com suas combinações, podem formar uma gama deles. A técnica é vantajosa, pois faz simultaneamente uma análise nos domínios tempo–frequência, sendo análoga ao princípio da transformada que utiliza banco de filtros *wavelets*.

Na escala temperada, tem-se a seguinte classificação, conforme ilustrado na Tabela 1:

Tabela 1: Frequências Sonoras: observe que ao passar de uma nota para outra, sua frequência fundamental é multiplicada por $\sqrt[12]{2}$.

Notas		1	2	3	4	5	6	7
0	C	32,703196	65,40639	130,8128	261,6256	523,2511	1046,502	2093,005
1	C#	34,647829	69,29566	138,5913	277,1826	554,3653	1108,731	2217,461
2	D	36,708096	73,41619	146,8324	293,6648	587,3295	1174,659	2349,318
3	D#	38,890873	77,78175	155,5635	311,127	622,254	1244,508	2489,016
4	E	41,203445	82,40689	164,8138	329,6276	659,2551	1318,51	2637,02
5	F	43,653529	87,30706	174,6141	349,2282	698,4565	1396,913	2793,826
6	F#	46,249303	92,49861	184,9972	369,9944	739,9889	1479,978	2959,955
7	G	48,99943	97,99886	195,9977	391,9954	783,9909	1567,982	3135,964
8	G#	51,913088	103,8262	207,6524	415,3047	830,6094	1661,219	3322,438
9	A	55	110	220	440	880	1760	3520
10	A#	58,270471	116,5409	233,0819	466,1638	932,3275	1864,655	3729,31
11	B	61,735413	123,4708	246,9417	493,8833	987,7666	1975,533	3951,066

Para implementação deste sistema foram utilizados arquivos no formato *wav* do tipo monofônico e gravados a uma taxa de amostragem de 8 kHz, adequada para registros produzidos por violão.

Embora outros instrumentos possam ser usados, os testes de validação foram conduzidos exclusivamente usando gravações de violão. A identificação das notas/acordes se faz em escala temperada (incluindo # e β), através de um mapeamento de identificação tonal, conforme apresentado nas Tabelas 2,3 e 4.

Tabela 2: Função identificadora de notas musicais em função das suas respectivas cifras com código representativo de cada uma delas.

Cifras	C	D	E	F	G	A	B
Notas	Dó	Ré	Mi	Fá	Sol	Lá	Si
CodMusical	0	2	4	5	7	9	11
Cifras	C#	D#		F#	G#	A#	
Acidentes	Dó#	Ré#		Fá#	Sol#	Lá#	
CodMusical	1	3		6	8	10	

Tabela 3: Exemplos de acordes básicos formados de tônica, terça e quinta. Com qualquer escala maior ou menor, forma-se o campo harmônico semelhante ao indicado na tabela anterior.

[1] *Registered trademarks of the MathWorks.™*

Tabela 4: Exemplos de acordes dissonantes básicos formados por cinco notas.

Acordes Dissonantes	Escala: Dó	Ré	Mi	Fá	Sol	Lá	Si	Dó	Ré	Mi	Fá	Sol	Lá
C7M/9	♪		♪		♪		♪		♪				
Dm7/9		♪		♪		♪		♪		♪			
Em7			♪		♪		♪		♪				
F7M/9				♪		♪		♪		♪		♪	
G7/9					♪		♪		♪		♪		♪
Am7						♪		♪		♪		♪	
Bdim7							♪		♪		♪		♪

Uma vasta quantidade de testes foi conduzida com arquivos gerados através do *software Curso de Violão ETM 6.3*[2], simulando notas e acordes produzidas por um violão (com taxa de acerto acima de 99%). Implementou-se toda a abordagem usando a plataforma *Matlab*. Como mencionado, o processo de identificação pode ser dividido em cinco etapas, conforme diagrama de blocos mostrado na Figura 1.

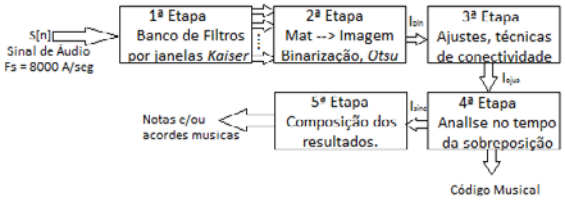


Figura 1: Diagrama de blocos do processo de reconhecimento.

2.1 Decomposição

A primeira etapa consiste em decompor o sinal digital amostrado com uma taxa de 8 kHz em 84 componentes a serem dispostas em uma matriz de tamanho 84×*N*, em que *N* denota o número de amostras armazenados.

No primeiro passo desta etapa (1° bloco da Figura 1) é inicializado um banco de filtros digitais por janelas de *Kaiser,* tal como explanado a seguir:

2.1.1. Implementação do filtro digital usando janela de Kaiser

As magnitudes mais expressivas, no domínio frequencial, concentram-se em torno da condição $\omega = 0$. A atenuação (diferença entre as alturas dos lóbulos principal e secundário) é um importante para a escolha da janela. A solução adotada envolveria funções de ondas dificilmente computáveis; portanto, não é uma ideia atrativa para ser implementada. *Kaiser* [12] (1966, 1974) descobriu que uma janela quase ótima poderia ser formada usando a função de *Bessel* de ordem zero, uma função facilmente computável. A janela de *Kaiser* é definida através da relação:

$$\omega[n] = \begin{cases} I_0\left\{\beta\left[1-\left(\left(n-\alpha\right)/\alpha\right)^2\right]^{1/2}\right\}, 0 \le n \le M \\ 0, caso\ contrário \end{cases} \quad (1)$$

em que $\alpha = M/2$, e $I_0(.)$ representa a função de *Bessel* de ordem zero do primeiro tipo [12].

A janela de *Kaiser* possui dois parâmetros: O comprimento (M+1) e o parâmetro de forma β.

Variando-se (M+1) e β, tanto o comprimento da janela como a forma podem ser ajustados. Assim, conhecendo as especificações do filtro, seja F_s (frequência de amostragem), A (Atenuação/Ganho), f_c (frequência de corte) e Δf (largura da região de transição) pode-se ajustar os parâmetros do filtro a partir das seguintes fórmulas:

$$\beta = \begin{cases} 0{,}1102(A-8{,}7), & A > 50 \\ 0{,}5842(A-21)^{0{,}4} + 0{,}07886(A-21), & 21 \le A \le 50 \\ 0{,}0, & 21 < A. \end{cases} \quad (2)$$

$$M = \frac{A-8}{2{,}285\ \Delta\omega}, A \ge 21, \quad (3)$$

$$\Delta\omega = \omega_p - \omega_s, \quad (4)$$

em que ω_p e ω_s são as frequências de corte (inferior e superior) do passa-faixa.

A proposta consiste em projetar um banco de filtros passa-faixas, de atenuação igual a -65 dB, utilizando janelas de *Kaiser*. Cada janela espectral deverá estar centrada numa frequência correspondente à frequência característica de uma nota musical (Dó, Dó#, Ré, ..., Si), repetindo-se até um total de sete oitavas musicais, conforme Tabela 1.

A largura espectral de cada janela não é fixa; esta deverá crescer à medida que a frequência aumenta (filtros com fatores de qualidade constantes), tendo largura correspondente a um semitom. A faixa da região de transição foi assumida constante, com largura de 20 Hz.

Como as frequências indicadas na Tabela 1 crescem em razão de uma P.G. (Progressão Geométrica) à medida que se avança um semitom, é possível definir:

$$f_N(i) := 32{,}703196.(\sqrt[12]{2})^i\ Hz. \quad (5)$$

Sendo a frequência $f_N(0)$ = 32,703196 Hz correspondente ao primeiro Dó (Escala 1), tem-se:

- <u>Frequência central de cada janela</u>

$$f_c(i) := 32{,}703196.(\sqrt[12]{2})^i\ Hz\ . \quad (6)$$

- <u>Largura espectral de cada janela</u>

$$\delta f(i) = f_s(i) - f_l(i) = 31{,}7308795.\left[(\sqrt[12]{2})^{i+1} - (\sqrt[12]{2})^i\ \right]Hz. \quad (7)$$

- <u>Frequências de corte (nota inferior)</u>

$$f_{inf}(i) = 31{,}7308795.(\sqrt[12]{2})^i\ Hz. \quad (8)$$

- <u>Frequências de corte (nota superior)</u>

$$f_{sup}(i) = 31{,}7308795.(\sqrt[12]{2})^{i+1}\ Hz. \quad (9)$$

- <u>Resposta ao impulso do filtro digital</u>

$$h[i,n] = \frac{sen[2\pi f_c(i).(n-\alpha)]}{\pi(n-\alpha)}.\omega[n]. \quad (10)$$

O código musical (#CM) e sua respectiva escala musical (#EM) podem ser vistos na Tabela 1 e calculados a partir do índice *i* das equações (11) e (12), usando a fórmula de recorrência:

$$\#\ CM = i\ mod\ 12, \quad (11)$$

[2] *Registrado por Airton de Igaratinga, MG, Brasil.* ™

$$\# EM = \left\lfloor 1 + \frac{i}{12} \right\rfloor, \quad (12)$$

em que $\lfloor x \rfloor$ denota o maior inteiro menor do que x.

2.1.2. Construção de uma matriz utilizando filtros de Kaiser

O sinal digital de áudio pode ser representado como um vetor com N amostras de mesmo comprimento, e cada amostra está espaçada por um intervalo de 1/8000 segundos e denota-se o sinal de entrada por $S[n]$. Desta forma, pode-se construir uma matriz de tamanho $84 \times N$ efetuando-se a convolução discreta entre cada filtro $h[i,n]$ e o sinal $S[n]$. Assim, a matriz gerada assume a seguinte forma:

$$M[i,n] = h[i,n] * S[n]; 1 \leq i \leq 84, 0 \leq n \leq N-1. \quad (13)$$

Sendo cada elemento da matriz denotado por $a_{i,n}$, em que i e n identificam à qual linha e coluna, respectivamente, pertence o elemento. As componentes de energia podem ser obtidas elevando-se cada elemento da matriz ao quadrado:

$$M_E[i,n] = (a_{i,n}^2); 1 \leq i \leq 84, 0 \leq n \leq N-1. \quad (14)$$

2.2 Conversão Matriz - Imagem Binária

A segunda etapa do processo consiste em duas partes: primeiramente, converte-se a matriz $M_E[i,n]$ em uma *imagem digital* monocromática, com 256 níveis de tons de cinza; em seguida, faz-se uma binarização desta utilizando o limiar de *Otsu* [13], fazendo com que as componentes com alto nível de energia assumam nível lógico 1, enquanto que componentes com baixo nível de energia assumam o nível lógico 0.

Vale ressaltar que notas musicais puras devem apresentar uma grande amplitude na componente de frequência fundamental na qual elas são definidas. Acordes também devem apresentar grandes amplitudes nas respectivas frequências de suas notas. A binarização faz a separação entre as baixas e altas amplitudes.

2.2.1. Conversão da Matriz de Áudio para Imagem Monocromática

Para conversão da matriz $M_E[i,n]$ em uma imagem digital monocromática com 256 níveis de energia é utilizado um algoritmo de varredura que localiza o valor de maior energia contido nesta matriz e a partir dele, calcula o fator de conversão da seguinte forma:

$$F_{cov} = \frac{255}{\max(M_E[i,n])}. \quad (15)$$

Assim, a matriz que representa a imagem é convertida utilizando a expressão:

$$M_{Gray} = \left[F_{cov} \begin{pmatrix} [a_{1,1}^2] & \cdots & [a_{1,N}^2] \\ \vdots & \ddots & \vdots \\ [a_{84,1}^2] & \cdots & [a_{84,N}^2] \end{pmatrix} \right], \quad (16)$$

em que $[x]$ denota o inteiro mais próximo de x (arredondamento). Vale ressaltar que cada elemento desta matriz é quantizado por um número binário de 8 bits (1 byte/elemento), alocando na memória um espaço equivalente a $84 \times N$ bytes.

2.2.2. Binarização por limiar de *Otsu*

O processo de binarização ocorre aplicando-se a limiarização de Otsu [13] na matriz M_{Gray}.

O método de Otsu é um método de limiarização global para escolher o melhor *threshold*. Este se baseia no histograma normalizado (da imagem) como uma função de densidade de probabilidade discreta, na forma, $p_r(r_q) = m_q/m,\ q = 0, 1, 2, \ldots, L-1$; em que m é o número total de pixels na imagem, m_q é o número de pixels que tem intensidade r_q e L representa o número total de níveis de intensidade na imagem, neste caso, 256.

O método de Otsu constrói duas classes denotadas por C_0 e C_1 e escolhe para limiar um valor de k como sendo o nível de intensidade para o qual $C_0=[0,1,....,k-1]$ e $C_1=[k,k+1,...,L-1]$), maximizando a variância entre classes σ^2_B, que é definida como:

$$\sigma^2{}_B := \prod_0(\mu_0 - \mu_T)^2 + \prod_1(\mu_0 - \mu_T)^2, \quad (17)$$

em que

$$\prod_0 = \sum_{q=0}^{k-1} p_q(r_q), \quad (18)$$

$$\prod_1 = \sum_{q=k}^{L-1} p_q(r_q), \quad (19)$$

$$\mu_0 = \sum_{q=0}^{k-1} \frac{q p_q(r_q)}{\prod_0}, \quad (20)$$

$$\mu_1 = \sum_{q=k}^{L-1} \frac{q p_q(r_q)}{\prod_1}, \quad (21)$$

$$\mu_T = \sum_{q=0}^{L-1} q p_q(r_q). \quad (22)$$

O limiar de Otsu é o valor calculado k, dessa maneira a matriz gerada assumirá a forma:

$$M_{Bin} = \begin{bmatrix} b_{1,1} & \cdots & b_{1,N} \\ \vdots & \ddots & \vdots \\ b_{84,1} & \cdots & b_{84,N} \end{bmatrix}, \quad (23)$$

em que

$$b_{i,n} = \begin{cases} 0, se\ a_{i,n}^2 \leq k \\ 1, se\ a_{i,n}^2 > k. \end{cases} \quad (24)$$

2.3 Conexão entre os pixels

É bastante natural esperar que cada linha da matriz binarizada apresente oscilações (alternâncias entre 0's e 1's) devido à natureza da energia instantânea do sinal analisado ser oscilatória. Uma vez que a energia instantânea do sinal é definida como sendo o módulo elevado ao quadrado do mesmo, logo, a frequência de oscilação da energia deverá ser estimada em média como sendo o dobro da frequência da nota à qual se deseja analisar.

Analisando "componente a componente" de acordo com a Tabela 1, a nota de menor frequência está em torno dos 32 Hz, ao passo que a maior por volta dos 4 kHz; isto significa que a cada um segundo podem aparecer de 64 a 8000 oscilações. Assim, sabendo que existem 8000 amostras em um segundo, pode ocorrer desde 1 nível lógico baixo para cada 1 nível lógico alto até 128 níveis lógico baixo para cada 128 níveis lógicos alto.

Na ausência de nota é assumido nível lógico baixo durante todo tempo em que esta ocorrer (cenário de baixo ruído).

A técnica de conectividade entre pixels [13] permite remover todas as oscilações presentes analisando linha a linha a matriz binária. Ela é baseada numa máscara de tamanho 1×150; assim, havendo uma conectividade de pixels (nível lógico alto ou baixo) de comprimento superior a 150, este é caracterizado como presença de silêncio. Caso contrário, caracteriza-se como presença de informação (nota musical) nas componentes da linha analisada.

Portanto, todo trecho do sinal que tiver oscilações que satisfaçam critério enunciado no parágrafo anterior, assumirá nível lógico alto (presença de informação), caso contrário, o nível lógico permanecerá inalterado (ausência de informação).

2.4 Sobreposição temporal

O processo de sobreposição é utilizado como critério de decisão para resolver se a sobreposição temporal entre duas ou mais notas será caracterizada como acorde ou simplesmente como notas sobrepostas. Caso o critério de decisão decida por acorde musical, este deverá ser sincronizado, caso contrário, todas as sobreposições deverão ser removidas.

Para melhor entendimento da técnica, supõe-se a ocorrência do acorde Dó maior (de acordo com a Tabela 3) a partir de t_1 e das notas isoladas Ré, Mi e Sol a partir de t_4, conforme Figura 2.

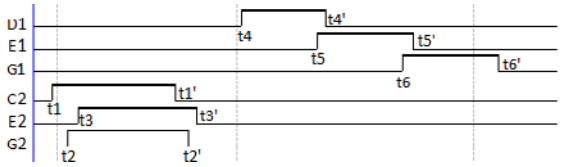


Figura 2: Exemplo ilustrativo de identificação de notas/acordes em trecho de um sinal qualquer.

Para o exemplo, o algoritmo desenvolvido visa analisar as sobreposições ocorridas entre C2, E2 e G2 e entre D1, E1 e G1 a fim de classificar estas como um acorde musical ou meramente como um conjunto de notas sobrepostas. A técnica utilizada se baseia em analisar a largura das notas sobrepostas. Isto é feito efetuando-se um *AND* lógico entre elas, como segue:

Na escala 1, tem-se:

$$\#1 = D1\ and\ E1\ and\ G1 = 0, \forall t$$

Na escala 2, tem-se:

$$\#2 = C2\ and\ E2\ and\ G2 = \begin{cases} 1, se\ t_3 \leq t \leq t_1' \\ 0, \text{caso contrário.} \end{cases}$$

Desta forma, as três notas presentes na escala 1 formam um conjunto de notas puras e sobrepostas; estas deverão ter suas sobreposições eliminadas, ao passo que as três notas presentes na escala 2 apresentam largura de sobreposição igual à diferença t_1'–t_3, que se for superior a um determinado limiar e ainda satisfizer o critério de formação dos acordes (Tabela 3), a mesma deverá ser classificada como um acorde musical e então sincronizada, conforme ilustra a Figura 3.

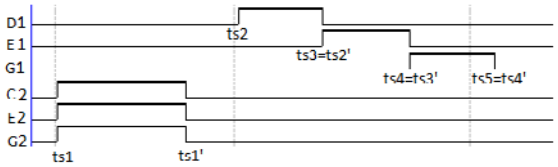


Figura 3: Exemplo de identificação de notas/acordes em trecho de um sinal qualquer sincronizado e ajustado.

De uma forma geral, para *m* sobreposições encontradas, assumindo que o método somente identifica acordes formados por 3,5 ou 7 notas, as *m* sobreposições devem ser combinadas de acordo com o seguinte critério:

- Caso $m \geq 7$, realiza-se a função "*and*" em todas as combinações *m* tomada 7 a 7, 5 a 5 e 3 a 3, a fim de identificar algum acorde presente entre elas.
- Caso $5 \leq m < 7$, realiza-se a função "*and*" em todas as combinações *m* tomada 5 a 5 e 3 a 3 possíveis.
- Caso $3 \leq m < 5$, realiza-se a função "*and*" em todas as combinações *m* tomada 3 a 3 possíveis.
- Caso $m < 3$, não há acorde neste trecho de sobreposição, apenas a presença de notas puras.

2.5 Composição dos resultados

A 5ª e última etapa deste processo é a mais simples de todas. Ela é responsável pela composição dos resultados obtidos na saída da etapa anterior (cf. Figura 1), a fim de exibir ao usuário final um resultado mais amigável.

Pela 4ª etapa, sabe-se que a matriz final sincronizada apresenta níveis lógicos alto em instantes de tempo (colunas) onde a informação está presente e o número da linha da matriz informa qual nota está ali presente, inclusive sua escala musical, conforme identificado nas "equações 11 e 12". A partir de então, pode-se formar os resultados a serem exibidos em tela.

No exemplo ilustrado na Figura 2, chega-se à Figura 3, e a partir dela podem-se inferir facilmente os resultados, assim sendo:

$$Cifras = \begin{cases} Acorde\ Dó\ Maior, & \text{quando } t_{s1} \leq t \leq t_{s1}'; \\ Nota\ Ré_1, & \text{quando } t_{s2} \leq t \leq t_{s2}'; \\ Nota\ Mi_1, & \text{quando } t_{s3} \leq t \leq t_{s3}'; \\ Nota\ Sol_1, & \text{quando } t_{s4} \leq t \leq t_{s4}'; \\ \text{Ausência de notas / acordes, caso contário.} \end{cases}$$

3 CONCLUSÕES

Este trabalho introduz uma nova técnica para a detecção e reconhecimento de notas ou acordes musicais (com ou sem sobreposição temporal) produzidas por um violão e descreve a implementação de um método inovador em relação àquele publicado em [4] pelos mesmos autores. Vários métodos introduzidos podem ser adaptados para uso

em cenários como classificação de instrumentos e transcrição automática. [11, 16].

A técnica foi inteiramente implementada e simulada em *Matlab*, e foram realizados cinquenta testes, com duração entre 1" a 35", resultando em taxa de acerto superior a 99% e precisão temporal em torno de 250 ms. O sistema conduz uma alternativa atrativa para aprendizado e serve como base para afinação de violão. A extensão para outros instrumentos parece ser relativamente simples. De fato, os testes incluíram acordes produzidos por outros instrumentos (e.g., piano, cello) sem degradar a taxa de acerto.

4 REFERÊNCIAS BIBLIOGRÁFICAS

[1] A.L. Szczupack e L.W.P. Biscainho, "Identificação de notas musicais em registros de violão solo" em anais do *7º Congresso de Engenharia e Áudio*, AES Brasil 2009, pp. 26 – 33.

[2] A.S. Ogasawara, "*Reconhecedor de notas musicais em sons polifônicos*", Departamento de Eletrônica e Computação, UFRJ, Abril de 2008.

[3] S.A.S. Ferreira, "*Sistema especialista para reconhecimento de acordes musicais em tempo real para violão elétrico utilizando técnicas de DSP*", Dissertação de Mestrado, UFBA, Dezembro de 2006.

[4] M.M. Cordeiro Jr e H.M. de Oliveira, "Sistema de Reconhecimento Automatizado de Notas Musicais Isoladas" em *anais XXXII CNMAC*, 2009, Vol. 2, Mato Grosso, Brasil, pp. 1038 – 1039.

[5] K. Lee, *"*Automatic Chord Recognition from Audio Using Enhanced Pitch Class Profile*"*, in *proc. of the International Computer Music Conference*, 8p., 2006.

[6] K. Lee and M. Slaney, *"*Automatic chord recognition from audio using an HMM with supervised learning*"*, *Proc. ISMIR*, 5p., 2006.

[7] K. Lee and M. Slaney, *"*Automatic chord recognition from audio using a supervised HMM trained with audio-from-symbolic data*"*, in *proc. of the 1st ACM workshop on Audio and music computing multimedia*, pp.11–19, 2006.

[8] H-T. Cheng et al., *"*Automatic chord recognition for music classification and retrieval*"*, in *IEEE International Conference on Multimedia and Expo*, pp. 1505–1508, 2008 doi 10.1109/ICME.2008.4607732.

[9] C. Harte and M. Sandler, *"*Automatic chord identification using a quantised chromagram*"*, AES, in *Proceedings of the Audio Engineering Society*. Spain: Audio Engineering Society, May 2005.

[10] T. Fujishima, "Realtime chord recognition of musical sound: A system using Common Lisp Music", In *Proc. of the Int. Computer Music Conf.*, Beijing: International Computer Music Association, 1999.

[11] T.F. Tavares, J.G.A. Barbedo and A. Lopes, "Towards the evaluation of automatic transcription of music" em anais do *6º Congresso de Engenharia e Áudio*, AES Brasil 2008, pp. 96 – 99.

[12] A.V. Oppenheim, R.W. Schafer, with J.R. Buck, "Filter Design Techniques" em *Discrete-Time Signal Processing*, 2ª Ed. Upper Saddie River: Prentice Hall, 1999, cap. 7, seção 7.2.3, pp. 474 – 477.

[13] R.P. Gonzales e R.C. Woods, *Processamento Digital de Imagens*, 3ª Ed. São Paulo: Prentice Hall, 2010.

[14] H.M. de Oliveira, "Wavelets: Uma Evolução na Representação de Sinais" em *Análise de Sinais para Engenheiros - uma Abordagem via Wavelets,* 1ª Ed. Rio de Janeiro, Brasil: Brasport, 2007.

[15] H. M. de Oliveira, "Espectro de Sinais" em *Análise de Fourier e Wavelets: Sinais Estacionários e não Estacionários*, 1ª Ed., Recife, Brasil: Editora Universitária da UFPE, 2007.

[16] J.G.A. Barbedo and A. Lopes, "Short-Term Classification of Musical Instruments: A Critical View" em anais do *6º Congresso de Engenharia e Áudio*, AES Brasil 2008, pp. 100 – 107.

Sociedade de Engenharia de Áudio
Artigo de Congresso

Apresentado no 9º Congresso de Engenharia de Áudio
15ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2011, 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.

Pitch tracking in polyphonic audio by clustering local fundamental frequency estimates

Martín Rocamora,[1] and Pablo Cancela[1]

[1] Universidad de la República, FING, IIE, Departamento de Procesamiento de Señales
Montevideo, Julio Herrera y Reissig 565, C.P. 11200, Uruguay

rocamora@fing.edu.uy, pcancela@fing.edu.uy

ABSTRACT

A novel way of performing pitch tracking by means of clustering local fundamental frequency (f0) candidates is described. The technique is based on an existing pitch salience representation for polyphonic music called F0gram which relies on the Fan Chirp Transform [1]. The grouping is performed by applying a Spectral Clustering method, since it can handle filiform shapes such as pitch contours. The approach seems appealing since many sound sources can be tracked simultaneously and the number of contours and sources is derived from the data. Results of a melody detection evaluation indicate the introduced method is promising, despite that various aspects of the technique deserve further work.

1 INTRODUCTION

Multiple fundamental frequency (f0) estimation is one of the most important problems in music signal analysis and constitutes a fundamental step in several applications such as melody extraction, sound source identification and separation. In our previous work [1] the Fan Chirp Transform (FChT) was applied to polyphonic music analysis, a technique based on decomposing the audio signal into harmonically related chirps. In addition, a pitch salience representation for music analysis called F0gram was proposed that provides a set of local fundamental frequency candidates together with a pitch change rate estimate for each of them. To continue the analysis temporal integration of local pitch candidates has to be performed, which is the problem tackled herein. There is a vast amount of research on pitch tracking in audio, often comprising an initial frame by frame f0 estimation followed by formation of pitch contours exploiting estimates continuity over time. Techniques such as dynamic programming, linear prediction, hidden Markov models, among many others (see [2] for a review), were applied to the temporal tracking.

The herein proposed technique for pitch contours formation does not involve a classical temporal tracking algorithm. Instead the pitch tracking is performed by unsupervised clustering of F0gram peaks. The spec-

tral clustering method [3] is selected for this task as it imposes no assumption of convex clusters, thus being suitable for filiform shapes such as pitch contours. The pitch change rate estimates provided by the FChT analysis play an important role in the definition of similarity between pitch candidates. The clustering is carried out within overlapped observation windows corresponding to several signal frames. Then contours are formed by simply joining clusters that share elements. This short-term two-stage processing proved to be more robust than aiming a straightforward long-term clustering.

There are very few applications of spectral clustering for tracking a sound source. Blind one-microphone separation of two speakers is tackled in [4] as a segmentation of the spectrogram. A method is proposed to learn similarity matrices from labeled datasets. Several grouping cues are applied such as time-frequency continuity and harmonicity based. A simple multiple pitch estimation algorithm is part of the feature extraction. The mixing conditions are very restrictive (equal strength and no reverberation). Performance is assessed through a few separation experiments.

Clustering of spectral peaks is applied in [5], for partial tracking and source formation. Connecting peaks over time to form partials and grouping them to form sound sources is performed simultaneously. The problem is modeled as a weighted undirected graph where the nodes are the peaks of the magnitude spectrum. The edge weight between nodes is a function of frequency and amplitude proximity (temporal tracking) and a harmonicity measure (source formation). Clustering of peaks across frequency and time is carried out for windows of an integer number of frames ($\sim$ 150 ms) using a spectral clustering method. Clusters from different windows are not connected for temporal continuation. The two more compact clusters of each window are selected as the predominant sound source.

The rest of this document is organized as follows. In the next section the application of the FChT to build a pitch salience representation is briefly discussed. Section 3 summarizes the fundamental aspects of the Spectral Clustering methods. Section 4 describes the proposed algorithm for pitch contours formation. Experimental results are presented in section 5. The paper ends with a critical discussion on the present work.

2 PITCH SALIENCE COMPUTATION

The Short Time Fourier Transform is the standard method for time-frequency analysis. In this representation the signal is supposed to be stationary within the analysis frame. However, music audio signals such as a singing voice typically exhibit rapid pitch fluctuations that are troublesome for the analysis.

A different approach to perform the analysis is considering the projection over frequency modulated sinusoids (chirps), in order to obtain a non-Cartesian tiling of the time-frequency plane. The modulation rate of the chirp can be selected in order to closely match the pitch change rate of the musical sound source. Among the chirp based transforms, the FChT offers optimal resolution simultaneously for all the partials of a harmonic chirp (harmonically related chirps). This is well suited for music analysis because many sounds have an harmonic structure. The FChT can be formulated as [1],

$$X_w(f,\alpha)=\int_{-\infty}^{\infty} x(t)\; w(\phi_\alpha(t))\; \phi'_\alpha(t)\; e^{-j2\pi f \phi_\alpha(t)} dt \quad (1)$$

where $\phi_\alpha(t) = (1 + \frac{1}{2}\alpha t)\, t$, is a time warping function and $w(t)$ stands for a time limited window, such as Hann. Notice that by the variable change $\tau = \phi_\alpha(t)$, the formulation can be regarded as the Fourier Transform (FT) of a time warped version of the signal $x(t)$, which enables an efficient implementation based on the FFT. Given the previous formulation, the FChT representation of a harmonic linear chirp is composed of deltas convolved with the FT of the window, provided the appropriate chirp rate α is applied in the warping.

Computing the FChT for consecutive short time signal frames a time-frequency representation in the form of a spectrogram can be built. For polyphonic music analysis the approach followed in [1] is to compute several FChT instances with different α values. This produces a multidimensional representation made up of various time-frequency planes.

The time-frequency representation described above can be applied to obtain a detailed description of the melodic content of an audio signal, as proposed in [1]. Pitch salience is computed for each signal frame in a certain range of fundamental frequency values. Given S_α the power spectrum produced by the FChT for a chirp rate value α, salience of fundamental frequency f_0 is obtained by gathering the log-spectrum at the positions of the corresponding harmonics, $\rho(f_0,\alpha) = \frac{1}{n_H}\sum_{i=1}^{n_H} log|S_\alpha(if_0)|$, where n_H is the number of harmonics that are supposed to lie within the analysis bandwidth. Some postprocessing steps are carried out in order to attenuate spurious peaks at multiples and submultiples of the true pitches, and to balance different fundamental frequency regions [1]. Finally, for each f0 in the grid the highest salience value is selected among the different available α values. In this way an F0gram is obtained, that shows the evolution of pitch for all the harmonic sounds in the signal. An example of an F0gram is depicted in Figure 1 for an audio fragment that is used throughout the paper.

3 SPECTRAL CLUSTERING

The goal of clustering can be stated as dividing data points into groups such that points in the same cluster are similar and points in different clusters are dissimilar. An useful way of representing the data is in the form of a similarity graph, each vertex corresponding to a data point. Two vertices of the graph are connected if their similarity is above certain threshold, and the edge between them is weighted by their similarity value. In

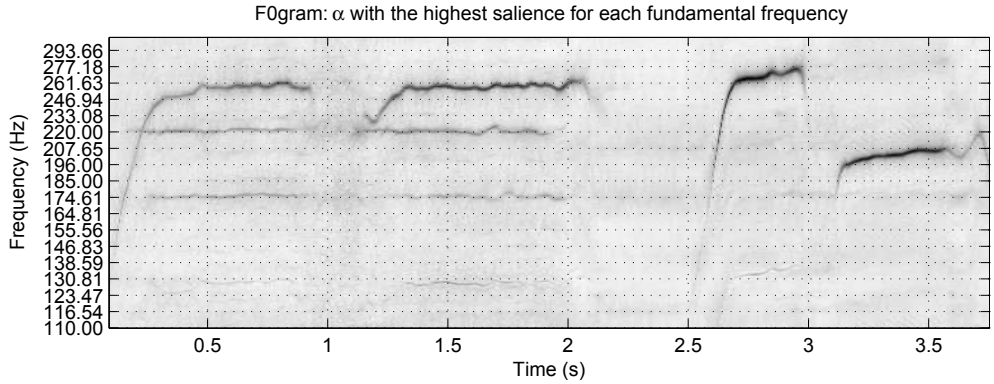


Figure 1: F0gram for a fragment of the audio file pop1 (used throughout the paper) from the MIREX melody test set. It consist of three simultaneous singing voices followed by a single voice, with a rather soft accompaniment.

terms of the graph representation, the aim of clustering is to find a partition of the graph such that different groups are connected by very low weights whereas edges within a group have high weights.

The simplest way to construct a partition is to solve the mincut problem. Given a number of clusters k, it consist in finding a partition $A_1, \ldots, A_k$ that minimizes

$$\text{cut}(A_1, \ldots, A_k) = \frac{1}{2} \sum_{i=1}^{k} W(A_i, \bar{A}_i), \qquad (2)$$

where $W(A, B) = \sum_{i \in A, j \in B} w_{ij}$ is the sum of weights of vertices connecting partitions A and B, and $\bar{A}$ stands for the complement of A. This corresponds to finding a partition such that points in different clusters are dissimilar to each other. The problem with this approach is that it often separates one individual vertex from the rest of the graph. An effective way of avoiding too small clusters is to minimize the Ncut function,

$$\text{Ncut}(A_1, \ldots, A_k) = \sum_{i=1}^{k} \frac{\text{cut}(A_i, \bar{A}_i)}{\text{vol}(A_i)}, \qquad (3)$$

where $\text{vol}(A) = \sum_{i \in A} d_i$ is the sum of the degree of vertices in A. The degree of a vertex is defined as $d_i = \sum_{j=1}^{n} w_{ij}$, so $\text{vol}(A)$ measures the size of A in terms of the sum of weights of those edges attached to their vertices. The Ncut criterion minimizes the between cluster similarity (in the same way as mincut), but it also implements a maximization of the within cluster similarities. Notice that the within cluster similarity can be expressed as, $W(A, A) = \text{vol}(A) - \text{cut}(A, \bar{A})$ [6]. In this way the Ncut criterion implements both objectives: to minimize the between cluster similarity, if $\text{cut}(A, \bar{A})$ is small, and to maximize the within cluster similarity, if $\text{vol}(A)$ is large and $\text{cut}(A, \bar{A})$ is small.

The mincut problem can be solved efficiently. However with the normalization term introduced by Ncut it becomes NP hard. Spectral clustering is a way to solve relaxed versions of this type of problems. Relaxing Ncut leads to the normalized spectral clustering algorithm. It can be shown [6] that finding a partition of a graph with n vertices into k clusters by minimizing Ncut, is equivalent to finding k indicator vectors $h_j = (h_{1j}, \ldots, h_{nj})'$ with $j = 1, \ldots, k$ of the form, $h_{ij} = 1/\text{vol}(A_j)$ if vertex $v_i \in A_j$ and zero otherwise. In this way, the elements of the indicator vectors point out to which cluster belongs each graph vertex. This problem is still NP hard, but can be relaxed by allowing the elements of the indicator vectors to take, instead of two discrete values, any arbitrary value in $\mathbb{R}$. The solution to this relaxed problem corresponds to the first k generalized eigenvectors of $(D - W)\, u = \lambda\, D\, u$, where D is an n by n diagonal matrix with the degrees of the graph vertices $d_1, \ldots, d_n$ on the diagonal, and $W = (w_{ij})_{i,j=1\ldots n}$ is the matrix of graph weights.

The vectors u of the solution are real-valued due to the relaxation and should be transformed to discrete indicator vectors to obtain a partition of the graph. To do this, each eigenvalue can be used in turn to bipartition the graph recursively by finding the splitting point such that Ncut is minimized [3]. However, this heuristic may be too simple in some cases and most spectral clustering algorithms consider the coordinates of the eigenvectors as points in $\mathbb{R}^k$ and cluster them using an algorithm such as k-means [6]. The change of representation from the original data points to the eigenvector coordinates enhances the cluster structure of the data, so this last clustering step should be very simple if the original data contains well defined clusters. In the ideal case of completely separated clusters the eigenvectors are piecewise constant so all the points belonging to the same cluster are mapped to exactly the same point.

Finally, the algorithm can be summarized as [6],

input: `similarity matrix` $S \in \mathbb{R}^{nxn}$, `number of clusters k`

steps:

1. `build a similarity graph using matrix` S
2. `compute the unnormalized Laplacian of the graph` $L = (D - W)$
3. `compute the first` k `generalized eigenvectors of` $(D - W)\, u = \lambda\, D\, u$

4. consider the eigenvectors $u_1, \ldots, u_k$ as columns of a matrix $U \in \mathbb{R}^{nxk}$
5. consider the vectors $y_i \in \mathbb{R}^k$ $i = 1, \ldots, n$ corresponding to the rows of U
6. cluster the points (y_i) in $\mathbb{R}^k$ with k-means into clusters $C_1, \ldots, C_k$

output: clusters $A_1, \ldots, A_k$ / $A_i = \{j \mid y_j \in C_i\}$.

4 PITCH CONTOURS FORMATION

In order to apply the spectral clustering algorithm to the formation of pitch contours several aspects must be defined. In particular, the construction of the graph involves deciding which vertices are connected. Then, a similarity function has to be designed such that it induces meaningful local neighbours. Besides, an effective strategy has to be adopted to estimate the number of clusters. In what follows, each of these issues are discussed and the proposed algorithm is described.

4.1 Graph construction

Constructing the similarity graph is not a trivial task and constitutes a key factor in spectral clustering performance. Different alternatives exist for the type of graph, such as k-nearest neighbor, ϵ-neighborhood or fully connected graphs, which behave rather differently. Unfortunately, barely any theoretical results are known to guide this choice and to select graph parameters [6]. A general criteria is that the resulting graph should be fully connected or at least should contain significantly fewer connected components than the clusters we want to detect. Otherwise, the algorithm will trivially return connected components as clusters.

To include information on temporal proximity a local fixed neighborhood is defined, such that f0 candidates at a certain time frame are connected only to candidates in their vicinity of a few frames (e.g. two neighbor frames on each side). In this way the graph is in principle fully connected, as can be seen in Figure 2, and resulting connected components are determined by similarity between vertices. Two candidates distant in time may nevertheless belong to the same cluster by their similarity to intermediate peaks. Note that in Figure 2 only one neighbor frame on each side is taken into account to link peaks. In this case, if a peak is missing the given contour may be disconnected. For this reason, a local neighbourhood of two or three frames on each side is preferred. Similarity of not connected components is set to zero, so a sparse similarity matrix is obtained.

In addition, a contour should not contain more than one f0 candidate per frame. To favour this, candidates in the same frame are not connected. Specifying cannot-link constrains of this type is a common approach for semi-supervised clustering [7]. However, this not strictly prohibits two simultaneous peaks to be grouped in the same cluster if their similarity to neighbor candidates is high. For this reason, clusters should be further processed to detect this situation and select the most appropriate candidate in case of collisions.

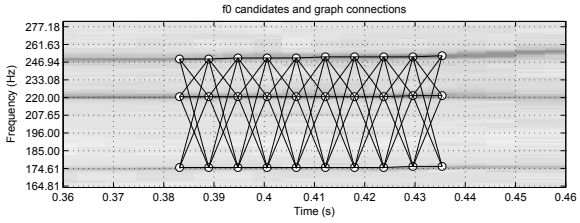


Figure 2: Graph connections considering only one neighbor frame on each side for an observation window of 10 frames. The resulting graph is fully connected.

4.2 Similarity measure

To define a similarity measure between F0gram peaks it is reasonable to base it on the assumption of slow variation of pitch contours in terms of fundamental frequency and salience (as defined in section 2).

Fundamental frequency distance between two graph vertices v_i and v_j may be better expressed in a logarithmic scale, that is as a fraction of semitones. To do this, pitch value of a vertex is expressed as the corresponding index in the logarithmically spaced grid used for pitch salience computation[1]. Then, this can be converted to a similarity value $s_{f0}(v_i, v_j) \in (0, 1]$ using a Gaussian radial basis function,

$$s_{f0}(v_i, v_j) = e^{-\frac{d^2_{f0}(v_i, v_j)}{\sigma^2_{f0}}} \tag{4}$$

where $d_{f0}(v_i, v_j) = |f0_i - f0_j|$ stands for pitch distance and σ_{f0} is a parameter that must be set which defines the width of local neighborhoods. In a similar way, a similarity function can be defined that accounts for salience proximity. To combine both similarity functions they can be multiplied, as in [5].

Although this approach was implemented and proved to work in several cases, the similarity measure has some shortcomings. Pitch based similarity is not able to discriminate contours that intersect. In this case, salience may be useful but it also has some drawbacks. For instance, points that are not so near in frequency and should be grouped apart, may be brought together by their salience similarity. This suggest the need for a more appropriate way of combining similarity values.

A significant performance improvement was obtained by combining the pitch value of the candidates and the chirp rates provided by the FChT. The chirp rate can be regarded as a local estimation of the pitch change rate. Thus, the pitch value of the next point in the contour can be predicted as $\overrightarrow{f0}^k_i = f0^k_i\,(1+\alpha^k_i\,\Delta t)$, where $f0^k_i$ and α^k_i are the pitch and chirp rate values, i and k are the candidate and frame indexes respectively, and Δt is the time interval between consecutive signal frames. Figure 3 depicts most prominent f0 candidates and their predictions for a short region of the example. Note that there are some spurious peaks in the vicin-

[1] In which a 16th semitone division is used (192 points per octave).

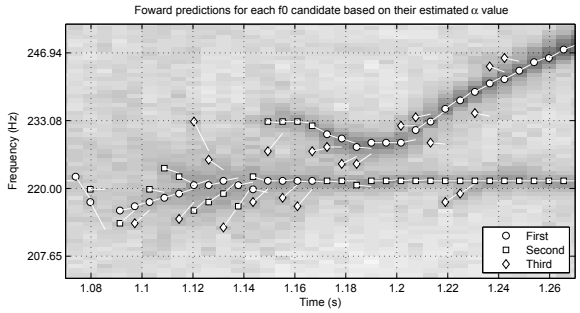


Figure 3: Forward predictions of the three most prominent f0 candidates for a short interval of the example. Although the clusters seem to emerge quite defined, spurious peaks may mislead the grouping. This can be improved if a backward prediction is also considered.

ity of a true pitch contour whose estimate lie close to a member of the contour and can lead to an incorrect grouping. A more robust similarity measure can be obtained by combining mutual predictions between pitch candidates. This is done by computing for each candidate also a backward prediction $\overleftarrow{f0}$ in the same way as before. Then, distance among two candidates v_i^k and v_j^{k+1} is obtained by averaging distances between their actual pitch values and their mutual predictions,

$$d_{f0}(v_i^k, v_j^{k+1}) = \frac{1}{2}\left[|f0_i^k - \overleftarrow{f0}_j^{k+1}| + |\overrightarrow{f0}_i^k - f0_j^{k+1}|\right] \quad (5)$$

Using this mutual distance measure the similarity function is defined as in Equation (4). Additionally, the same reasoning can be extended to compute forward and backward predictions for two or three consecutive frames. This similarity values are used as graph weights for candidates in their temporal proximity.

Still remains to set the value σ_{f0}, which plays the role of determining the actual value assigned to points in the vicinity and to outlying points. Self tunning sigma for each pair of data points was tested based on the distance to the k-th nearest neighbor of each point, as proposed in [8]. This approach can handle cluster with different scales, but applied to this particular problem it frequently grouped noisy peaks far apart from each other. It turned out that, given the filiform shape of clusters that are to be detected, a fixed value for σ_{f0} was more effective. Since pitch predictions become less reliable as the time interval grows, a more restrictive value for σ_{f0} is used for measuring similarity to points at the second and third consecutive frame (reported results correspond to $\sigma_{f0}^1 = 0.8$, $\sigma_{f0}^2 = 0.4$).

Figures 4 and 5 show two different examples of the local clustering, which correspond to three well-defined clusters and two clusters with spurious peaks. An observation window of 10 signal frames is used and the three most prominent F0gram peaks are considered. A neighborhood of two frames on each side is used. Similarity matrix is sorted according to the detected clusters, producing a sparse band diagonal matrix, where clusters can be visually identified as continuous bands.

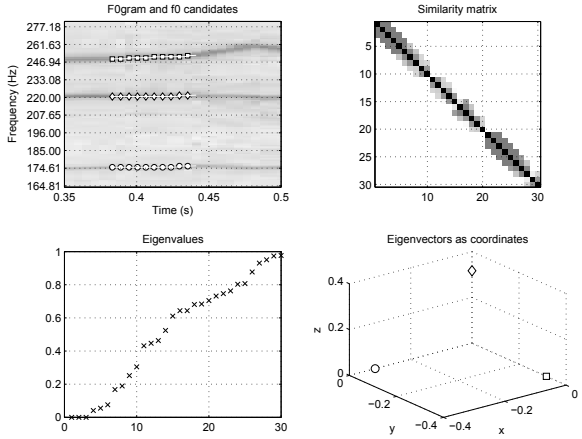


Figure 4: Local clustering for a short time interval of the audio example. Similarity matrix, eigenvalues and eigenvectors as coordinates are depicted. Three well-defined clusters can be identified in the data, as well as the corresponding bands in the similarity matrix. The multiplicity of eigenvalue zero coincides with the number of connected components. All members of a cluster are mapped to the same point in the transformed space.

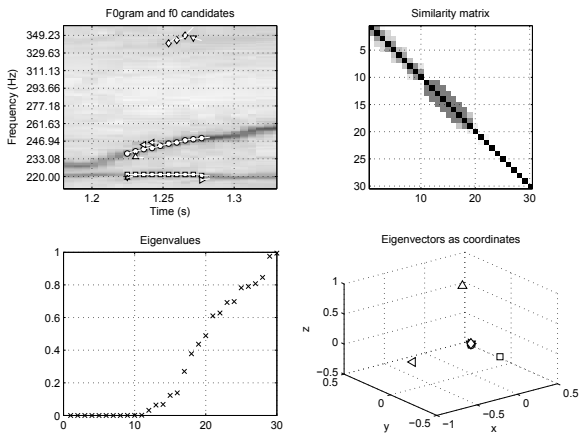


Figure 5: Local clustering example with two true pitch contours and several spurious peaks. The two corresponding bands in the similarity matrix can be appreciated. Multiplicity of eigenvalue zero not only indicates the relevant connected components but also isolated points. The true contours are correctly identified by the algorithm and spurious peaks tend to be isolated.

4.3 Number of clusters determination

Automatically determining the number of clusters is a difficult problem and several methods have been proposed for this task [7]. A method devised for spectral clustering is the eigengap heuristic [6]. The goal

is to choose the number k such that all eigenvalues $\lambda_1, \ldots, \lambda_k$ are very small, but λ_{k+1} is relatively large. Among the various justifications for this procedure, it can be noticed that in the ideal case of k completely disconnected components, the graph Laplacian has as many eigenvalues zero as there are connected components, and then there is a gap to the next eigenvalue.

This heuristic was implemented, but it sometimes failed to detect the correct number of cluster (e.g. when clusters are not so clear there is no well-defined gap). The following iterative strategy gave better results. It consist in firstly estimating the number of connected components using the multiplicity of eigenvalue zero by means of a restrictive threshold. Then, the compactness of the obtained clusters is evaluated. To do this, different measures were tested and a threshold on the sum of distances to the centroid in the transformed space was selected. As mentioned before, in case of completely separeted connected components all members of the same cluster are mapped to a single point in the transformed space. For this reason, the detection of poor quality clusters showed not to be too sensitive to the actual value used for thresholding. Each of the not compact clusters is further divided until all the obtained clusters conform to the threshold. This is done repeatedly by running k-means only to points in the cluster, starting with $k = 2$ for a bipartition and incrementing the number of desired clusters until the stop condition is met. This strategy tends to isolate each spurious peak as a single cluster (see Figure 5), what in turn favours to ignore them in the formation of pitch contours.

4.4 Filtering simultaneous members

Despite of the introduction of cannot-link constrains some clusters can occasionally contain more than one member at the same time instant. The best f0 candidate can be selected based on pitch distance to their neighbors. This approach was explored but difficulties were encountered for some particular cases. For instance, when a contour gradually vanishes F0gram peaks are less prominent, their pitch change rate estimate is less reliable and spurious peaks appear in the nearby region. Therefore, under the assumption of slow variation of contour parameters, salience similarity was introduced as another source of information. To do this, the most prominent peak of the cluster is identified and the cluster is traversed in time from this point in both directions, selecting those candidates whose salience is closest to the already validated neighbors.

4.5 Formation of pitch contours

The above described local clustering of f0 candidates has to be extended to form pitch contours. Increasing the length of the observation window showed not to be the most appropriate option. The complexity of the clustering is increased for longer windows, since a higher number of clusters inevitably arise mainly because of spurious peaks. Additionally, computational burden grows exponentially with the number of graph vertices. Thus, an observation window of 10 signal frames was used in the reported simulations ($\sim$ 60 ms).

Neighboring clusters in time can be identified based on the similarity among their members. A straightforward way to to this is by performing local clustering on overlapped observation windows and then grouping clusters that share elements. Figure 6 shows the clustering obtained using half overlapped observation windows for the two previously introduced examples.

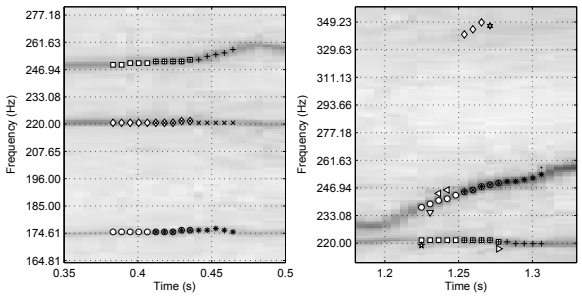


Figure 6: Examples of clustering using half overlapped observation windows. The pitch contours are correctly continued since several of their members are shared.

5 RESULTS AND DISCUSSION

The contours obtained by applying the proposed algorithm to the example audio excerpt are depicted in Figure 7. The three most prominent peaks of the F0gram are considered for pitch tracking. Several issues can be noted from these results. Firstly, the main contours present are correctly identified, without the appearance of spurious detections when no harmonic sound is present (e.g. around $t = 1.0\,s$). The example shows that many sound sources can be tracked simultaneously with this approach. No assumption is made on the number of simultaneous sources, which is only limited by the number of pitch candidates considered. The total number of contours and concurrent voices at each time interval is derived from the data.

It can also be seen that the third voice of the second note (approximately at $t = 1.0 - 2.0\,s$) is only partially identified by two discontinued portions. Because of the low prominence of this contour some of the pitch candidates appear as secondary peaks of the more prominent sources. This situation can be improved by increasing the number of prominent peaks considered.

Apart from that, there are the three short length contours detected at interval $t = 2.1 - 2.5\,s$ that seem to be spurious. However, when carefully inspecting the audio file it turned out that they correspond to harmonic sounds from the accompaniment. Although this contours have a very low salience they are validated because of their structure. It depends on the particular problem where this algorithm finds application if these contours may be better filtered out based on salience.

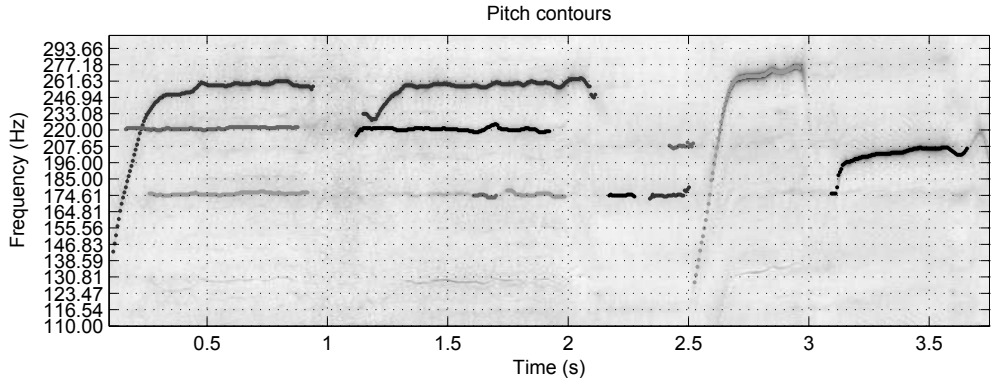


Figure 7: Pitch contours for the audio example obtained by considering the three most prominent F0gram peaks.

A melody detection evaluation was conducted following a procedure similar to the one applied in [1]. The vocal files of the 2004-2005 MIREX melody extraction test set were considered, which is a publicly labeled database available from `http://www.music-ir.org/mirex/`. It comprises 21 music excerpts for a total duration of 8 minutes.

The three most prominent F0gram peaks were selected as pitch candidates to form contours using the herein described algorithm. All the identified pitch contours were considered as main melody candidates and the ones that better match the labels were used to asses performance. Only those frames for which the melody was present according to the labels were taken into account to compute the evaluation measure according to,

$$\text{score}(f_0) = \min\{1, \max\{0, (\text{tol}_\text{M} - \Delta f_0)/(\text{tol}_\text{M} - \text{tol}_\text{m})\}\}$$

where $\Delta f_0 = 100|f_0 - f_0^{gt}|/f_0^{gt}$ is the relative error between the pitch contour value and the ground truth, and the tolerances tol_M and tol_m correspond to 3% and 1% respectively. This represents a strict soft thresholding.

The performance obtained in this way is compared to an equivalent evaluation that considers F0gram peaks as main melody estimates without performing any type of grouping into contours (as reported in [1]). Grouping the F0gram peaks into contours involves the determination of where does a contour starts and when does it ends, necessarily leaving some time intervals without melody estimation. This is avoided when isolated F0gram peaks are considered as main melody estimates, since for every melody labeled frame there is always a pitch estimation. Therefore, this performance measure can be considered as a best possible reference.

Results of the evaluation are presented in table 1. Two different values are reported for the pitch contours formation corresponding to a single run of the k-means algorithm and 10 repetitions. When the clusters in the transformed space are not well defined the k-means algorithm can get stuck in a local minima. This can be improved if several executions are performed but with different set of initial cluster centroid positions and the best performing solution is returned (i.e. lowest centroid distances). It can be noticed that the k-means repetition consistently gives a slight performance increase.

In addition, precision and recall values are reported. Precision is computed as the mean score value of the estimations within the 3% threshold. Remaining frames are considered not recalled items, as well as melody labeled frames for which there is no pitch contour.

When visually inspecting the results for individual files it turned out that most melody labeled regions for which there were no estimated contours correspond to low salience portions of the F0gram (for instance, when a note vanishes). It seems that labels are produced from monophonic files containing only the vocal melody and when mixed into a polyphonic track some regions are masked by the accompaniment. Figure 8 shows a detail of the current example where this situation can be appreciated. In order to take this into account the evaluation was repeated but ignoring low prominent melody frames. To do this a salience estimation was obtained for each labeled frame by interpolating the F0gram values. Then a global threshold was applied to discard those frames whose salience was below 30% of the F0gram maximum value (26% of the total frames).

Table 1: Results for the melody detection evaluation. The pitch contours are obtained from the three most prominent f0 candidates. An evaluation using F0gram peaks (1st to 3rd) without tracking is also reported.

F0gram	no salience threshold			30% salience threshold		
peaks	score	precision	recall	score	precision	recall
1	83.38	96.93	86.03	97.22	99.01	98.19
1-2	88.24	97.59	90.42	99.20	99.59	99.61
1-3	90.33	97.91	92.26	99.61	99.74	99.87
Pitch	no salience threshold			30% salience threshold		
contours	score	precision	recall	score	precision	recall
1 k-means	81.99	90.37	84.77	96.69	98.27	97.63
10 k-means	83.21	90.38	85.53	97.20	98.40	98.15
frames		100%			74%	

The performance of the pitch contours formation

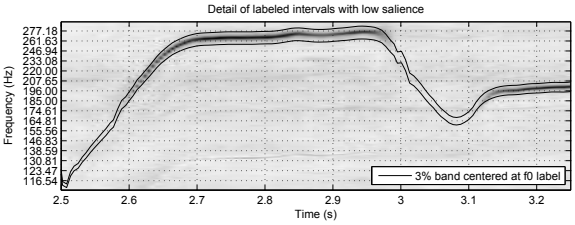


Figure 8: Some melody labeled regions of the example exhibit a very low salience (2.5-2.6 and 3.0-3.1 s).

by itself is quite encouraging. However, it decreases considerably compared to the values obtained before grouping F0gram peaks. The gap is reduced by restricting the evaluation to the most prominent peaks, which seems to confirm that low salience regions are troublesome for the algorithm. Visually inspecting the estimations for individual files gives the idea that most pitch contours are correctly identified. However, the evaluation results indicate the algorithm seems not to take full advantage of the information given by the F0gram peaks. Blindly relying on estimated α values no matter their corresponding salience is probably the most important shortcoming of the proposed algorithm.

6 CONCLUSIONS AND FUTURE WORK

In this work a novel way of performing pitch tracking by means of clustering local f0 candidates is described. The technique is based on an existing pitch salience representation called F0gram suited for polyphonic music [1]. This makes use of the Fan Chirp Transform which can produce precise representation of non stationary sound sources like singing voice.

The grouping is performed by applying a Spectral Clustering method since it can handle filiform shapes such as pitch contours. The similarity measure proposed takes advantage of the pitch change rate estimate provided by the FChT based F0gram. The determination of the number of clusters is tackled by an iterative approach, where the number of connected components is taken as an initial estimate and not compact enough clusters are further divided into an increasing number of groups. This strategy tends to isolate each spurious peak in a single cluster, what in turn favours to ignore them in the formation of pitch contours. Clustering is carried out for overlapped observation windows of a few hundred milliseconds and clusters from different time windows are linked if they share elements. In this way, groups that exhibit a coherent geometric structure emerge as pitch contours while the others are discarded.

The clustering approach to the tracking problem seems appealing because the solution involves the joint optimization of all the pitch contours present in a given time interval. Therefore, many sound sources can be tracked simultaneously and the number of contours and simultaneous sources can be automatically derived from the data. This differs from most classical multiple f0 tracking techniques in which each source is tracked in turn. In addition, the algorithm is unsupervised and relies on a few set of parameters. The influence of each parameter has not been fully assessed and the determination of optimal values will be tackled in future work. Preliminary results indicate that performance is not too sensitive to a particular configuration of some of them (e.g. number of candidates, k-means repetitions), but as it would be expected the values for σ_{f0} have to be set with more care. It is important to notice that the algorithm has low computational cost given that efficient algorithms exists for solving generalized eigenvector problems as well as for the k-means step.

Results of a melody detection evaluation indicate the introduced technique is promising for pitch tracking and can effectively distinguish most singing voice pitch contours. There is some room for improvement and the main shortcomings will be tackled in our future work. In particular, other sources of information should be included in the similarity measure in order to take full advantage of the local pitch candidates. The estimation of the pitch change rate is less reliable for low salience peaks. This could be taken into account when computing similarity, for example by adjusting the σ_{f0} value in accordance with the salience of the candidate.

REFERENCES

[1] P. Cancela, E. López, and M. Rocamora, "Fan chirp transform for music representation," in *13th Int. Conf. on Digital Audio Effects, Austria*, sep. 2010.

[2] A. de Cheveigné, "Multiple F0 estimation," in *Computational Auditory Scene Analysis: Principles, Algorithms and Applications*, D. Wang and G. Brown, Eds., pp. 45–79. IEEE / Wiley, 2006.

[3] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. on PAMI*, vol. 22, no. 8, pp. 888–905, aug. 2000.

[4] F. R. Bach and M. I. Jordan, "Learning spectral clustering, with application to speech separation," *JMLR*, vol. 7, pp. 1963–2001, 2006.

[5] M. Lagrange, L. G. Martins, J. Murdoch, and G. Tzanetakis, "Normalized cuts for predominant melodic source separation," *IEEE Trans. on ASLP*, vol. 16, no. 2, pp. 278 –290, feb. 2008.

[6] Ulrike von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, pp. 395–416, 2007.

[7] A. K. Jain, "Data Clustering: 50 Years Beyond K-Means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.

[8] L. Zelnik-Manor and P. Perona, "Self-tuning spectral clustering," in *Advances in NIPS*. 2004, pp. 1601–1608, MIT Press.

Sociedade de Engenharia de Áudio
Artigo de Congresso
Apresentado no 9º Congresso de Engenharia de Áudio
15ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2011, 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.

DUETO: Synthesis of Singing Duets in Real Time

Adriano Monteiro, Jônatas Manzolli
Interdisciplinary Nucleus for Sound Studies (NICS)
Music Department, Arts Institute
University of Campinas (UNICAMP)
Campinas, São Paulo, Brasil
monteiro.adc@gmail.com jotamanzo@hotmail.com,

DUETO comprises a singing voice synthesis module, a real time interactive gestural control and an evolutionary composition model emulating the *Tintinnabulation* technique created by the Estonian composer Arvo Pärt. The synthesis module is a variant of a linear source-filter model for singing voice. DUETO singing duets are generated in real time concatenating spectral shapes of vowels using an Interactive Genetic Algorithm (IGA) and the performer's gestural control via the Nintendo Wiimote Interface.

1. INTRODUCTION

DUETO development is in line with recent research on computer music, interactive performance and design of digital instruments for live performances [1][2][3]. It was implemented assembling an evolutionary algorithmic composition, a computer based synthesis and a gestural interface. The algorithmic composition is based on the *Tintinnabulation* technique of Arvo Pärt, described in [4][5]. The evolutionary model is an application of genetic algorithms [6][7][8]. There are studies on how evolutionary methods suited computer generated images [9][10] and on creative evolutionary systems [11]. There are also systems developed for music composition and improvisation [12][13].

In DUETO, the genetic algorithm interacts with the user. It is called in literature Interactive Genetic Algorithm (IGA) or Human-Based Genetic Algorithm (HBGA)[14]. It uses human arbitrariness to guide the computer generated synthetic evolution [15].

Nowadays, methods for singing voice are related to three approaches mainly: 1) *Physical Model*, based on modelling of voice production mechanisms; 2) *Spectral Model*, based on modelling the acoustic properties and hearing perception of voice utterance; 3) *Concatenative Synthesis*, in which the voice sound is learned directly from recorded samples and the samples are sequenced and/or modified to produce the synthesis [16][17][18].

In order to emulate the vocal tract, considering the resonance of vowels, we created a database of smoothed spectral shapes that were extracted from recordings of two female voices. DUETO is a software synthesizer similar to other spectral information methods such as: Linear Predictive Coding [19], FOF (Fomant d' Onde Formatique) [20], EpR Model (Excitation plus Resonance) [21].

2. COMPOSITION MODEL

Suiting our generative model we studied a vocal composition method called *Tintinnabulation* created by the Estonian composer Arvo Pärt. This technique was deeply influenced by medieval music and Gregorian Chant. Consequently, the melodic structure of Pärt's music is based on the use of Greek modes.

The Greek Modes can be described by pattern of musical intervals between adjacent notes. The pattern of the first Greek mode, called Ionian, is expressed by the set S={2,2,1,2,2,2,1}, in which the numerical values describe

the amount of halftones between two adjacent notes[1]. The other modes (*Dorian, Phrygian, Lydian, Mixolydian, Aeolian and Locrian*) are obtained by rotating the pattern in the set S. In the next sub-section, this rotation process is presented in the matrix $\mathbf{S_{7x7}}$

The Tintinnabulation term used by Arvo Pärt is a Latin expression for the sound produced by little bells. Pärt uses triad chord as sound of little bells. Thus he called tintinnabulation a technique for triad's voice-leading in vocal pieces [22].

Triads are chords composed of three notes. Given the jth note of a Greek mode, denoted by p_j, a triad is described here by the following set:

$$T = \{p_j, p_{j+2}, p_{j+4}\} \quad \textbf{(1)}$$

Two voices are generated with the *Tintinnabulation* technique: 1) one moving stepwise from and to a central note is called *M-voice*, 2) another constructed only with notes of a triad is called *T-voice.*

Frequently Pärt uses another voice called *Drone* that is a low pitch note sounding throughout most or all of a musical piece. A Drone produces long time resonances of the mode/scale fundamental note. In DUETO, we implemented the Drone as a synthetic male voice producing low frequency singing vowels.

The next sub-sections present a formal description of the algorithmic composition model related to Arvo Pärt's composition system.

2.1 Greek Modes Model

In order to establish a relationship between a given drone note and a Greek mode we defined a matrix $\mathbf{S_{7x7}}$. Each row of $\boldsymbol{S}$ corresponds to a rotation of the first row. The S_{mk} k=1,2…7, elements of mth row of matrix **S,** are the intervals of the mth Greek mode.

$$S = \begin{bmatrix} 2 & 2 & 1 & 2 & 2 & 2 & 1 \\ 2 & 1 & 2 & 2 & 2 & 1 & 2 \\ 1 & 2 & 2 & 2 & 1 & 2 & 2 \\ 2 & 2 & 2 & 1 & 2 & 2 & 1 \\ 2 & 2 & 1 & 2 & 2 & 1 & 2 \\ 2 & 1 & 2 & 2 & 1 & 2 & 2 \\ 1 & 2 & 2 & 1 & 2 & 2 & 2 \end{bmatrix}$$

Given a Drone note **N** represented by a MIDI Note Number, it is simple to construct a Greek mode associating elements of $\mathbf{S_{7x7}}$ to **N.** The notes of the mode are obtained by

$$p_{k+1}^m = p_k^m + S_{mk} \quad \textbf{(2)}$$

where $S_{mk} \in S_{7\times7}$ with $1 \le k \le 7$ and $1 \le m \le 7$. Parameter m is input by the user and $p_0^m = N$ is the Drone note in MIDI Note Number value.

2.2 Melodic Algorithm: M-voice

M-voice stepwise patterns are obtained by sequencing adjacent notes of a given mode. Pärt defined four possibilities [4][5]: *a) ascending pattern, from I to V note of the mode, b) descending, from I to IV note, c) descending, from V to I note and d) ascending, from IV to I note.* Our melodic algorithmic uses four index functions to generate the stepwise patterns used by Pärt.

In order to generate these four melodic patterns, we define four index functions $f_i(j): A \to P$ with A={0,1,2,3,4} and $P = \{p_0, p_1, ..., p_7\}$ constructed using **Eq. 2**, $1 \le i \le 4$ and $0 \le j \le 4$ as follows:

$$\begin{aligned} f_1(j) &= p_j \\ f_2(j) &= p_{(7-j)} \\ f_3(j) &= p_{(5-j)} \\ f_4(j) &= p_{(4+j)} \end{aligned} \quad \textbf{(3)}$$

In order to construct M-voice MIDI note patterns $M^a = \{m_1^a, m_2^a, ..., m_b^a\}$, we define a melodic chromosome or melodic control pair g=(a,b) with $1 \le a \le 4$ and $0 \le b \le 4$. The elements m_j^a are calculated as follows:

$$m_j^a = f_a(j) - 12((a-1)\bmod 2) \quad \textbf{(4)}$$

with $f_a(j)$ defined in **Eq. 3** and $0 \le j \le b$.

2.3 Harmonic Algorithm: T-voice

The T-voice is generated using a single triad and the M-voice pattern as note reference. The triad set pattern T defined in **Eq. 1** can be constructed using a fundamental note $p_0 \in P$ generated with **Eq.2**. In this way, a triad related to the fundamental note is given by

$$T = \{p_0, p_2, p_4\} \quad \textbf{(5)}$$

After defining the fundamental triad, the next control parameter of T-voice is the layer position **L**={1,2,3} where 1 means superior position, 2 inferior and 3 alternate.

Given the melodic control pair g(a,b) (see **Eq. 4**) the T-voice pattern denoted $T = \{t_1^a, t_2^a, ..., t_b^a\}$ is generated by:

$$t_j^a = h(j) - 12((a-1)\bmod 2) \quad \textbf{(6)}$$

where $0 \le j \le b$ and $h(j): A \to P$ is defined below.

when layer position **L**=1
If (j=0 or j=1) then ***h(j)***= p_2
If (j=2 or j=3) then ***h(j)***=p_4
If j=4 then ***h***(j)=p_0

when layer position **L**=2
If (j=0 or j=1) then ***h(j)***=p_4
If (j=2 or j=3) then ***h***(j)=p_0
If j=4 then ***h(j)***= p_2

when layer position **L**=3
If (j=0 or j=1) then ***h(j)***= $p_{(RND\ (2,4))}$
If (j=2 or j=3) then ***h(j)***= $p_{(RND(0,4))}$
If j=4 then ***h(j)***= $p_{(RND(0,4))}$

where RND(x, y) denotes a random choice between x or y.

[1] Here we use the Chromatic Scale with twelve notes in an octave and the minimum interval between two adjacent notes is a halftone = 1.

2.4 Rhythmic Algorithm

The rhythmic algorithm is interactive and it uses a pattern selected by the computer and a looping procedure. While the user controls changes on tempo, the computer, using genetic operations, selects a set of pre-defined pattern of numbers. This pattern is looped until another one is chosen by the computer.

Given a tempo ***t*** input by the user and a set $R = \{r_1, r_2, ...r_n\}$ selected by the computer with ***n*** elements, the sequence of durations looped by the algorithm is:

$$S = \{t.r_1, t.r_2 ....t.r_n\} \quad \textbf{(7)}$$

2.5 Duet Note Event

Finally, we call Duet Note Event (DNE), denoted by **E**(j,a), a quadruple of two notes and two durations automatically generated by the computer. Given the melodic chromosome g(a,b) the DNE is expressed by

$$E(j,a) = (m_j^a, t.r_j, t_j^a, t.r_j) \text{ with } 0 \le j \le b \quad \textbf{(8)}$$

where m_j^a is defined in **Eq. 4**, t_j^a in **Eq. 6** and $t.r_j$ in **Eq. 7** where t and m are tempo input by the user.

3. GENETIC ALGORITHM MODEL

DUETO is an interactive system that generates patterns according to genetic operations and user's input. It is a semi-automatic decision system in which man and machine collaborate to each other. It generates note patterns based on an algorithmic and evolutionary model that is not completely strict to pre-defined rules.

3.1 Population of Control Parameters

We define a population as a matrix $D_{n \times m}$ where **n,** that is the number of rows, is the number of individuals in the population. Each row of D is a list of **m** control parameters. Particularly, in the implementation presented here, m=8.

We start with an initial population $\mathbf{D^{(0)}}$ and then we iteratively construct a sequence of j generations of that initial population. It is given by $\mathbf{D^{(1)}}$, $\mathbf{D^{(2)}}$,…, $\mathbf{D^{(j)}}$ where the jth generation is a matrix defined as

$$D^j = \begin{bmatrix} d_1 \\ d_2 \\ ..... \\ ..... \\ d_n \end{bmatrix}$$

where each row is a chromosome with eight parameters:

$$d_k = (g_k, a_k, b_k, mr_k, tr_k, mv_k, tv_k, f_k) \quad \textbf{(9)}$$

Where g_k is the Greek mode, a_k and b_k is the melodic control pair, $0 \le mr_k, tr_k \le 7$ are the durations for the M-voice and T-voice respectively, $0 \le mv_k, tv_k \le 7$ are the vowels indexes for M-voice and T-voice respectively and $0 \le f_k \le 1$ is the fitness evaluation the individual d_k.

It is possible also to define de best individual of $D^{(j)}$. It is denoted by:

$$d_j^* = \max(f_k) \text{ for } 1 \le k \le m \quad \textbf{(10)}$$

3.2 Reproduction Process

Reproduction process consists of renewing completely the population applying crossover and mutation over all individuals.

3.2.1 Crossover

Given two individuals d_p and d_q that are randomly chosen in the population and a crossover rate α, with 0 ≤ α ≤ 1, all individuals are renewed in pairs as follows:

$$d_p = \alpha.d_q + (1-\alpha).d_p \quad \textbf{(11)}$$

with $1 \le p,q \le n$ and the crossover is applied to eight parameters of each individual and this is a convex combination of d_p and d_q.

3.2.2 Mutation

Given the j-th matrix $\mathbf{D^{(j)}}$ with **n** rows or individuals and a mutation rate 0 ≤ ß ≤ 1 , we choose by random (β*n) rows in the matrix $\mathbf{D^{(j)}}$, mutation is assigning a random number for each one of the 8 parameters of an individual. Only f_k=0.0 is assigned by definition.

3.3 Selection, Reproduction and Feedback

DUETO works in an IGA manner and therefore the user commands selection of a new individual and the reproduction process. Two buttons from the Wiimote interface are used to interfere in the genetic cycle: buttonA makes the computer selects the best individual and operate reproduction (first mutation and after crossover are applied over the population); buttonB feedbacks the last selected individual in the population and after that only crossover is applied over the population. The individual d_k feedback by the user receives fitness score f_k=1.0 and mutated individuals receive fitness score f_k=0.0 (zero).

Fitness is affect by the performer's feedback that reinforces a direction in the search space. In oposition, mutation assigns random values to all individual cromossoms and it induces new directions in the search space (see figure 1).

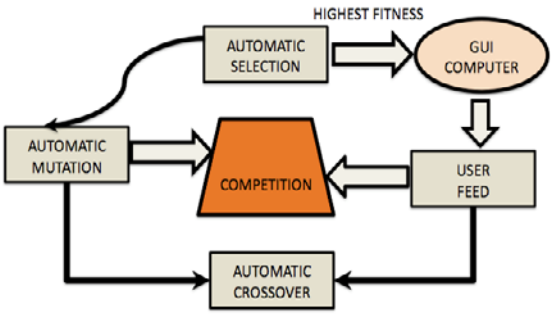


FIGURE 1– DIAGRAM OF THE INTERACTIVE GENETIC ALGORITHM (IGA) IMPLEMENTED IN DUETO.

3.4 Genetic Parametric Assignment

The rhythmic duration pattern, synthesis parameters, frequency of the excitation signal and the vowel spectral formant, are controlled automatically by the IGA in real

time. The rhythmic pattern pair (mr_k,tr_k), melodic control pair (a_k,b_k) and Greek mode (g_k) are determined by the chromosome of the best individual (see 3.1) that is selected according to **Eq. 10.** Using the best individual parameters, new duration values are selected, see 2.4, and a new pitch is generated, see 2.2 and 2.3. After these, new values are sent to the synthesis module to control the frequency of the excitation signal of the two synthetic voices: M-voice and T-voice. Finally, the pair (mv_k, tv_k) determines the vowel spectral formants of them.

4. SYNTHESIS METHOD

We adopted a formant synthesis method based on source-filter decomposition. First, we extract formant spectral envelopes from recorded voice samples by a decovolution process, applying a cepstral filtering (see an introduction in [19]). Then, we use the spectral data to re-synthesis by convolution in the frequency domain with an excitation signal.

Assuming a model for representing the vocal sound production as a linear source-filter model, it is possible to separate the spectral components related to the excitation signal produced by the vocal chords from the spectral components related to the resonance of the vocal tract (formant characteristics) by a filtering process in the cepstral domain.

In the time domain, the excitation signal $x(t)$ is convoluted with the vocal tract impulse response system $h(t)$.

$$y(t) = x(t) * h(t) \qquad (12)$$

Analysing $y(t)$ in the frequency domain, operating a Fourier Transform, the spectrum of the excitation signal $X(f)$ is multiplied by the frequency response of the vocal tract $H(f)$.

$$Y(f) = X(f) \times H(f) \qquad \textbf{(13)}$$

$X(f)$ and $H(f)$ can be separated by a simple subtractive operation if switched from $Y(f)$ to $\log(Y(f))$:

$$\log(Y(f)) = \log(X(f)) + \log(H(f)) \qquad \textbf{(14)}$$

Knowing that $\log(Y(f))$ is the sum of fast periodic functions caused by the excitation signal and slow periodic functions caused by the resonance of the vocal tract, we treat the spectral shape's periodicity as it was a function of time and apply the Fourier Transform to separate its rapid components (the higher cepstral bins) from the slow components (the lower cepstral bins) in the cepstral domain.

Finally, the high cepstral bins are filtered out and the signal is converted again to the spectral domain and operated an exponentiation to rescaling the spectrum. The results, in a smoothed spectral shape, are related to the frequency response of the vocal tract filter.

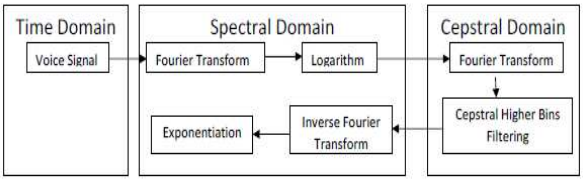


FIGURE 2 - SCHEMA OF THE SPECTRAL DECONVOLUTION PROCESS BY FILTERING IN THE CEPSTRAL DOMAIN.

In order to find a model for the vocal tract filters we created a database of spectral shapes extracted from recordings of two females voices singing eight vowels, see Table 4.

We recorded both voices singing each vowel in the same pitch and after this process was repeated for other pitches. Expressed in MIDI Note Numbers the pitches were 53, 57, 60, 65, 69 and 72. Finally, based on the idea that "the vocal cord or glottis signal is rather constant in shape and that this shape is roughly triangular" [19] we window our excitation with a bell-shape like signal and convolve all spectral shape in the database with that bell-shape excitation to obtain a synthesised vocal sound.

5. Performance Controls

A graphic user interface (GUI) (see Figure 3), a Wii Nintendo remote-control and the performer's voice are the interactive real time control of our system. DUETO was developed in Pure Data and the Bluetooth port was used to establish communication with the wiimote interface through the OSC protocol. The wiimote and its accessory, the nunchuck, are used in DUETO as gestural control. They embed a three-dimensional accelerometer that retrieves acceleration parameters for the three coordinates of space, in real-time and transmits these data wirelessly – via Bluetooth.

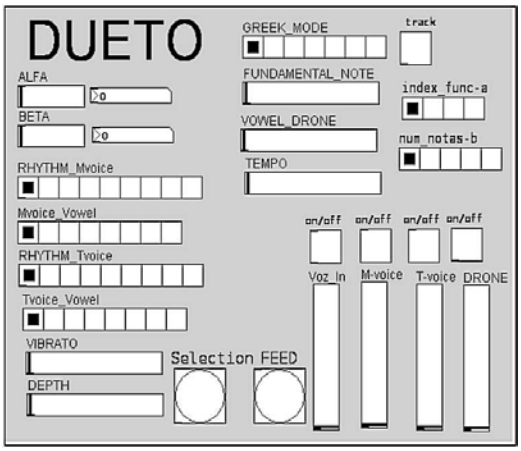


FIGURE 3 – GRAPHIC USER INTERFACE OF DUETO

The gestural control of DUETO consists of seven control buttons of the Wiimote [A, B, minus, home, plus, 1, 2] and buttons [Z, C] of the Nunchuck. Continuous data produced by the [Pitch, Roll] parameters and the joystick of the Nunchuk is used in combination with the buttons (see Table 1 and Table 2).

A	B	minus, home, plus	1	2
genetic selection	genetic feedback	select volume	on/off tracking	on/off drone
Pitch		**Roll**		
fund. frequency		volume		

TABLE 1. CONTROLLERS OF THE WIIMOTE INTERFACE

The gesture control of DUETO works in pairs: a button selects a parameter and the *Pitch* or *Roll* changes its value. In Table 2, *minus, home, plus* are selection buttons. They control volume of the performer's voice, the synthetic voices and the drone respectively. While one of these three buttons is hold, the value from the *Roll* is used to change the volume of the selected signal. The Z button and the nunchuk *Pitch* parameter are used to change tempo and the C button and the nunchuk *Roll* parameter are used to change the vowels formants of the drone voice. The vibrato frequency and its depth are controlled by the joystick x and y respectively.

Z	C		
Select tempo control	Select Drone vowels control		
Pitch	**Roll**	**Joy-x**	**Joy-y**
tempo	Drone vowels	Freq. Vibrato	Depth Vibrato

TABLE 2. CONTROLLERS OF THE NUNCHUK INTERFACE

6. RESULTS AND EVALUATION

Our objective here is to simulate the real time behaviour of DUETO without direct human interaction, evaluate how the genetic algorithm controls the synthesis machine and the effect of this mechanism on the synthetic voice M-voice and T-voice.

The used presets are: a) fundamental pitch: 60 MIDI Note Number, b) vibrato frequency: 6 Hz, c) depth: 1/8 of tone, d) tempo: 500 ms, e) duration of each simulation: 60 seconds, f) population: 20 individuals and the sound output of the M-voice and T-voice were recorded separately. Every 5 seconds *selection* was automatic applied over the population and *feed* was programmed to operate every 7 seconds. We run two simulations as presented in Table 3.

#	Alfa	Beta	Feed
1	0.5	0.1	Off
2	0.5	0.1	On

Table 3. Parameters of two simulations

6.1 Evaluation of the Genetic Algorithm

In simulation #1 (solid line plot in Figure 4) only two individuals (10% of the population) were mutated and the crossover rate was 50%. In simulation #2 (dash line plot in Figure 4) reproduction and mutation rates were maintained and every 7 seconds feedback was applied in order to simulate reinforcement.

Two plots in Figure 4 describe the evolution of the best individual d^*_k. It is possible observed that the number of different best individuals changes from 7 in simulation #1 to 10 in simulation #2 and the distribution of individuals is more homogenous in simulation #2.

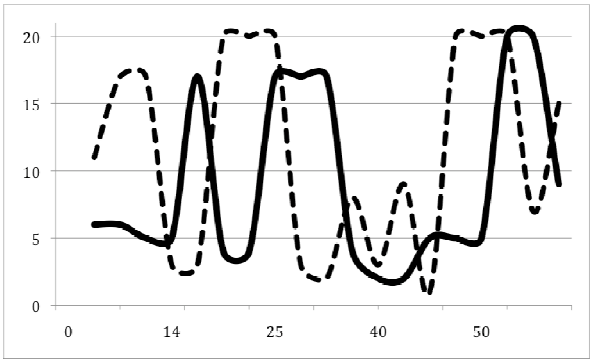


FIGURE 4 – GRAPHIC OF THE 2 SIMULATIONS TOGETHER. SOLID LINE IS SIMULATION #1 AND DASH LINE IS SIMULATION #2. THE Y-AXIS IS THE BEST INDIVIDUAL AND X-AXIS IS TIME FROM 0 TO 60 SECONDS.

The distribution of individuals in simulation #1 is presented in Figure 5: [(2,2),(4,3),(5,5),(6,2),(9,1),(17,4),(20,2)].

In the second simulation the computer chose 10 different individuals and the distribution is [(1,1),(2,1),(3,4),(7,1),(8,1),(9,1),(11,1),(15,1),(17,2),(20,6)] as presented in Figure 6.

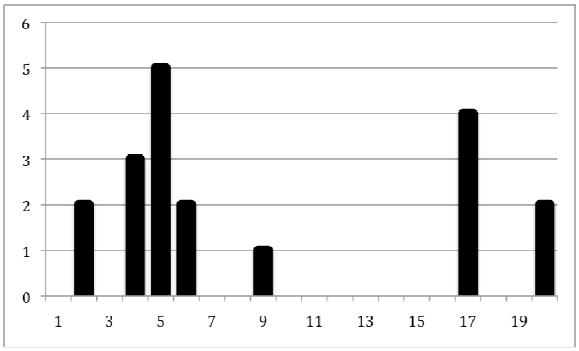


FIGURE 5 – SIMULATION #1, DISTRIBUTION OF INDIVIDUALS.

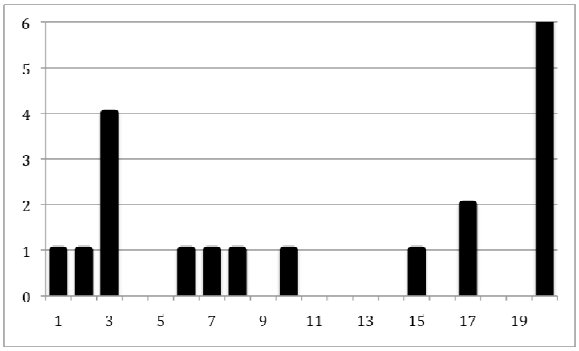


FIGURE 6 – SIMULATION #2, DISTRIBUTION OF INDIVIDUALS.

6.2 Synthesis Evaluation

It is possible now to compare the evolution of the genetic algorithm with the synthetic voices. As defined in 3.1, the population matrix is

$$D^j = \begin{bmatrix} d_1 \\ d_2 \\ \ldots\ldots \\ \ldots\ldots \\ d_{20} \end{bmatrix}$$

where each row contains the following array:

$$d_k = (g_k, a_k, b_k, mr_k, tr_k, mv_k, tv_k, f_k).$$

As already said, we generated synthetic voices using a database of two female voices recordings. Table 4 presents the vowels associated to eight samples. The first set of spectral formants, used to synthesize the M-voice, was produced on pitch = 60. The second set, used in the T-voice, was produced with another female voice for pitch = 65.

1	2	3	4	5	6	7	8
[a]	**[æ]**	**[e]**	**[ε]**	**[ə]**	**[o]**	**[i]**	**[u]**

TABLE 4. ARRAY OF SPECTRAL FORMANTS IN THE VOWEL DATABASE USED IN DUETO.

In the next 4 graphics we describe how the genetic algorithm controlled evolution of the pair (mv_k, tv_k) that concatenates vowel spectral shapes in real time. In simulation #1, presented in Figures 7 and 8, we observed that the computer produced a variation around vowels [3,4] respectively **[e]** and **[ε].**

In simulation #2, reproduction rates were maintained and feedback was performed every 7 seconds. Therefore crossover was applied over the population more frequently. The observed result showed that the vowels were confined to a region between [3,4,5] respectively **[e], [ε]** and **[ə]** as showed in Figures 9 and 10.

With these results it was possible to observe: the genetic algorithm influences voice synthesis according to the evolution of the individuals in the population, and the emulation of performer's feedback drives the synthesis producing a homogenous spread over the population, as observed in simulation #2.

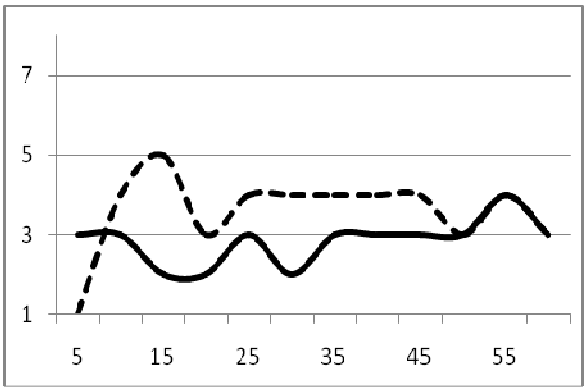


FIGURE 7– EVOLUTION OF VOWELS IN SIMULATION #1.

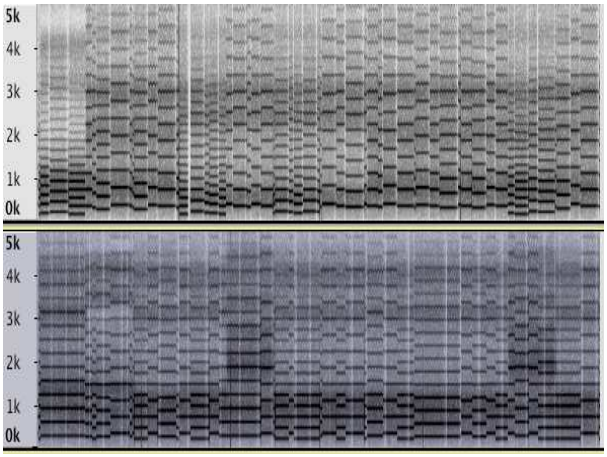


FIGURE 8 – SPECTROGRAM OF SIMULATION #1.

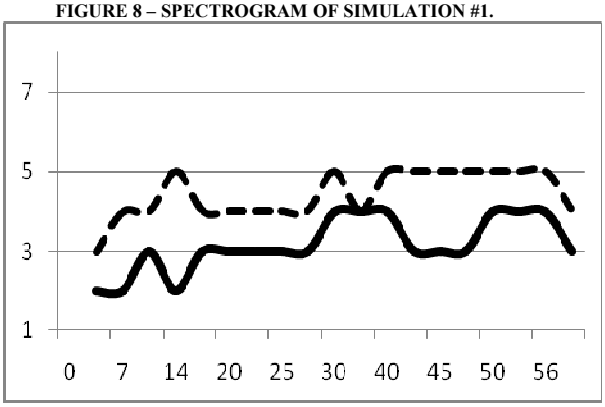


FIGURE 9– EVOLUTION OF VOWELS IN SIMULATION #2.

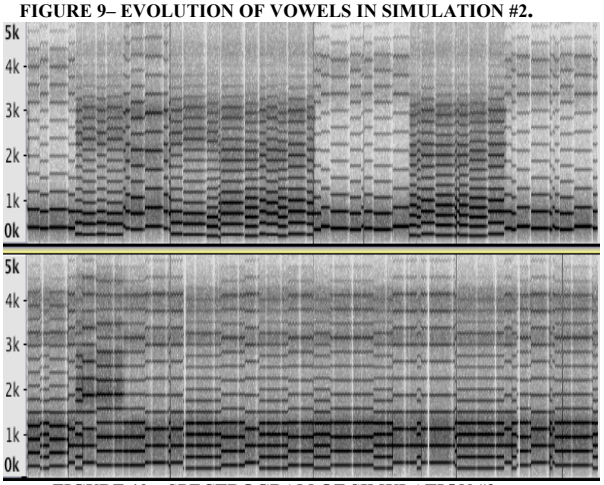


FIGURE 10 – SPECTROGRAM OF SIMULATION #2.

7. CONCLUSION

This article presents an interactive computer system to generate synthetic duets. It is a semi-automatic decision system in which an interactive genetic algorithm (IGA) controls two synthetic voices in real time. DUETO is a hybrid system combining human selection to computer automatic evolution.

On one hand the development of singing voice synthesis has improved in the last years and nowadays it is possible to encapsulate synthesis processes in real time programming language such as MAX/MSP or Pure Data. On the other hand, it is still a need for expressiveness in voice synthesis, and also there still is a search for composition algorithms to create synthetic singing music.

Finally, our contribution on this paper combines three aspects: a) a simple gesture interface to control in real time

synthesis of singing duets, b) an algorithmic model for generating vocal music with computers and c) an evolutionary model to evolve the synthesis according to performer's interaction.

ACKNOWLEDGMENTS

We thank the NICS staff for supporting this research and Prof. Adolfo Maia Jr. for reviewing the mathematical formulation. We also thank Aline da Silva Alves and Nathália Yuri Kato da Silva for recording the voice samples. The Brazilian agency CNPq and FAPESP support Monteiro and Manzolli.

REFERENCES

[1] Bown, O., Eldridge, A., MacCormack, J. (2009) "Understanding Interaction in Contemporary Digital Music: from instruments to behavioural objects" Organised Sound, 14(2), pp. 188–196, Cambridge University Press.

[2] Drummond, J. "Understanding Interactive Systems". Organised Sound, 14(2), pp. 124–133, Cambridge University Press, 2009.

[3] Miranda, E.R. & Wanderley, M. (2006). New Digital Musical Instruments: Control And Interaction Beyond the Keyboard. Publisher: A-R Editions, Inc. ISBN 0-89579-585-X.

[4] Hillier, Paul. Arvo Pärt. New York: Oxford University Press, 2002.

[5] Votta Jr., A. (2009) A Música Tintinabular de Arvo Pärt. Master Dissertation, Department of Music, University of Campinas (Unicamp).

[6] Holland, J. H. Adaptation in Natural and Artificial Systems. University of Michigan Press, 1975.

[7] Holland, J. H. "Genetic Algorithms", Scientific American, July: 1992.

[8] Sims, K. (1993) Interactive Evolution of Equations for Procedural Models. The Visual Computer 9, 466—476.

[9] Todd, S., Latham, W. (1992) Evolutionary Art and Computers. Academic Press.

[10] Baluja, S., Pomerleau, D. and Jochem, T. (1994) "Towards automated artificial evolution for computer generated images". Connection Science, 6(2–3), 325–354.

[11] Bentley, P., Corne, D. (2002) Creative Evolutionary Systems. Morgan Kaufmann, San Francisco.

[12] Biles, J. (1994). Genjam: A genetic algorithm for generating jazz solos. In Proceedings of the International Computer Music Conference.

[13] Todd, P. M., Werner, G. M.: Frankensteinian Methods for Evolutionary Music Composition. In: Griffith, N., Todd, P. M. (eds) Musical Networks: Parallel Distributed Perception and Performance, pp. 313—339. The MIT Press, Cambridge (1999).

[14] Kosorukoff, Alex (2001). Human-based genetic algorithm. In IEEE Transactions on Systems, Man, and Cybernetics, SMC-2001, 3464-3469.

[15] Moroni, A.; Von Zuben, F.; Manzolli, J. "ArTbitration: Human-Machine Interaction in Artistic Domain". Leonardo Music Journal, Cambridg Mass, MIT Press, v. 35, n. 2, p. 185-188, 2002.

[16] Cook, P. (1996). Singing voice synthesis: History, current work, and future directions. Computer Music Journal, 20(3).

[17] Cook, P. (1993). SPASM, a real-time vocal tract physical model controller; and singer, the companion software synthesis system. Computer Music Journal, 17(1):pp. 30–44.

[18] Childers, D. G. 1994. "Measuring and Modeling Vocal Source-Tract Interaction," *IEEE Transactions on Biomedical Engineering 1994.*

[19] Tempelaars, S. Signal processing, speech, and music. Lisse: Swets & Zeitlinger. 1996. pp. 267.

[20] Rodet, X. (1984). Time-domain formant-wave-function synthesis. Computer Music Journal, 8(3):pp. 9–14.

[21] Bonada, J., Celma, Ó., Loscos, À., Ortolà, J., Serra, X., Yoshioka, Y., Kayama, H., Hisaminato, Y., and Kenmochi, H. (2002). Singing voice synthesis combining excitation plus resonance and sinusoidal plus residual models. In Proc. Int. Computer Music Conf, pages 139–146, Havanna, Cuba.

[22] Langager, Graeme. (1997) The Tintinnabuli compositional style of Arvo Pärt. Master Dissertation. California State University. Long Beach.

Sociedade de Engenharia de Áudio
Artigo de Congresso

Apresentado no 9º Congresso de Engenharia de Áudio
15ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2011, 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.

Redução de Ruído em Sinais de Voz Usando Função de Limiar *SoftSoft* para Minimizar a Distorção Log-Espectral

Irineu Antunes Júnior[1]

[1] Universidade Federal do ABC, CECS, Engenharia de Informação
Santo André, SP, CEP 09.210-170, Brasil

irineu.antunes@ufabc.edu.br

RESUMO

A redução de ruído em sinais com *AWGN* pode ser feita pela aproximação por uma versão suavizada. Num domínio transformado, pode-se obter tal aproximação pela modificação de coeficientes, a exemplo da função *Soft* empregada para sinais suaves por trechos, que não é o caso da voz. Este trabalho inova ao estudar como obter funções especializadas na modificação de coeficientes de voz. Por meio de uma abordagem variacional, são propostas funções (*LogSoft* e *SoftSoft*) que obedecem a um compromisso entre medidas relacionadas com a percepção auditiva. Simulações constatam que a *SoftSoft* resulta em menor distorção log-espectral do que a *Soft* convencional.

0 INTRODUÇÃO

Em muitos sistemas de comunicação e aplicações multimídia, o sinal de voz apresenta contaminação por ruído que pode ser modelado como *AWGN* (*additive white gaussian noise*). Pode-se aprimorar o desempenho destes sistemas por meio de técnicas de redução de ruído e de melhoria de sinais de voz. Como o *AWGN* não é correlacionado com o processo estocástico que representa o sinal de voz e possui densidade espectral de potência plana, uma das primeiras técnicas usadas foi a subtração espectral [1, p.333]. Nesta técnica, uma estimativa do espectro do ruído é subtraída do espectro do sinal. Posteriormente, foram consideradas outras formas de modificação dos coeficientes transformados do sinal ruidoso com a finalidade de obter um espectro com menos componentes de ruído. Dada a relativa insensibilidade de sinais de voz face a modificações de fase [1, p.336], pode-se considerar apenas a modificação da magnitude dos coeficientes. Neste trabalho, empregam-se os coeficientes reais providos pela *DCT* (*discrete cosine transform*), uma vez que o uso da *FFT* ou de outra transformada complexa não traria benefício significativo. Deve-se comentar que o emprego transformadas mais sofisti-

cadas, como o uso de *lapped orthogonal transforms* ou de *wavelets* foi averiguado, tendo-se constatado que podem promover alguns melhoramentos adicionais como a eliminação de pequenos estalos (*clicks*) devido ao processamento em blocos, além de uma representação com menor número de coeficientes.

Este trabalho tem como finalidade investigar o uso de funções de limiar (*thresholding functions*) para redução de ruído em sinais de voz. As funções de limiar mais conhecidas são as funções *Soft* e *Hard* [2], inicialmente, concebidas para selecionar coeficientes de uma transformada *wavelet*, mas que também podem ser aplicadas aos coeficientes de outras transformadas. Normalmente, o sinal reconstruído usando apenas coeficientes selecionados por essas funções fornece uma aproximação (ou "outra versão do sinal") que pode ter interesse prático. Por exemplo, no caso de aplicações de compressão, há interesse numa aproximação que tenha poucos componentes, ou seja, poucos coeficientes transformados. Já no caso de redução de ruído, o número de coeficientes não é importante, sendo esperado um sinal aproximado que contenha menos ruído, podendo o resultado ser adotado como uma estimativa do sinal original.

Neste trabalho, a redução de ruído num sinal de voz corrompido por *AWGN* é obtida aplicando-se funções de limiar aos coeficientes *DCT*. Inicialmente, na Seção 1, é detalhado o processamento em blocos aqui empregado e, ainda, são apresentadas algumas funções de limiar existentes, concebidas para diversas aplicações, mas que podem eventualmente ser usadas para redução de ruído em sinais de voz. Na Seção 2, apresenta-se uma abordagem variacional para se obter uma dada função de limiar, que pode ser vista como o funcional que minimiza a relação de compromisso entre a fidelidade (preservação de coeficientes) e a suavidade (redução de coeficientes) do resultado. A classe de aproximações provida por uma função de limiar depende de como esse compromisso é expresso matematicamente. Neste ponto, inova-se ao propor funções de limiar especializadas na redução de ruído e melhoria de sinais de voz. Por fim, na Seção 3, constata-se por meio de simulações computacionais que o emprego da função de limiar especializada *SoftSoft* fornece um resultado com menor distorção do que a função convencional *Soft*.

1 Descrição do Processamento Empregado

O sinal de voz pode ser considerado um processo estocástico não estacionário. No entanto, para permitir o uso de processamento no domínio transformado, pode-se presumir que o mesmo é estacionário para intervalos da ordem de 30 ms de duração ([1, p.13]) e procede-se ao processamento em blocos de curta duração. Mais especificamente, dado um trecho de voz ruidoso $\mathbf{y} = \{y(n)\}_{n=0}^{N-1}$, com N amostras, este é dividido em M blocos, sendo cada bloco ($\mathbf{y}_m$) de comprimento $K = N/M$.

Em cada bloco, o sinal ruidoso é dado por $\mathbf{y}_m = \mathbf{s}_m + \mathbf{w}_m$, a soma do sinal de voz ($\mathbf{s}$) e do *AWGN* ($\mathbf{w}$) de média nula. No domínio transformado $\mathrm{DCT}(\mathbf{y}_m) = \mathrm{DCT}(\mathbf{s}_m + \mathbf{w}_m)$, ou seja, $\mathbf{Y}_m = \mathbf{S}_m + \mathbf{W}_m$, na qual a densidade espectral de potência do ruído é constante e igual à sua variância σ_w^2.

Para obter uma aproximação do trecho de voz sem ruído ($\mathbf{s}_m$), neste trabalho, consideram-se apenas funções que modificam individualmente cada um dos coeficientes, de maneira que uma estimativa do coeficiente livre de ruído é calculada por

$$\hat{S}_m[k] = g_t(Y_m[k]), \tag{1}$$

onde $Y_m[k]$ é o coeficiente ruidoso de índice k do m-ésimo bloco.

Após a aplicação da função $g_t(.)$ a cada um dos coeficientes ruidosos, calcula-se a transformada inversa de cada bloco e recompõe-se o sinal com menor nível de ruído ($\hat{\mathbf{s}}$). A eficácia do método pode ser verificada por audição informal e, de maneira mais objetiva, por medidas espectrais de distorção que refletem a qualidade psicoacústica da voz, como por exemplo a distorção log-espectral (*LSD*, [3, p.158]), considerada mais adiante nos resultados experimentais.[1]

1.1 Funções de Limiar para Modificação de Coeficientes

Uma das primeiras funções de limiar usadas para redução de ruído em sinais de voz foi a função *Soft* que possui um parâmetro ajustável t para estabelecer o limiar de seleção dos coeficientes segundo a expressão

$$g_t^{(S)}(X) = \begin{cases} \mathrm{sgn}(X)(|X| - t), & |X| > t \\ 0, & |X| \leq t \end{cases} \tag{2}$$

na qual empregam-se X para representar de maneira genérica um dado coeficiente a ser modificado e $\mathrm{sgn}(.)$ é a função *signum*. Nesta expressão, pode-se observar que os coeficientes com magnitude menor do que o limiar (t) são anulados; já aqueles acima do limiar, são mantidos, sofrendo apenas um pequeno encolhimento. Tal encolhimento, não introduz muita distorção no resultado pois, normalmente, grande parte da energia do sinal de voz é representada por um pequeno número de coeficientes de grande magnitude e, além do mais, $t \ll X_{\max}$ (o maior coeficiente em magnitude). Eventualmente, para compensar esse encolhimento, pode-se introduzir um ganho no sinal resultante para que este tenha a mesma intensidade sonora que o sinal original, permitindo a comparação subjetiva por ouvintes.

[1] Escolheu-se uma medida de distorção espectral porque os métodos considerados atuam no domínio transformado. Tal escolha possibilita determinar qual método proporciona melhor estimativa de espectro. O uso de medidas genuínas de qualidade psicoacústicas pode ser considerado numa avaliação mais geral do método quando comparado com outras técnicas de redução de ruído.

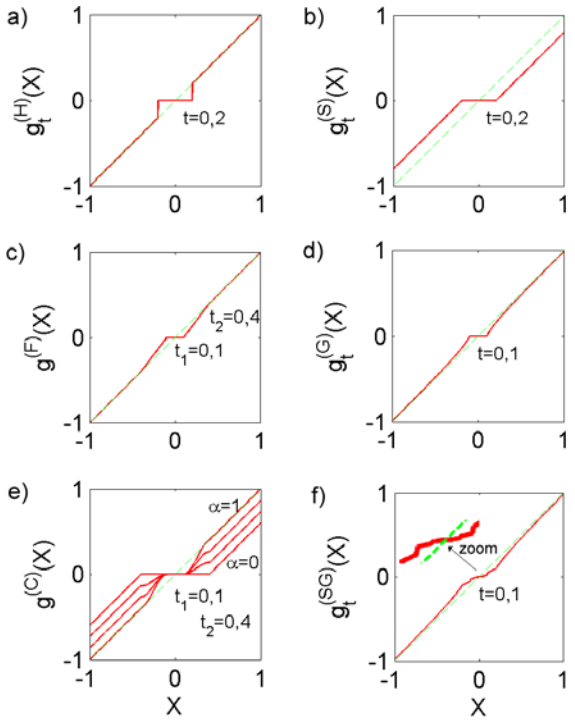


Figura 1: Funções de limiar **a)** *Hard* [2], **b)** *Soft* [2], **c)** *Firm* (*Semi-Soft*) [4], **d)** *garrote* [5], **e)** *Custom* [6] e **f)** *Step-Garrote* [7].

Na Fig. 1, são exibidas várias funções de limiar sugeridas por diversos autores, todas semelhantes ao formato da curva *Soft*. Deve-se ressaltar que em nenhuma delas são preservados coeficientes de pequena magnitude.

2 Obtenção das Funções de Limiar

Uma maneira alternativa de se chegar ao conceito de *thresholding* é considerar uma abordagem variacional ([8] *apud* [9]). Neste caso, a função de limiar $g(.)$ é o funcional que minimiza uma relação de compromisso entre dois objetivos, por exemplo, uma medida da fidelidade da aproximação e outra da suavidade do resultado. A fidelidade, em geral, é tomada como o erro médio quadrático; já a suavidade é mais difícil de se quantificar por critérios matemáticos. De qualquer modo, uma vez escolhidos os objetivos e fixados parâmetros de compromisso, obtém-se uma dada função de limiar.

2.1 Obtenção da função *Soft*

A regra *Soft* é o funcional que minimiza o erro quadrático penalizado pela norma L_1 do resultado [10, p.451], isto é,

$$\sum_{k=0}^{K-1}\left\{\left(S_m[k]-\hat{S}_m[k]\right)^2+2t\left|\hat{S}_m[k]\right|\right\}, \qquad (3)$$

ou seja, para cada bloco, dados os coeficientes do sinal sem ruído $S_m[k]$ e fixado o parâmetro t, a aplicação da regra $\hat{S}_m[k]=g_t^{(S)}(Y_m[k])$ é a solução que minimiza (3), com relação de compromisso igual a $2t$.

Neste trabalho, as funções $g(.)$ são restritas àquelas que modificam individualmente cada um dos coeficientes, por isto, a minimização de (3) pode ser feita de maneira unidimensional. Por exemplo, para um dado valor de coeficiente X e relação de compromisso $2t$ fixados, o valor G que minimiza

$$Q_1(G)=\left(\frac{X}{a_1}-\frac{G}{b_1}\right)^2+2t\left|\frac{G}{b_1}\right|, \qquad (4)$$

para $a_1=b_1=1$, é justamente a conhecida curva *Soft* com limiar t. Observe-se que, para maior generalidade, introduziram-se as constantes a_1 e b_1 que podem ser consideradas como um ganho (ou atenuação) aplicados, respectivamente, ao coeficiente de entrada e de saída. Normalmente, toma-se $a_1=b_1=1$ e admite-se que o sinal de saída sofra um encolhimento de amplitude.[2]

Na Fig. 2.a, é apresentado o resultado obtido por minimização numérica de (4) para limiar $t=0{,}2$ e $a_1=b_1=1$. Deve-se comentar que a variação dos ganhos (a_1, b_1) permite modificar a inclinação do trecho acima do limiar, mas ainda proporciona uma curva com o aspecto da função *Soft* pois o patamar permanece em zero. Desta maneira, os parâmetros a_1 e b_1 podem ser considerados como ajustes de ganho que normalmente não são incluídos na função *Soft*. Esses parâmetros podem ser usados para, por exemplo, corrigir as amplitudes dos coeficientes de saída para que o resultado não sofra encolhimento de amplitude.

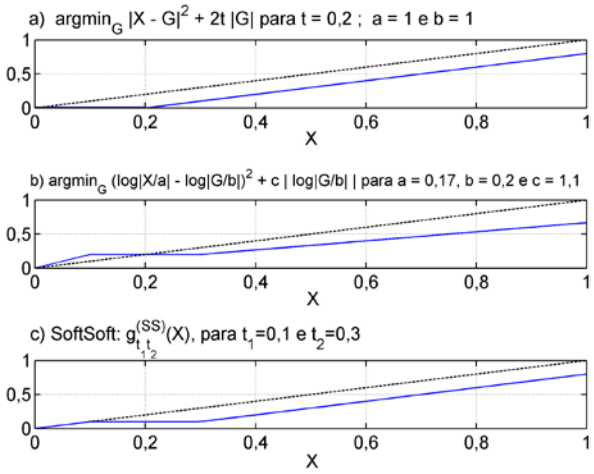


Figura 2: a) Função *Soft* para $t=0{,}2$ e $a_1=b_1=1$; b) Função *LogSoft* para $a_2=0{,}13$, $b_2=0{,}2$ e $c_2=1{,}1$; c) Função *SoftSoft* para $t_1=0{,}1$ e $t_2=0{,}3$.

[2] Na realidade, pode-se multiplicar (4) por $\sqrt{a_1}$ e obter expressão equivalente $\left(X-\frac{a_1}{b_1}G\right)^2+\frac{2t}{\sqrt{a_1}}\left|\frac{a_1}{b_1}G\right|$, na qual apenas a saída G é afetada por um fator $\frac{a_1}{b_1}$, sendo o limiar igual a $t/\sqrt{a_1}$. Alternativamente, sem perda de generalidade, poder-se-ia tomar $a_1=1$. O mesmo não ocorrerá para a função *SoftSoft*, definida a seguir, pois as constantes de ganho, neste caso, estão dentro de logaritmos.

2.2 Obtenção de função especializada para voz

A intensidade sonora percebida pelo sistema auditivo humano é mais bem descrita pelo logaritmo da potência do sinal. Motivado por este fato, é proposto investigar o valor G que minimiza a seguinte relação de compromisso

$$Q_2(G) = \left(\log\left|\frac{X}{a_2}\right| - \log\left|\frac{G}{b_2}\right|\right)^2 + c_2\left|\log\left|\frac{G}{b_2}\right|\right| \quad (5)$$

já escrita para o caso unidimensional. A Fig. 2.b exibe o formato da curva obtida por minimização numérica de (5) tomando parâmetros $a_2 = 0{,}13$, $b_2 = 0{,}2$ e $c_2 = 1{,}1$. Adicionalmente, no Apêndice I, é fornecida uma demonstração de que o argumento G que minimiza (5) é dado por

$$g^{(LS)}_{t,a_2,b_2}(X) = \begin{cases} b_2\frac{|X|}{a_2}e^{-t}\operatorname{sgn}(X), \text{ para } |X| > a_2e^t \\ b_2\operatorname{sgn}(X), \text{ para } a_2e^{-t} \le |X| \le a_2e^t \\ b_2\frac{|X|}{a_2}e^{t}\operatorname{sgn}(X), \text{ para } |X| < a_2e^{-t} \end{cases} \quad (6)$$

em que se empregou $c_2 = 2t$. A nova função de limiar, denominada *LogSoft*, depende de 3 parâmetros (t, a_2 e b_2) sendo a altura do patamar igual ao parâmetro b_2 e a largura igual a $2a_2\operatorname{senh}(2t)$.

É importante notar que como o logaritmo enfatiza coeficientes de pequena magnitude, resulta que estes coeficientes passam a ser considerados na composição do resultado, como se pode observar na parte próxima da origem da curva da Fig. 2.b. Para sinais de voz, em geral, os coeficientes de pequena magnitude representam sons não vocálicos que têm pouca energia mas que, em boa parte, contribuem para a inteligibilidade [11]. Tal característica não é observada em nenhuma das funções da Fig. 1, que sempre eliminam coeficientes abaixo de um certo limiar. Contudo, há uma regra de *thresholding*, já empregada para redução de ruído em sinais de voz [11], que pode ser usada como aproximação da *LogSoft* com a vantagem de necessitar do ajuste de apenas dois limiares.

2.3 Função *SoftSoft* especializada para voz

A função *SoftSoft* introduzida em [12] é assim chamada pois pode ser escrita pela combinação de duas funções *Soft*:

$$g^{(SS)}_{t_1,t_2}(X) = g^{(S)}_{t_2}(X) + \left(X - g^{(S)}_{t_1}(X)\right). \quad (7)$$

Na Fig. 2.c, pode-se observar que a curva *SoftSoft* aproxima o formato da Fig. 2.b com a vantagem de poder ser ajustada por apenas dois parâmetros: um limiar inferior (t_1) e um superior (t_2).

Neste trabalho, considera-se o uso da proposta *SoftSoft*, em comparação com o método *Soft*, para a redução de ruído em sinais de voz. Pelo fato de a *SoftSoft* se assemelhar à curva que minimiza de (5), espera-se que consiga reduzir a distância quadrática entre os logaritmos. Para avaliar esta distância, emprega-se uma medida de distorção log-espectral, a *LSD*.

A seguir, são apresentados resultados de simulação computacional para ambos os métodos, sendo os parâmetros das funções ajustados empiricamente, por meio de técnica de otimização numérica, de maneira a minimizar a *LSD* entre o resultado e o sinal sem ruído. O ajuste empírico é possível pois se supõe disponível o sinal sem ruído para avaliação de desempenho dos métodos comparados.[3]

3 Avaliação Experimental dos Métodos *Soft* e *SoftSoft*

3.1 Base de dados de voz e medidas de desempenho usadas

Do banco de dados de voz TIMIT [13], foi escolhida uma sentença foneticamente balanceada para ser usada como sinal original (sem ruído). Escolheu-se a frase em inglês *"clear pronunciation is appreciated"* falada por uma voz feminina.

Como o nível de ruído das gravações é muito baixo, pode-se considerar este sinal "limpo". O sinal teve a sua taxa de amostragem convertida para 8000 amostras/s, que é a taxa empregada nas simulações. Adicionalmente, tomou-se $M = 142$ blocos de $K = 128$ amostras e, como resultado, obteve-se um sinal com duração de 2,272 segundos. O sinal ruidoso é obtido acrescentando-se a ele *AWGN* com nível σ_w.

A razão sinal-ruído do resultado é calculada a partir de

$$\mathrm{SNR} = 10\log_{10}\left(\frac{\sum_{n=0}^{N-1}|s(n)|^2}{\sum_{n=0}^{N-1}|s(n)-\hat{s}(n)|^2}\right) \quad (8)$$

e também pode ser calculada para a entrada, bastando usar o sinal ruidoso $y(n)$ no lugar de $\hat{s}(n)$.

Por ser uma medida global do nível de ruído, a SNR não reflete bem a qualidade do sinal de voz obtido na saída. A qualidade é mais bem representada pelo valor médio da razão sinal-ruído dos blocos

$$\mathrm{SegSNR} = \frac{1}{M}\sum_{m=0}^{M-1}\mathrm{SNR}_m, \quad (9)$$

em que SNR_m é a razão sinal-ruído do m-ésimo bloco. A medida definida por (9) é conhecida como razão sinal-ruído segmentada. Caso, para algum bloco, ocorra logaritmo de zero, tal problema é contornado limitando-se os valores de SNR_m a uma faixa dinâmica de 40 a -40 dB, de maneira semelhante ao recomendado em [14].

[3] Numa implementação prática do método, faz-se necessária uma estimativa para os limiares. Por exemplo, como sugestão, poder-se-ia experimentar os valores estimados em [11] obtidos para minimizar o MSE mas que, possivelmente, também devem servir como estimativa para mínimo LSD.

Considerando-se aspectos perceptuais do sistema auditivo humano, o nível de ruído face ao sinal é mais bem avaliado pelo logaritmo das potências, como na *LSD*, calculada conforme sugerido em [14],

$$\mathrm{LSD} = \frac{1}{M}\sum_{m=0}^{M-1} D_m, \tag{10}$$

$$D_m = \sqrt{\frac{1}{\frac{K}{2}+1}\sum_{k=0}^{K/2}\left(10\log_{10}\frac{\mathsf{S}_m[k]}{\hat{\mathsf{S}}_m[k]}\right)^2},$$

em que $\mathsf{S}_m[k] \triangleq \max\left\{\left|\{\mathrm{FFT}(s_m)\}_k\right|^2, 10^{-10}\mathsf{S}_{\max}^2\right\}$ é a potência dentro do faixa dinâmica presumida de 100 dB e $\mathsf{S}_{\max}$ é o coeficiente *FFT* de maior magnitude dentre todos os blocos, $\mathsf{S}_{\max} = \max_{k,m}\{|\{\mathrm{FFT}(s_m)\}_k|\}$. A *FFT* de cada bloco é calculada com $K = 128$ pontos. De maneira análoga, calcula-se $\hat{\mathsf{S}}_m[k]$ (para encontrar a *LSD* da saída) ou emprega-se $\mathsf{Y}_m[k]$ no lugar de $\hat{\mathsf{S}}_m[k]$ (quando se for determinar a *LSD* da entrada).

3.2 Determinação dos valores dos limiares

A *LSD* do resultado $\hat{s}(n)$ pode ser calculada a partir do sinal limpo $s(n)$ e do sinal ruidoso $y(n) = s(n) + w(n)$, que é obtido pela adição de uma realização $w(n)$ de *AWGN*, sendo o nível de ruído σ_w ajustado para proporcionar uma *SNR* de entrada desejada. Por exemplo, para sinal ruidoso de entrada com $SNR = 3$ dB, os valores de *LSD* podem ser calculados em função dos limiares t_1 e t_2 da função *SoftSoft*. A Fig. 3 exibe as curvas de nível assim obtidas. Nesta figura, deve-se notar que, para tornar os valores dos limiares independentes da amplificação do sinal, empregam-se limiar inferior e superior normalizados, respectivamente, $\bar{t}_1 \triangleq t_1/t_2$ e $\bar{t}_2 \triangleq t_2/Y_{\max}$, sendo $Y_{\max}$ a maior magnitude de coeficiente, $Y_{\max} = \max_{k,m}|Y_m[k]|$.

Ainda na Fig. 3, percebe-se que há um ponto de mínimo (a) que é alcançado empregando a função *SoftSoft* com $\bar{t}_1 = 0{,}0420$ e $\bar{t}_2 = 0{,}0504$. Já a função *Soft*, que corresponde à *SoftSoft* com limiar inferior nulo, atinge o mínimo *LSD* em (b), ou seja, $\bar{t}_1 = 0$ e $\bar{t}_2 = 0{,}0272$, que é igual ao parâmetro $\bar{t}$ da *Soft*.

Utilizando o método *SoftSoft,* a determinação dos pontos de mínimo *LSD* foi refeita para diversos níveis de ruído, SNR entre -3 e 20 dB, sendo as posições destes pontos expostas na Fig. 4. Novamente, deve-se alertar quanto à normalização dos limiares que, na Fig. 4.a, foram ambos normalizados em relação ao coeficiente de maior magnitude[4] e, na Fig. 4.b, segundo a normalização anteriormente introduzida e denotada por $\bar{t}_1$ e $\bar{t}_2$.[5] De maneira semelhante, também foram encontrados os pontos de mínimo *LSD* para o caso *Soft* e

[4]Esta normalização permite observar o comportamento do limiar inferior de maneira independente do limiar superior.

[5]As duas formas de normalização sugerem a possibilidade de encontrar uma expressão empírica relacionando os dois limiares, podendo ser útil numa implementação prática do método.

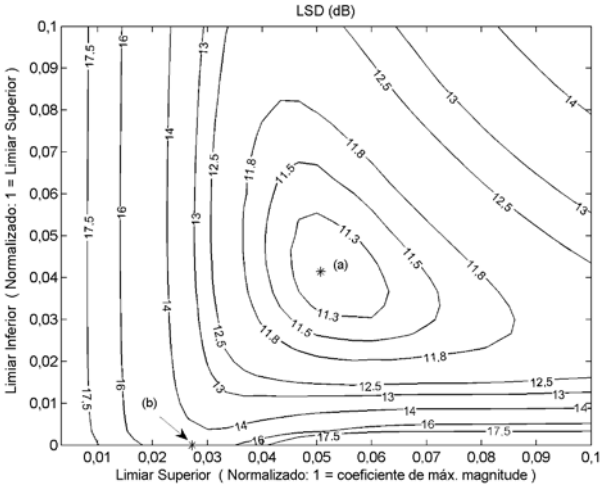


Figura 3: Curvas de nível da superfície da *LSD* em função do limiar inferior ($\bar{t}_1$) e superior ($\bar{t}_2$) para $SNR = 3$dB.

determinados os valores ótimos de limiar $\bar{t}$, normalizados em relação ao coeficiente de maior magnitude. Os resultados experimentais são apresentados e discutidos a seguir.

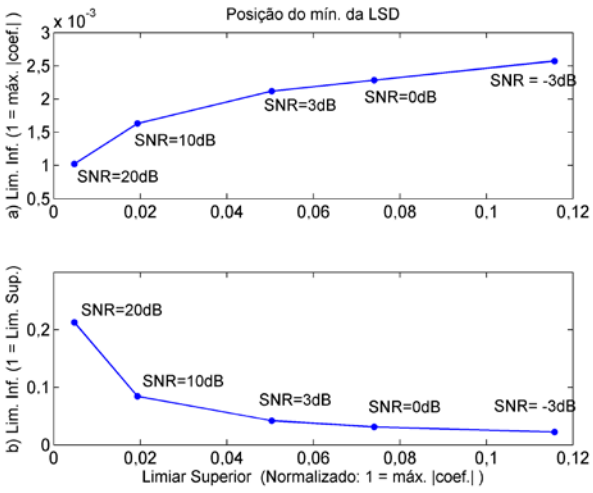


Figura 4: Valores de limiar da função *SoftSoft* que minimizam a LSD para $SNR = 20, 10, 3, 0$ e -3dB, com **a)** os dois limiares normalizados em relação ao coeficiente de maior magnitude e **b)** limiares $\bar{t}_1$ e $\bar{t}_2$.

3.3 Resultados experimentais: método *Soft versus SoftSoft*

A Tabela 1 apresenta os valores de SNR, SegSNR e LSD para o sinal ruidoso de entrada e, em comparação, os respectivos resultados proporcionados pelos métodos *Soft* e *SoftSoft* empregando limiares ótimos, isto é, que produzem mínima distorção (*LSD*).

Como se pode constatar na tabela, de fato, a *SoftSoft* consegue atingir uma distorção menor do que a

Tabela 1: Comparação entre os métodos *Soft* e *SoftSoft* para "Sinal Ruidoso de entrada" com diversas SNR. (Média de 10 realizações do ruído. Comprimento: $N = M \times K = 142 \times 128$.)

	SNR	SegSNR	LSD
Sinal Ruidoso de entrada	20,0	9,7	8,7
Soft, $\bar{t} = 0{,}0021$	21,3	11,3	8,5
SoftSoft, $\bar{t}_1 = 0{,}2128$, $\bar{t}_2 = 0{,}0048$	20,6	11,1	7,0
Sinal Ruidoso de entrada	10,0	-0,3	14,5
Soft, $\bar{t} = 0{,}0091$	13,1	3,4	12,6
SoftSoft, $\bar{t}_1 = 0{,}0841$, $\bar{t}_2 = 0{,}0194$	12,1	4,2	9,6
Sinal Ruidoso de entrada	3,0	-7,3	19,5
Soft, $\bar{t} = 0{,}0272$	8,1	-1,0	15,4
SoftSoft, $\bar{t}_1 = 0{,}0420$, $\bar{t}_2 = 0{,}0504$	7,4	0,8	11,2
Sinal Ruidoso de entrada	0,0	-10,2	21,9
Soft, $\bar{t} = 0{,}0430$	6,2	-2,6	16,5
SoftSoft, $\bar{t}_1 = 0{,}0308$, $\bar{t}_2 = 0{,}0741$	5,7	-0,3	11,8

Soft. A vantagem torna-se mais expressiva na medida em que se aumenta o nível de ruído. (Por exemplo, para $\mathrm{SNR} = 0\,\mathrm{dB}$, a diferença de LSD é igual a $16{,}5 - 11{,}8 = 4{,}7\,\mathrm{dB}$.) Já a SNR, que de fato não foi otimizada, não é significativamente maior no caso *Soft*. Além do mais, a SegSNR, que é uma medida mais adeguada de qualidade, sempre é maior para o caso *SoftSoft*, exceto para $\mathrm{SNR} = 20\,\mathrm{dB}$, que é um nível de ruído não muito alto.

Quanto à avaliação subjetiva dos sinais, ajustados para o mesmo nível de intensidade sonora, um ouvinte informal, em geral, avalia que o *SoftSoft* apresenta menor distorção e menos presença de ruídos espúrios (na forma de "ruído musical") do que o resultado do método *Soft*. Além do mais, no caso *SoftSoft*, o ruído de fundo é percebido como se fosse branco, possivelmente em virtude da preservação dos coeficientes transformados de pequena magnitude.

4 Conclusões e Perspectivas

Neste trabalho, ofereceu-se uma revisão da abordagem variacional para obtenção de funções de limiar, tendo-se destacado funções baseadas na *Soft*. Nessa revisão, ao se propor o uso do logaritmo da intensidade dos sinais, obteve-se a curva *LogSoft,* cujo formato, com valores de limiar normalmente empregados, assemelha-se à função *SoftSoft* especializada para voz. Através de simulações computacionais verificou-se que o método *SoftSoft* fornece menor distorção (log-espectral) sendo, portanto, mais indicado para aplicações de redução de *AWGN* em voz. Deve-se acrescentar que o *SoftSoft*, possivelmente, também deve ser mais indicado para um método que trate o caso de ruído não branco, no qual o ajuste dos limiares dependerá da potência de ruído em cada coeficiente.

É oportuno comentar que em [11] foi investigado o uso de *SoftSoft* para minimizar o erro médio quadrático (*MSE*), sendo oferecidos métodos para estimar os valores ótimos de limiar em [12]. E, quando o objetivo é minimizar o MSE, o *SoftSoft* também mostrou-se mais adequado para redução de ruído *AWGN* em voz do que o método *Soft*.

Os resultados aqui alcançados motivam a realização de um estudo do uso de *LogSoft* que, talvez, ainda possa aprimorar a qualidade do sinal obtido. No entanto, deve-se alertar quanto a algumas dificuldades, dentre elas: a maior complexidade de se otimizar três parâmetros; a inexistência de métodos para estimar os três parâmetros ótimos; o desconhecimento da topologia da superfície *LSD* (ou *MSE),* que pode não ser convexa. Estas questões são possíveis tópicos para investigação futura que motivam a busca por outros métodos especializados.

REFERÊNCIAS

[1] S. V. Vasegui, "Advanced Signal Processing and Noise Reduction", *John Wiley & Sons*, 2nd edition, 2000.

[2] D. L. Donoho, and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage", *Biometrika*, vol. 1, pp. 425-455, 1992.

[3] L. Rabiner, and B. H. Juang, "Fundamentals of speech recognition", *Prentice Hall*, 1st edition, 1993.

[4] H. Y. Gao, and A. G. Bruce, "Waveshink with firm shrinkage", *Statistica Sinica*, vol. 7, pp. 855-874, 1997.

[5] Leo Breiman, "Better subset regression using the nonnegative garrote", *Technometrics*, 37, 373-384, 1995.

[6] Byung-Jun Yoon, P.P. Vaidyanathan, "Wavelet-Based Denoising by Customized Thresholding", *ICASSP Conference*, 2004. IEEE.

[7] S. Aayat, M. T. Manzuri, and R. DIANNAT, "Wavelet based speech enhancement using a new thresholding algorithm", *Proc. of Int. Symp. on Intelligent Multimedia, Video and Speech Processing*, pp. 238-241, Oct. 2004.

[8] R. A. DeVore, and B. J. Lucier,"Fast wavelet techniques for near-optimal image processing", *IEEE Military Communications Conference Record*, SanDiego, Oct.11-14,1992, vol.3, pp.1129-1135.

[9] Byung-Jun Yoon, P.P. Vaidyanathan, "Wavelet-Based Denoising by Customized Thresholding", *ICASSP Conference*, 2004. IEEE.

[10] S. Mallat, "A Wavelet Tour of Signal Processing", *Academic Press*, 2nd edition, 1999.

[11] I. Antunes Jr., and P. M. S. Burt, "Speech Denoising by SoftSoft Thresholding", *Proc. ISIE, IEEE International Symposium on Industrial Electronics*, v.1., pp. 532-536, Montreal, 2006.

[12] I. Antunes Jr., "Redução de ruído em sinais de voz usando curvas especializadas de modificação dos coeficientes da transformada em co-seno.", *Tese (Doutorado)*, Escola Politécnica da Universidade de São Paulo, Departamento de Telecomunicações e Controle, 2006.

[13] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallet, and N. L. Dahlgren, "The darpa timit acoustic-phonetic continuous speech corpus cdrom", Cdrom, NIST, Gaithersburg, MD, 1996.

[14] I. Cohen, "Relaxed Statistical Model for Speech Enhancement and a Priori SNR Estimation", *IEEE Transactions on Speech and Audio Processing*, vol. 13, Issue 5, Part 2, Sept. 2005 pp. 870 - 881.

APÊNDICE I

Conforme visto, dados os números reais $\mathcal{X}$ e $t > 0$, o valor $\mathcal{G}$ que minimiza a expressão

$$Q_1(\mathcal{G}) = (\mathcal{X} - \mathcal{G})^2 + 2t\,|\mathcal{G}|, \tag{11}$$

é dado pela função *Soft*, ou seja, basta tomar $\mathcal{G} = g_t^{(S)}(\mathcal{X})$. Substituindo $\log\left|\frac{X}{a}\right| \triangleq \mathcal{X}$, $\log\left|\frac{G}{b}\right| \triangleq \mathcal{G}$ e $2t \triangleq c$ em (11), vem

$$Q_2(G) = \left(\log\left|\frac{X}{a}\right| - \log\left|\frac{G}{b}\right|\right)^2 + c\left|\log\left|\frac{G}{b}\right|\right| \tag{12}$$

que é justamente a expressão (5) cuja minimização numérica forneceu a função de limiar especializada para voz da Seção 2.2. Como, neste caso, a solução também é dada pela função *Soft*, basta tomar

$$\log\left|\frac{G}{b}\right| = g_t^{(S)}\left(\log\left|\frac{X}{a}\right|\right),$$

ou melhor, segundo (2),

$$\log\left|\frac{G}{b}\right| = \begin{cases} \log\left|\frac{X}{a}\right| - t, \text{ para } \log\left|\frac{X}{a}\right| > t \\ 0, \text{ para } -t \leq \log\left|\frac{X}{a}\right| \leq t \\ \log\left|\frac{X}{a}\right| + t, \text{ para } \log\left|\frac{X}{a}\right| < -t \end{cases}.$$

Agora, supondo que o logaritmo esteja na base e, esta solução pode ser reescrita, fornecendo uma expressão de função de limiar especializada para voz:

$$g_{t,a,b}^{(LS)}(X) = \begin{cases} b\frac{|X|}{a}e^{-t}\operatorname{sgn}(X), \text{ para } |X| > ae^{t} \\ b\operatorname{sgn}(X), \text{ para } ae^{-t} \leq |X| \leq ae^{t} \\ b\frac{|X|}{a}e^{t}\operatorname{sgn}(X), \text{ para } |X| < ae^{-t} \end{cases}. \tag{13}$$

onde $\operatorname{sgn}(.)$ é a função *signum*. Chamou-se a função $g_{t,a,b}^{(LS)}(x)$ de *LogSoft* pois ela foi obtida da solução *Soft* usando logaritmos. Observe-se que a altura do patamar, neste caso, é igual ao parâmetro b; já a largura, depende dos parâmetros t e a, podendo ser calculada por $ae^{t} - ae^{-t} = 2a\operatorname{senh}(t)$. Note-se, também, que o uso de logaritmo em outra base não altera a minimização de (12), uma vez que pode ser compensado por uma redefinição dos parâmetros.

Por fim, deve-se comentar que é possível confirmar que a *LogSoft*, definida em (13), é de fato a solução que minimiza $Q_2(G)$. Para isto, basta verificar que ela anula a primeira derivada de $Q_2(G)$ e, também, que resulta em segunda derivada de $Q_2(G)$ positiva.

Sociedade de Engenharia de Áudio
Artigo de Congresso

Apresentado no 9º Congresso de Engenharia de Áudio
15ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2011, São Paulo, SP

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

Implementações em *downmix* passivo de 5.1 para 2.0

José Augusto Mannis
Departamento de Música UNICAMP
Campinas, SP, 13083-970, Brazil
jamannis@uol.com.br

RESUMO

Este artigo apresenta dois tipos de *downmix* empregados na atualidade, com alguns desdobramentos e variantes, sobretudo uma implementação desenvolvida a partir de trabalhos práticos realizados entre 2007 e 2008 [1][2], discutida, analisada e sintetizada. Aspectos como correlações entre parâmetros de *downmix* e materiais sonoros, ambientação sonora da mixagem, gêneros e estilos das realizações sonoras são aqui discutidos tendo em vista desdobramentos desta pesquisa em futuros trabalhos.

0 INTRODUÇÃO

Em produções recentes em suporte multicanal realizadas em 2007 em Londrina (PR), Festival de Londrina, e 2008, no âmbito da disciplina de Composição do Depto. de Música do Inst. de Artes da Unicamp, surgiu a necessidade de realização de versão em dois canais para veiculação em radio ou reprodução em sistemas de áudio corriqueiros. Tratava-se de criações musicais com narração nas quais a presença e clareza da fala eram fundamentais. A redução 2.0 deveria proporcionar a melhor audição possível do jogo espacial realizado em 5.1. Depois de várias tentativas, incluindo indicações de *downmix* recomendadas pela ITU, a melhor solução acabou sendo uma regulagem um pouco mais complexa, apresentada neste artigo. Assim o objetivo deste trabalho é apresentar os procedimentos e considerações ocorridos durante as referidas produções buscando, sempre que possível, fundamentá-los em conhecimentos teóricos, bem como apontar para os próximos passos desta pesquisa e seus possíveis desdobramentos. Nesse sentido, outro trabalho complementar a este, focado num modelo linear mais geral do que o aqui apresentado, está sendo submetido a este congresso como extensão aos desenvolvimentos aqui expostos, envolvendo testes subjetivos com diversos ouvintes, podendo apontar para novas possibilidades de controle efetivo.

1 SISTEMAS ESTEREOFONICOS 5.1 SURROUND E 2.0

Considera-se aqui o principio da estereofonia como uma questão essencialmente centrada no relevo sonoro, envolvendo definição das fontes sonoras e profundidade da imagem, independente do número de canais colocados em obra. A disposição dada pela Recomendação ITU-R-BS.775-2 [3] complementada pela ITU-R-BS.1116-1 [4] compreende cinco caixas acústicas situadas ao redor do ouvinte, a saber: *L* (cana frontal esquerdo), *C* (canal central frontal), *R* (canal direito central), *Ls* (canal esquerdo *surround*), *Rs* (canal direito *surround*)[1]. Não trataremos aqui do

[1] Caixas posteriores do 5.1

sub-woofer pois este não influencia na distribuição espacial da imagem sonora, sendo os comprimentos de onda transmitidos abaixo de 120Hz maiores que a altura de um ouvinte e mais de 20 vezes maiores que a distância entre os ouvidos. Como não atua na percepção espacial do som, o sinal do *sub-woofer* não necessita ser tratado em *downmix*, podendo ser passado diretamente para o canal de *sub-woofer* em formato 2.1, quando for o caso. Os dispositivos considerados 5.1 e 2.0 são os mais corriqueiros, mas o que for estabelecido para 2.0 vale para 2.1, assim como o que for estabelecido para 5.1 vale para 5.0.

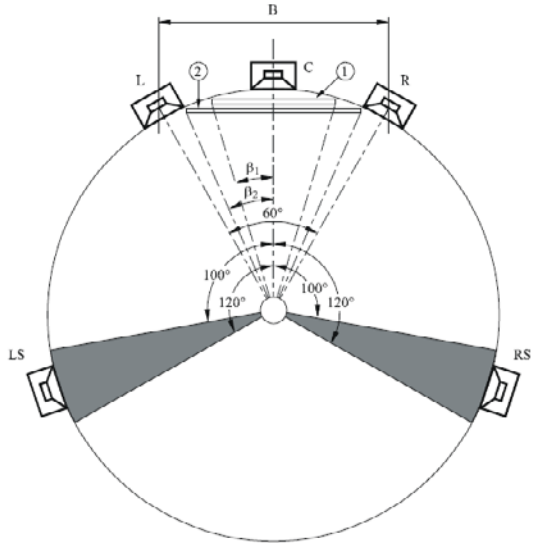


Figura 1 – Disposição de referência das caixas acústicas *L/C/R* e *Ls/Rs* [3]

A disposição das caixas *L* e *R* para escuta em 2.0 é exatamente a mesma que em 5.1, pois conforme Blauert [8] o ângulo frontal de ±30º a 60º são os mais estanques, permitindo a menor interferência da caixa esquerda no ouvido direito e vice-versa. Observa-se que as caixas *Ls* e *Rs* situam-se igualmente numa região com pouca interferência interaural, sobretudo próximo a 120º (cf. Figura 2) o que justifica parcialmente a escolha de alguns produtores, notadamente nos EUA e em sistemas de escuta pessoais, em estendê-las pouco além desse ângulo.

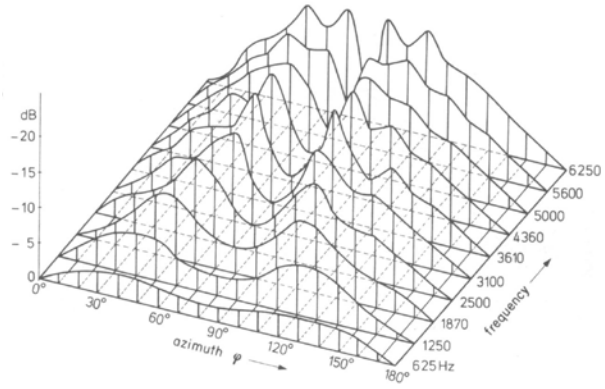


Figura 2 - Diferença interaural de nível sonoro.[6] (p.73)

2 REFERÊNCIAS : DOWNMIX DE 5.1 PARA 2.0

Downmix Left total/ Right total (L_t/R_t)

Destinado à redução de 5.1 para 2.0 com codificação (processamento de fase) de forma que o *downmix* possa ser lido e decodificado obtendo novamente uma reprodução em 5.1. Esta redução em 2.0 pode ser ouvida em equipamentos *hi-fi* ou fones de ouvido, mas, não é apropriada para escutas em mono, pois as informações dos sinais surround somadas eletricamente se cancelam em fase. A oposição de fase entre *Ls* e *Rs* desta configuração produz relevo sonoro na escuta somente quando cada um dos ouvidos recebe separadamente os sinais defasados, um em relação ao outro, por via aérea.

$$L_t = L + 0.71C + 0.71(-Ls - Rs) \qquad (1)$$

$$R_t = R + 0.71C + 0.71(Ls + Rs) \qquad (2)$$

Ls e *Rs* reproduzidos em *Lt* e *Rt* com diferença de fase de 90º

Downmix Left only / Right only (L_o/R_o)

Destinado à audição do produto sonoro sem a necessidade de recuperar o sinal para posterior *upmix* por decodificação. Nesta configuração temos a compatibilidade mono do registro sonoro, podendo ser ouvido integralmente, com todo seu conteúdo, em qualquer tipo de equipamento de áudio, mesmo através do alto-falante de celulares. Os canais frontais e *surround* não são separados por diferença de fase.

$$L_o = L + 0.71C + \delta Ls \qquad (3)$$

$$R_o = R + 0.71C + \delta Rs \qquad (4)$$

δ = 0.71 (-3dB); 0.50 (-6dB); 0.36 (-9dB)
(atenuação dos sinais surround)

Downmix Recomendação ITU

A partir do exposto na recomendação ITU-R-BS.775-2 [3], as equações de referência para *downmix* expressas na recomendação ITU-R BS.1116-1 [4] são:

$$L_0 = 1.00\ L + 0.71\ C + 0.71\ Ls \qquad (5)$$

$$R_0 = 1.00\ R + 0.71\ C + 0.71\ Rs \qquad (6)$$

Trata-se, portanto, de um caso particular de *Left only / Right only* (*Lo/Ro*). Os sinais *Ls* e *Rs* são reduzidos de 3dB (metade de sua energia: (-3dB=20.log0,71) e adicionados respectivamente a cada um dos canais *Lo* e *Ro*. O sinal do canal *C* é diminuído de 3dB e

adicionado a *Lo* e *Ro*. Na reprodução, ambos os sinais reduzidos se somam nos ouvidos do ouvinte formando uma imagem sonora exatamente ao centro e ganhando +3dB, voltando portanto a seu nível sonoro original. Os sinais originais dos canais *L* e *R* permanecem no 2.0 tal qual estavam no 5.1, sem nenhum processamento.

3 IMPLEMENTAÇÕES

Thompson et al.[5] descrevem um processo de *downmix* passivo (*Lt/Rt*) [6][7] com o seguinte sistema linear:

$$L_t = 1.00\ L + 0.71\ C + j(\cos\alpha\ .\ Ls + sen\alpha\ .\ Rs) \quad (7)$$

$$R_t = 1.00\ R + 0.71\ C - j(sen\alpha\ .\ Ls + \cos\alpha\ .\ Rs) \quad (8)$$

sendo *j* uma defasagem de 90º e α um ângulo de 30 a 35º.

Trata-se portanto de um *downmix Left total/ Right total (Lt/Rt).* Thompson [5] salienta que o desempenho dessa configuração depende das características dos sinais de *input*, o que é reforçado por Blauert [8][2] quando nos mostra que para um mesmo valor de parâmetro (intensidade ou atraso) aplicado a um som para situá-lo no espaço de escuta temos efeitos distintos conforme a natureza do material sonoro (fala, impulsos de banda larga, impulsos de banda estreita, sons puros (senoidais)). Portanto, a percepção espacial da escuta humana altera seu desempenho em função da natureza do som percebido e tem sua capacidade de análise alterada diante de diferentes materiais sonoros e timbres. Thomson mostra ainda como as características de mixagem influenciam no formato de *downmix* preferido na avaliação dos ouvintes, o que está de acordo com nossa hipótese de que o tratamento da mixagem e a ambientação sonora da mesma é um fator influente e a ser considerado no *downmix*.

Nosso procedimento de implementação em *downmix* 5.1 para 2.0 em 2007 e 2008 posicionou os canais *Ls* e *Rs* mais ao centro como no modelo *Lt/Rt* apresentado por Thompson [5], separando-os de uma posição coincidente com a imagem dos canais originais *L* e *R.*

Em nosso experimento exploramos o efeito de sombra acústica correspondendo a um ângulo de 120º (filtragem atenuando as bandas a partir de 2KHz) para o tratamento geral de cada sinal *surround*. Nesses primeiros ensaios a filtragem foi aplicada simultaneamente para os sinais *surround* em dirigidos a ambos os canais do *downmix* (*Lo* e *Ro*). Ulteriormente será experimentada uma filtragem diferenciada correspondendo mais realisticamente ao fenômeno da sombra acústica (cf. Figura 4), ou seja do sinal *Ls* uma filtragem mais acentuada para o canal *Ro* (pois o ouvido direito recebe um som mais apagado do canal *Ls*) e menos acentuada para o canal *Lo* (pois o ouvido esquerdo recebe o som de *Ls* mais diretamente).

Adotamos ainda, além da filtragem, uma atenuação de *Ls* e *Rs*, buscando estabelecer um plano dinâmico diferenciado da cena sonora frontal para representar o hemisfério posterior do ouvinte no seu espaço de escuta. O *downmix* realizado resume-se às seguintes fórmulas:

$$L_D = 1.00\ L + 0.71\ C + f_{FLT}\,(2\text{KHz})\ [\,f_{PAN}\,(0.75)\ Ls + f_{pan}\,(0.25)\ Rs\,] \quad (9)$$

$$R_D = 1.00\ R + 0.71\ C + f_{FLT}\,(2\text{KHz})\ [\,f_{PAN}\,(0.25)\ Ls + f_{pan}\,(0.75)\ Rs\,] \quad (10)$$

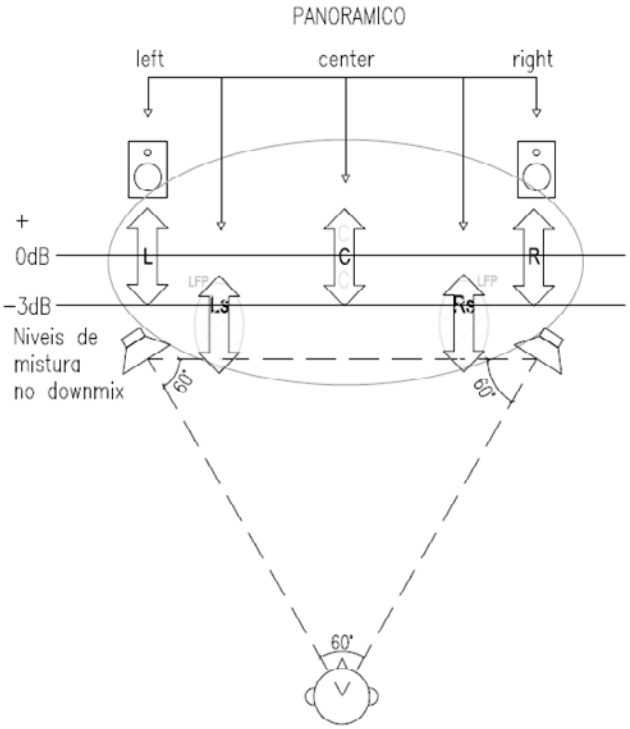


Figura 3 - Ilustração das implementações propostas para *downmix* de 5.1 para 2.0

Observamos nos experimentos que conforme a natureza do registro sonoro, de seu material, de seus timbres, de sua mixagem, das estratégias de espaço adotadas na mixagem, as regulagens dos parâmetros de *downmix* tiveram que ser afinadas.

No caso de um registro mesclando fala com música, o nivel da fala, predominantemente no canal central, teve que ser aumentado de 3dB para melhorar a inteligibilidade.

$$L_D = 1.00\ L + \chi\ C + f_{FLT}\,(2\text{KHz})\ [\,f_{PAN}\,(\phi)\ Ls + f_{pan}\,(1\text{-}\phi)\ Rs\,] \quad (11)$$

[2] p.204

$$R_D = 1.00\, R + \chi\, C + f_{FLT}\,(2\text{KHz})\,[\,f_{PAN}\,(1\text{-}\phi)\, Ls + f_{pan}\,(\phi)\, Rs\,] \quad (12)$$

$$0.71 \leq \chi \leq 1.41 \;\wedge\; 0.65 \leq \phi \leq 0.85$$

Interpretamos isso como uma compensação para o fechamento da cena sonora de 270º para 120º. As imagens das fontes sonoras estando muito mais sobrepostas necessitaram de reforço de intensidade para lhes dar mais clareza e inteligibilidade. Os parâmetros de intensidade e filtragem dos canais *surround* foram regulados empiricamente.

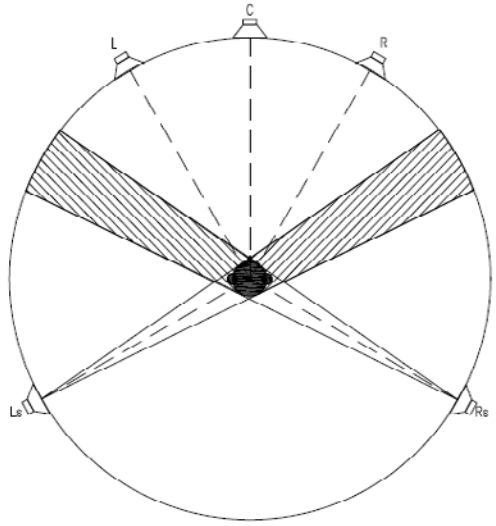


Figura 4 - Efeito de sombra acústica provocado pela cabeça humana sobre o sinal *Ls* para o ouvido direito e sobre o sinal *Rs* para o ouvido esquerdo.

4 DISCUSSÃO

Nenhum *downmix* de 5.1 para 2.0 substitui uma mixagem pensada para 2.0. As estratégias para uma mixagem em dois canais são diferentes daquelas aplicadas em situação multicanal. Neste trabalho priorizamos o *downmix* direcionado para uma melhor qualidade escuta, mais do que para uma mobilidade de *upmix* e *downmix* como o permite um formato *Lt/Rt* sobretudo codificavel e legível em sistema *Dolby Pro-Logic*. Assim, neste trabalho o *downmix* se caracteriza como um registro reduzido, uma imagem sonora de cinco[3] canais projetada em apenas dois canais, havendo compromisso com a inteligibilidade e clareza de seu conteúdo. Um dos ganhos em mixagem em 5.1 em relação a 2.0 é justamente a clareza, pois as fontes sonoras ganham um poderoso recurso para serem discernidas: sua localização espacial numa cena sonora de 270º contra uma cena de aproximadamente 120º em um registro 2.0. Em dois canais as fontes se mesclam e se amontoam umas sobre as outras havendo maior risco de perda de inteligibilidade. Em 5.1, estando tudo mais 'aberto', as fontes se acomodam mais confortavelmente no espaço. De forma que num *downmix* de 5.1 para 2.0 há um risco de perda de inteligibilidade. Nenhum *downmix* em dois canais chegaria a uma clareza e a uma separação de fontes tão boa quanto em uma mixagem em 2.0 pensada do começo ao fim para 2.0.

Os resultados obtidos na realização prática [1][2] foram satisfatórios do ponto de vista técnico (clareza e inteligibilidade da fala e da localização das fontes no espaço) e estético (não comprometendo a poética sonora de todos os trabalhos) envolvendo grande diversidade de autores, compositores, narradores e narradoras e uma ampla paleta tipo-morfológica de materiais sonoros.

Como desdobramento, talvez o processo de *downmix* possa ser usado como efeito ou um recurso de escritura sonora espacial em expressões artísticas, p.e. aplicado a cada par de caixas do *set* 5.0, onde um *downmix* diferente para cada par de caixas, proporcionaria um espaço sonoro plástico onde o percurso do ouvinte o colocaria diante de uma espécie de *Quadros de uma exposição*, o que caracterizaria uma realização em arte sonora.

5 CONCLUSÃO

Pretendemos a seguir proceder a uma avaliação por uma quantidade significativa de ouvintes de forma a estabelecer critérios para as regulagens dos parâmetros de *downmix* correlacionando-os à natureza do material sonoro processado bem como às situações e ambientes de mixagem dos arquivos sonoros originais em 5.1.

A próxima etapa deste trabalho envolve a implementação de um aplicativo associado a um dispositivo em tempo real permitindo a comparação do original 5.1 com seu *downmix* em 2.0 tendo os parâmetros reguláveis pelo ouvinte dentro de um campo restrito de possibilidades.

Além da busca pelos coeficientes mais adequados para determinados timbres, materiais sonoros e contextos de mixagem, poderemos também testar a eficiência do tratamento da sombra acústica aplicada mais fortemente no ouvido do lado oposto ao da caixa acústica emitindo o sinal *surround*. Este modelo é descrito pelo sistema linear[4]:

$$L_D = \alpha L + \beta R + \chi C + [f_{PAN}(\phi).\delta].Ls + f_{FLT}\,(2\text{KHz}).[f_{PAN}\,(1\text{-}\phi).\delta].Rs \quad (13)$$

$$R_D = \beta L + \alpha R + \chi C + f_{FLT}\,(2\text{KHz}).[f_{PAN}\,(1\text{-}\phi).\delta].Ls + [f_{PAN}(\phi).\delta].Rs \quad (14)$$

$$0.00 \leq \alpha, \beta, \chi, \delta, \phi \leq 1.00$$

[3]Lembramos mais uma vez que o *sub-woofer* não interfere na distribuição espacial dos sons, mas na coloração, no tônus e no calor da percepção.

[4] As variáveis destas equações estão normalizadas e, portanto, diferentemente das equações anteriores tem valores entre 0 e 1.

Este sistema será discutido amplamente em artigo complementar submetido a este congresso, propondo um suporte para experimentos com avaliação subjetiva.

BIBLIOGRAFIA

[1] MANNIS, J. A. *Downmix* de 5.1 para 2.0 de Criações sonoras em Atelier de composição. Festival de Londrina, PR. Julho de 2007.

[2] MANNIS, J. A. *Downmix* de 5.1 para 2.0 de Criações sonoras em Atelier de composição para veiculação radifônica e reprodução em equipamentos 2.0. Disciplina MU771 Depto. de Mùsica Unicamp. Campinas, SP, Junho de 2008.

[3] INTERNATIONAL TELECOMMUNICATION UNION. *Multichannel stereophonic sound system with and without accompanying picture*. Recommendation ITU-R BS.775-2. Genève: ITU, 2006. 11p.

[4] INTERNATIONAL TELECOMMUNICATION UNION. *Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems*. Recommendation ITU-R BS.1116-1. Genève: ITU, 1997. 26p.

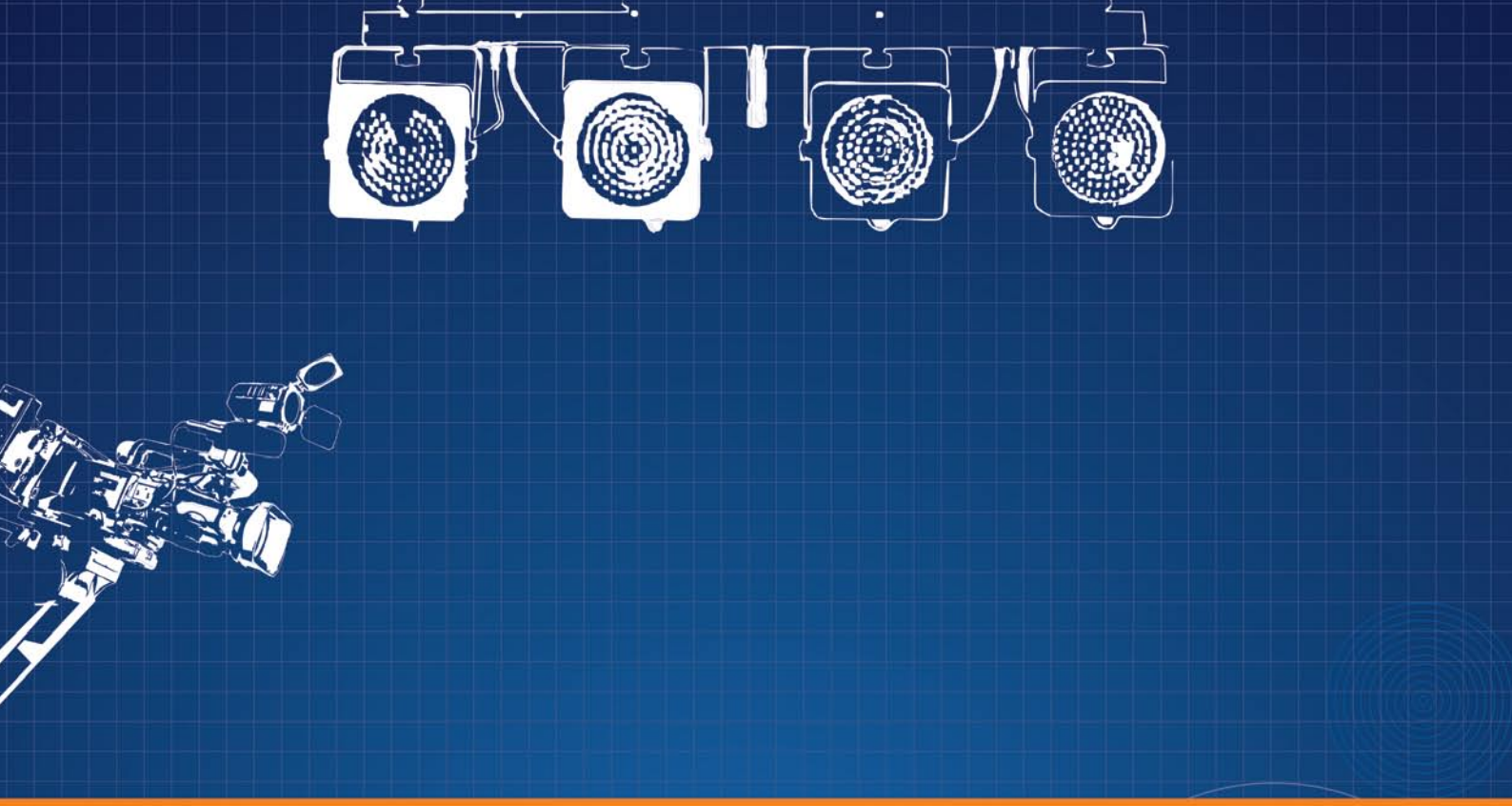
[5] THOMPSON; J. WARNER; A., SMITH, B. *An Active Multichannel Downmix Enhancement for Minimizing Spatial and Spectral Distortions,* 127th AES Convention, New York, NY, USA. October, 2009.

[6] *Dolby Surround Mixing Manual*, Issue 2, available: http://www.dolby.com/technicallibrary.aspx?id=306

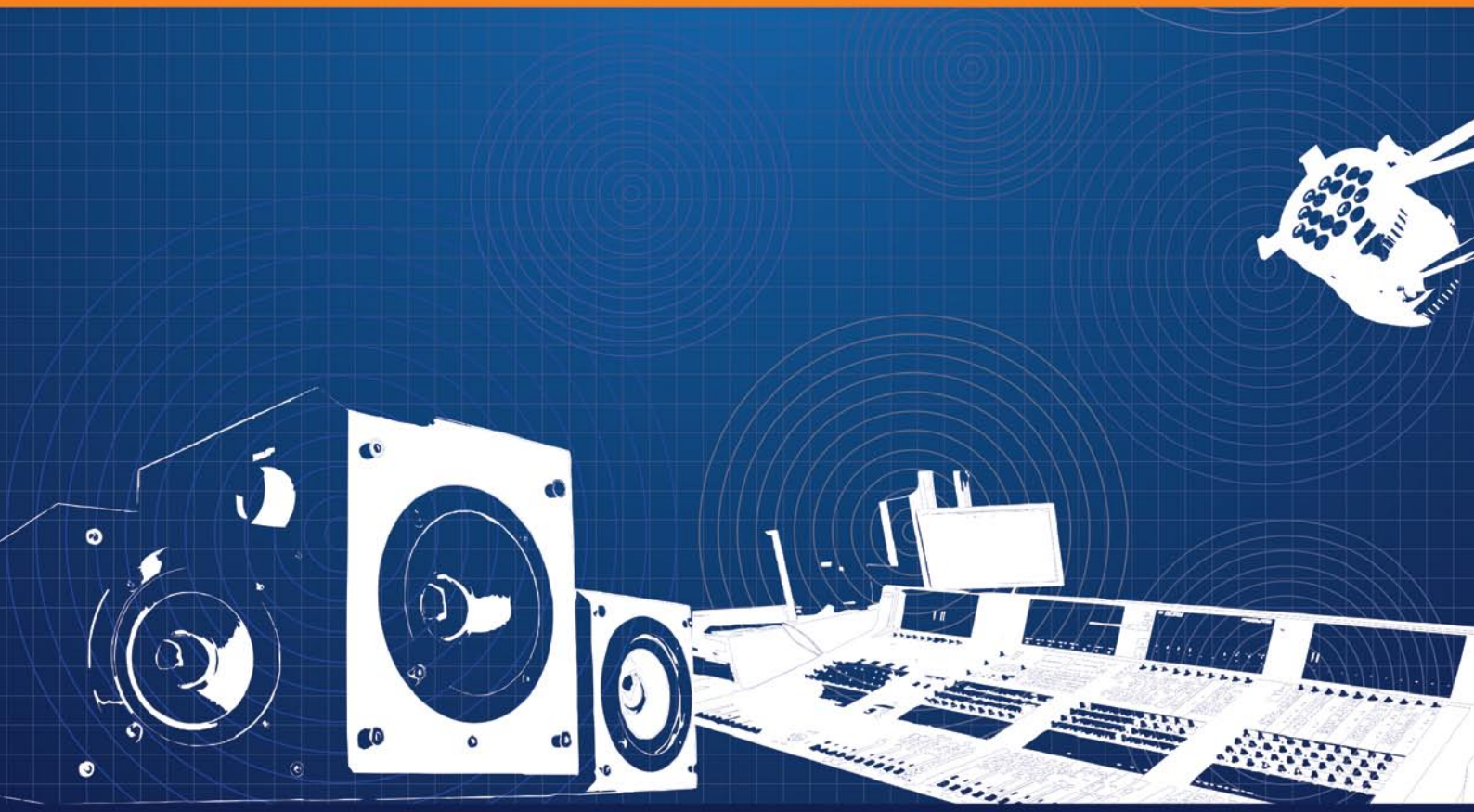
[7] GUNDRY, *A New Active Matrix Decoder for Surround Sound,* 19th International Conference:Surround Sound - Techniques, Technology, and Perception, (2001 June) (APUD THOMPSON et al., 2009).

[8] BLAUERT, J. *Spatial Hearing: the psycophysics of human sound localization.* Massachusetts: The MIT Press, 1997. 494p.

[9] ZIELINSKI, S. K., RUMSEY, F., and BECH, S., *Effects of Down-Mix Algorithms on Quality of Surround Sound,* J. Audio Eng. Soc., Vol. 51, pp. 780-798, (2003 Sept.) (APUD THOMPSON et al., 2009).



STUDENT PAPERS



Sociedade de Engenharia de Áudio

Artigo de Congresso

Apresentado no 9º Congresso de Engenharia de Áudio
15ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2011, 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.

Proposta de Estudo do Potencial Evocado Auditivo P300 para Averiguação dos Processos Cognitivos em Indivíduos com Ouvido Absoluto

Rodrigo Munhoz[1], César Augusto Mateus[1], Francisco J. Fraga[1], Patricia Vanzella[12]
[1]Universidade Federal do ABC (UFABC)
Santo André, São Paulo, CEP 09210-170, Brasil
[2]Universidade de Brasília (UnB), Departamento de Música
Brasília, Distrito Federal, CEP 70910-900
{cesar.amateus, rodrigo.munhoz, francisco.fraga}@ufabc.edu.br
pvanzella@unb.br

RESUMO

O ouvido absoluto é um traço cognitivo raro, que pode ser identificado em cerca de 0,01% da população. Portadores desse traço são capazes de identificar e/ou cantar, de forma espontânea e sem esforço, qualquer uma das notas musicais sem necessidade de compará-las a uma nota de referência. Neste artigo descrevemos a origem e os métodos de captação do potencial evocado auditivo de longa latência, P300, como resposta relacionada às funções mentais ligadas à percepção e atenção e, em seguida, propomos uma metodologia inicial para estudo do comportamento do P300 em portadores de ouvido absoluto. Para propiciar uma melhor compreensão dos conceitos envolvidos, apresentamos também um breve histórico do registro das atividades elétricas cerebrais, com foco nos potenciais evocados de córtex cerebral, mais especificamente dos potenciais evocados auditivos de longa latência.

0 INTRODUÇÃO

O registro da atividade elétrica cerebral de um indivíduo é atualmente conhecido como eletroencefalograma (EEG). Os sinais registrados pelo EEG resultam da somatória dos campos elétricos produzidos principalmente pelos potenciais pós-sinápticos de milhares de neurônios interconectados [1].

A transferência de informações entre os neurônios ocorre na forma de trens de pulso modulados em freqüência (potenciais de ação). Esta atividade neural em determinada região do cérebro depende de qual foi o sentido (audição, visão, tato, olfato ou paladar) ativado e do local do corpo que foi estimulado. A resposta elétrica ao estímulo sensorial surge em várias partes do cérebro e, dependendo do tipo de resposta e da região do corpo, também produz atividade elétrica na parte do corpo sensibilizada ou até mesmo em outras.

As modificações no EEG produzidas por um estímulo fixo e síncrono são consistentes e sincronizadas, enquanto a atividade eletroencefalográfica em geral (EEG "de fundo"), devida às sinapses de milhares de neurônios conectados em uma estrutura tridimensional, pode ser considerada como praticamente aleatória. Portanto, a

resposta cerebral (elétrica) a uma estimulação repetida e síncrona, que será a soma algébrica dos sinais consistentes gerados (potenciais evocados), aumenta em proporção ao número de sinais somados, enquanto que a amplitude da atividade aleatória do EEG tende a zero (cancelamento das ondas positivas com as negativas).

Desta maneira, se várias amostras de atividade elétrica neuronal, contendo tanto os potenciais evocados quanto os registros eletroencefalográficos aleatórios são digitalizadas, armazenadas e somadas pelo computador, considerando uma janela temporal (pós-estímulo) de duração pré-estabelecida, a amplitude dos potenciais evocados aumenta enquanto o traçado eletroencefalográfico de origem aleatória diminui, tornando-se praticamente nulo. Este processo de obtenção dos potenciais evocados a partir do EEG é conhecido na literatura como promediação [1].

O potencial evocado auditivo é o registro da atividade elétrica que ocorre no sistema auditivo, da orelha interna até o córtex cerebral, em resposta a um estímulo acústico. Nos últimos anos, a viabilidade técnica de registrar com precisão potenciais elétricos em vários níveis do sistema nervoso em resposta a estímulos acústicos tem produzido um grande número de aplicações clínicas relevantes para as áreas de otorrinolaringologia, audiologia e neurologia [1][2][3]. Estes potenciais evocados podem ser registrados por técnicas não invasivas, sem desconforto para o paciente e possibilitam avaliar as manifestações dos vários complexos eletrofisiológicos associados à ativação do sistema auditivo.

Em particular, nos últimos anos tem crescido muito a confiança dos profissionais da área de saúde auditiva nos procedimentos clínicos de captação e avaliação dos potenciais evocados auditivos de tronco encefálico (ABR – *Auditory Brainstem Response*). Os atuais testes clínicos de BERA – *Brainstem Evoked Response Audiometry* – têm a vantagem de permitir a avaliação objetiva e imparcial (independentemente da vontade do sujeito) da capacidade de audição em freqüências específicas, por meio da seleção de estímulos apropriados, e tem sido amplamente utilizada em audiologia, principalmente com a finalidade de detectar e avaliar precocemente as perdas auditivas em bebês [3].

Ao contrário do ABR, que exige equipamento especializado e mais preciso, o potencial evocado auditivo do córtex cerebral (CAEP - *Cortical Auditory Evoked Potential*) pode ser obtido a partir de um exame de EEG comum e reflete o processamento neural do estímulo acústico no córtex. Quando um estímulo auditivo alcança a orelha interna e é captado pelas células ciliadas no interior da cóclea, inicia-se a atividade auditiva neural no VIII par de nervos cranianos (nervo auditivo), passando pelo núcleo coclear e pelo corpo trapezoidal e seguindo para o complexo olivar superior e o lemnisco lateral. A resposta neuronal segue então para o colículo inferior e o núcleo geniculado medial, alcançando finalmente o córtex auditivo primário aproximadamente 100 ms após o início do estímulo [4].

De acordo com Näätänen e Picton (1987), para a maioria dos humanos adultos as propriedades do sinal de CAEP são semelhantes quando é usada a mesma metodologia de medição e quando o processamento auditivo central é normal [5]. O CAEP tem uma forma de onda com latências (atrasos) específicas relativas ao início do estímulo. Os picos desta onda recebem denominações de acordo com a sua polaridade (positiva ou negativa) e a seqüência com que são registrados (da menor para a maior latência), assim temos: P1, N1, P2, N2, P3 e N3.

As porções N1 e P2 do CAEP são as mais estáveis, com picos de latência em 100 ms e 175 ms, respectivamente. Variações tanto na latência como na amplitude dos picos dependem do estímulo auditivo, além de diversos outros fatores: idade, sexo, orelha estimulada, capacidade auditiva, atenção ao experimento, intensidade, freqüência e taxa de repetição do estímulo. À medida que o nível do estímulo diminui e se aproxima do limiar de audição, as ondas CAEP aparecem cada vez mais fracas e mais atrasadas, pois o processamento auditivo no córtex torna-se mais lento para sons mais fracos (Figura 1) [6].

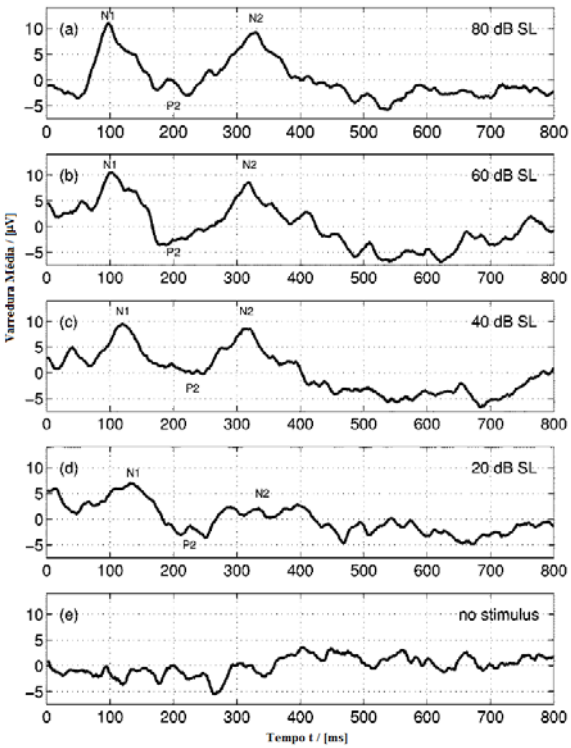


Figura 1 – Sinais calculados por promediação de EEG sobre 100 varreduras, obtidos de um mesmo indivíduo: (a)–(d) sinais de CAEP para estímulo (aplicado em t=0) com intensidades de 80, 60, 40, e 20 dB acima do limiar auditivo; (e) ausência de estímulo e, portanto, de CAEP (extraído de [6].)

1 POTENCIAIS EVOCADOS AUDITIVOS

O CAEP é a resposta do córtex cerebral a um estimulo auditivo. A medição desta resposta se dá pela gravação da atividade elétrica cortical, sendo que respostas evocadas (PE) refletem o potencial de campo elétrico que consiste do somatório da atividade elétrica entre membranas de milhares de neurônios, localizados em intervalos de fibras e nervos ou do núcleo do Sistema Nervoso Central.

O registro do PE pode ser feito de duas maneiras: utilizando-se eletrodos não invasivos, colocados sobre a superfície da pele (eletrodos de escalpo), ou utilizando-se eletrodos invasivos, colocados diretamente onde se quer medir o PE, e.g. em regiões intracranianas. Em humanos, geralmente opta-se pelo método não invasivo, pois medir potenciais em regiões internas do cérebro normalmente só é possível durante procedimentos cirúrgicos.

Quando o eletrodo é colocado próximo ao ponto do dipolo obtêm-se melhor resolução na gravação da forma de

onda do PE. Em estudos com eletrodos intracranianos [7][8] as respostas evocadas gravadas próximas ao ponto gerador do potencial de campo (dipolo) são chamadas de respostas de campo próximo (*near-field*) e respostas evocadas gravadas com maior distância do dipolo gerador são chamadas de respostas de campo distante (*far-field*).

Portanto, o fator determinante na captação destas respostas está na dificuldade em se obter respostas de campo próximo em humanos, pois para que se possa obter este tipo de sinal se faz necessária a utilização de eletrodos invasivos, o que torna sua obtenção praticamente inviável devido aos potenciais riscos e aos incômodos que o processo de inserção de eletrodos oferece.

Os PEs são classificados segundo o tipo (auditivos, visuais, cognitivos e sensoriais) e o tempo de latência. A latência é o tempo medido entre o início do estímulo e o aparecimento do potencial característico (pico do PE). Assim temos PEs de: i) curta latência, por volta dos 50ms; ii) média latência, entre 50 ms e 100 ms; iii) longa latência, com tempos superiores a 100 ms.

Os potenciais evocados auditivos de longa latência estão relacionados aos processos cognitivos: atenção, discriminação e memória, onde as principais ondas (potenciais) são N1, N2, P1, P2 e P3. O P3 ou P300 é o potencial mais tardio, como resposta relacionada às funções mentais de percepção e cognição. Este PE recebe este último nome porque em média, para adultos normais, ele surge cerca de 300 ms após o início do estímulo.

Como o P300 é o potencial mais influenciado pelo uso funcional que o indivíduo faz dos estímulos, ele recebe o nome de potencial cognitivo ou potencial relacionado a eventos (ERP – *Event Related Potential*). Este potencial evocado auditivo pode ser registrado quando um indivíduo é submetido, por exemplo, a uma atividade mental consciente de discriminar dois estímulos acústicos diferentes.

2 POTENCIAL COGNITIVO P300

O P300 se refere a uma resposta cognitiva ou um reflexo de atenção a um estímulo. A resposta P300 reflete a atividade elétrica de milhares de neurônios, mais especificamente, de campos elétricos oriundos de dipolos extracelulares gravados com técnicas de campo distante.

A resposta evocada P300 é mais evidente quando o indivíduo é submetido a experimentos com estímulos apresentados segundo o paradigma *oddball*. A tarefa cognitiva do *oddball*" consiste em oferecer estímulos frequentes ao indivíduo (repetidos de 70% a 90% das vezes) e com menos frequência (de 10% a 30% das vezes) oferecer alternativamente um estimulo diferente, ao qual o indivíduo deve responder quando o reconhecer. Para estímulos auditivos, a resposta é gravada com eletrodos de escalpo nas posições centrais da cabeça. A comunicação entre os dois hemisférios cerebrais através do corpo caloso, referente ao processamento das informações sensoriais, incluindo principalmente o processamento da atenção, tem um papel importante na geração do P300 [9].

O P300 é um indicador da passagem da sensação para a percepção, servindo como um marcador neurofisiológico da atenção. Uma pesquisa interessante é o estudo da relação entre P300 e ouvido absoluto. Registros de EEG de indivíduos portadores de ouvido absoluto não apresentam o P300, ou apresentam um P300 com amplitude significativamente reduzida quando comparados aos de indivíduos sem ouvido absoluto [10]. Este pode ser um indício de que portadores de ouvido absoluto não utilizam a memória de trabalho ou memória operacional (*working memory*) para realizar tarefas cognitivas que requerem a discriminação entre as alturas dos sons.

3 OUVIDO ABSOLUTO E P300

Ouvido absoluto é a habilidade que alguns indivíduos possuem de reconhecer com precisão e sem esforço algum a frequência de um determinado som relacionando-a a um nome (rótulo) específico ("Lá" para 440 Hz; "Dó" para 261 Hz, etc.), assim como de entoar notas musicais sem a necessidade de recorrer a qualquer parâmetro como referência [11]. Esta habilidade parece não se limitar apenas à percepção de sons de instrumentos musicais; relatos informais descrevem alguns portadores de ouvido absoluto capazes de identificar notas nos sons de qualquer fonte como, por exemplo, na voz, nos sons da natureza, nos ruídos de máquinas, entre outros.

A prevalência de portadores de ouvido absoluto na população em geral é muito baixa. Alguns pesquisadores relatam que cerca de apenas uma em cada dez mil pessoas possui esse traço cognitivo [11][12]. Tal raridade provavelmente se deve ao fato de que só é possível avaliar se um indivíduo é portador de ouvido absoluto se ele tiver algum treinamento musical. Indivíduos sem treinamento musical podem possuir uma memória bastante precisa das alturas musicais [13][14], mas não sabem dar nomes às notas.

Não se sabe ao certo qual a origem dessa habilidade. Aparentemente há fatores tanto genéticos como ambientais envolvidos na aquisição do ouvido absoluto [15]. Além disso, há uma forte associação entre idade de início do treinamento musical e a presença do traço cognitivo em questão [16]. Alguns trabalhos mostram que a incidência de portadores de ouvido absoluto é maior entre indivíduos que iniciaram o treinamento musical até os 6 ou 7 anos de idade, sugerindo a existência de um período crítico para aquisição da habilidade [17][18]. Ao mesmo tempo, nem todos os que iniciam cedo o treinamento musical adquirem ouvido absoluto, o que sugere que talvez seja necessária uma predisposição genética para que o traço se manifeste [15].

É necessário ressaltar que músicos que não possuem ouvido absoluto, assim como não-músicos, normalmente processam as notas musicais de maneira relativa. Essa é a maneira usual de se perceber as alturas musicais e é ela que nos permite, por exemplo, aprender uma canção ou identificar uma melodia familiar. Ao ouvir uma sequência de notas musicais o "ouvido relativo" percebe as relações intervalares formadas entre as notas, sem no entanto reconhecer cada uma das notas pelos seus nomes.

Estudos de neuroimagem recentes avaliaram o processamento de alturas musicais através do ouvido absoluto e do ouvido relativo e encontraram diferenças anatômicas e funcionais significativas: portadores de ouvido absoluto apresentam uma pronunciada assimetria cerebral esquerda quando comparados a não-músicos e a músicos sem ouvido absoluto [19][20].

Faremos a captação e avaliação dos potenciais evocados de longa latência (P300) em indivíduos pertencentes a três grupos: músicos portadores de ouvido absoluto, músicos sem esta habilidade e não músicos. A captação e análise dos potenciais evocados serão feitas com auxílio de um

aparelho digital de vinte e um canais de EEG, com capacidade para captar e visualizar também PEs de longa latência usando quatro canais poligráficos. O posicionamento dos eletrodos será direcionado para as regiões corticais de processamento cognitivo, ou seja, as regiões centrais, parietais e frontais.

Os indivíduos participantes desta pesquisa devem ter idade entre 20 e 60 anos. Todos devem ser submetidos a exame audiométrico a fim de verificar a existência de limiar auditivo médio nas frequências de 500 a 4000 Hz melhor ou igual a 25 dB NA (NA – nível de audição).

Utilizaremos como critérios de exclusão os indivíduos que se enquadrem em quaisquer dos seguintes casos: i) possuam perda auditiva condutiva, neurosensorial ou mista unilateral ou bilateral; ii) tenham afecções de orelha externa ou média; iii) apresentem doenças neurológicas ou demências; iv) sejam portadores de diabetes *mellitus*.

Ainda falta definir claramente o protocolo dos experimentos (paradigma *oddball*) bem como dos estímulos auditivos a serem usados. Nosso objetivo inicial é o de verificar as alterações de amplitude e latência do P300 nos três grupos de voluntários, bem como se existem variações dependentes do tipo de estímulo utilizado. Uma de nossas hipóteses é de que deve haver diferenças no P300 de portadores de ouvido absoluto quando submetidos a estímulos auditivos que carreguem algum tipo de informação lingüística (e.g. voz cantada), visto que nesse caso pode haver outros processos cognitivos envolvidos (por exemplo, processamento da linguagem).

Esta hipótese baseia-se nos resultados obtidos por um dos autores deste trabalho, que verificou, com base em um teste auditivo realizado em cerca de 200 portadores de ouvido absoluto de diferentes nacionalidades, que existem diferenças significativas de desempenho na tarefa de identificação de notas musicais em função do timbre do instrumento, evidenciando de forma significativa uma especial dificuldade em identificar notas cantadas, tanto para voz natural como para voz sintetizada [21].

4 CONCLUSÃO

Neste artigo propusemos uma metodologia inicial para o estudo do P300 em indivíduos portadores e não portadores de ouvido absoluto, utilizando um equipamento específico para este tipo de exame, que permite analisar as respostas evocadas quanto a sua amplitude e tempo de latência para cada grupo. Considerando que os sinais sonoros são processados no lobo temporal do córtex cerebral, e que fisiologicamente, o hemisfério esquerdo do cérebro está associado a funções verbais e analíticas e o hemisfério direito à intuição e às artes, um ponto relevante para a continuidade destes estudos seria identificar e deduzir, utilizando exames de EEG com mapeamento (localização de fonte), se música e linguagem são processadas de forma integrada ou independente.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Luck, Steven J. An introduction to the event-related potential technique. 1. ed. Massachusetts: Cambridge University Press Advances Learning, 2005.

[2] Hoth S., Computer-aided hearing threshold determination from cortical auditory evoked potentials, Scand. Audiol., vol. 22, no. 3, pp. 165–177, March 1993.

[3] Hall, J. W. New Handbook of Auditory Evoked Responses, Ed. Pearson, Allyn & Bacon, 2006

[4] Durrant, J. D.; Lovrinic, J. H. Bases of hearing science, 3rd ed., Baltimore, MD: Williams & Wilkins, 1995.

[5] Näätänen, R.; Picton, T. The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an analysis of the component structure, Psychophysiology, vol.24, pp. 375–425, 1987.

[6] Hoppe, U.; Weiss, S.; Stewart, R.W.; Eysholdt, U. An automatic sequential recognition method for cortical auditory evoked potentials, IEEE Transactions on Biomedical Engineering, vol.48, no.2, pp.154-164, 2001.

[7] Lee ,Y.S.,Lueders, H., Dinnner, D. S., Lesser, R. P., Hahn, J., & Klemm, G. Recording of auditory evoked potentials in man using chronic subdural electrodes, Brain,107, 115-131.1984.

[8] Moller, A. R. Latency in the ascending auditory pathway determined using continuous sounds: Comparison between transient and envelope latency. Brain Research, 207, 184-188. 1981.

[9] Polich, J., Hoffman, L. P300 and handedness: On the possible contribution of callosal size to ERPs. Psychophysiology, 35,497-507. 1998.

[10] Klein, M., Coles, M. G. H., and Donchin, E. People with absolute pitch process tones without producing a P300. Science, 233:1306-1309, 1984.

[11] Bachem, A.. Various types of absolute pitch. Journal of the Acoustical Society of America, New York, v. 9, p. 146-151, Dec. 1937.

[12] Takeuchi A.H., Hulse S.H. Absolute pitch. Psychol Bull 113: 345-361, 1993.

[13] Schellenberg E.G., Trehub S.E. Good pitch memory is widespread. Psychol Sci 14: 262-266, 2003.

[14] Smith N.A., Schmuckler, M.A. Dial A440 for absolute pitch: Absolute pitch memory by non-absolute pitch possessors. J Acoust Soc Am 123: EL77-EL84, 2008.

[15] S, Johnston P.A., Service S.K., Gitschier J., Freimer N.B. Absolute pitch: An approach for identification of genetic and nongenetic components. Am J Hum Genet 62: 224-231, 1998.

[16] Gregersen, P.K., Kowalsky, E., Kohn, N., Marvin, E.W. Early childhood music education and predisposition to absolute pitch: Teasing apart genes and environment. Am J Med Genet 98: 280-282, 2000.

[17] Sergeant, D.C.. Experimental investigation of absolute pitch. Journal of Research in Music Education, v. 17, n. 1, p. 135-143, 1969.

[18] Trainor, L. J. Are there critical periods for musical development? Developmental Psychobiology, 46, 262-278, 2005.

[19] Schlaug, G., Jäncke, L., Huang, Y., & Steinmetz, H. In vivo evidence of structural brain asymmetry in musicians. Science, 267, 699-701, 1995.

[20] Wilson, S. J., Lusher, D., Wan C. Y., Dudgeon, P., & Reutens, D. C. The neurocognitive components of pitch processing: Insights from absolute pitch. Cerebral Cortex, 19, 724-732, 2009.

[21] Vanzella, P., Schellenberg, E.G. Absolute Pitch: Effects of Timbre on Note-Naming Ability. PLoS ONE 5(11): e15449,2010.

Sociedade de Engenharia de Áudio
Artigo de Congresso

Apresentado no 9º Congresso de Engenharia de Áudio
15ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2011, 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.

Sistemas de Composição Audiovisuais Duais com Síntese Granular

Fernando Falci de Souza[1,2], Claudiney Carrasco[1,2], Adolfo Maia Jr.[2,3]
[1]Departamento de Música - UNICAMP
Campinas, SP, 13083-970, Brasil
[2]Núcleo Interdisciplinar de Comunicação Sonora - UNICAMP
Campinas, SP, 13083-970, Brasil
[3]Departamento de Matemática Aplicada – UNICAMP
Campinas, SP, 13083-970, Brasil
fernando_ffs@yahoo.com.br, carrasco@iar.unicamp.br, adolfo@nics.unicamp.br

RESUMO

Inspirados no trabalho de Norman McLaren sobre animações gráficas sonorizadas e com a experiência obtida no nosso projeto anterior, temos como objetivo principal deste projeto o desenvolvimento de um conjunto de sistemas áudio visuais duais para composição de imagens gráficas a partir do conteúdo informacional da síntese granular e, inversamente, a composição de fluxos sonoros granulares a partir do conteúdo e seqüenciamento de imagens gráficas. Mais especificamente, queremos estudar e desenvolver métodos de geração e controle de fluxos de sons granulares associados ao conteúdo informacional de vídeos sincronizados com os fluxos sonoros em tempo real.

0 INTRODUÇÃO

Estamos na época áurea da experimentação musical. A comunicação cada vez mais intensa e imediata e a tecnologia digital de ponta, tanto em computação como em instrumentos digitais sofisticados, tem levado a novos paradigmas em tecnologias do som e da imagem. Estas modernas tecnologias requerem naturalmente novos modelos de organização do conteúdo material e informacional, com desafios crescentes à criatividade humana.

Em música contemporânea buscam-se também modelos de organização sonora com suporte tecnológico. Obras de Pierre Schaeffer, Iannis Xenakis, Karlheinz Stockhausen, e John Cage, bem como de outros mais contemporâneos, atestam muitas vezes a preocupação destes compositores com a pesquisa de novas possibilidades de organização do material sonoro. Por exemplo, em Síntese Granular, técnica inventada pelo físico Dennis Gabor e desenvolvida por vários pesquisadores [1-3], a preocupação com a organização do conteúdo sonoro, bem como da organização temporal, parte da simples observação de que é impraticável o controle total do compositor sobre cada grão sonoro possível e disponível. Assim põe-se naturalmente a questão de um controle macro para construção de estruturas sonoras com identidades e características que possam estar dentro do domínio de manipulação macro do compositor. As possibilidades são virtualmente infinitas [4-6]. Levando em conta estas premissas, obtivemos alguns resultados com o desenvolvimento de dois softwares: ***EVOGRAIN*** e ***GranularStreamer***. Estes aplicativos geram fluxos de sons granulares que podem ser usados diretamente em performances em tempo real ou arquivados como material

pré-composicional [7-8]. Eles possuem uma interface gráfica que possibilita um controle macro do músico enquanto produz em tempo real uma série de processos de construção do fluxo sonoro. A concepção, os resultados teóricos e as interfaces gráficas (GUI) destes softwares foram implementadas no Núcleo Interdisciplinar de Comunicação Sonora (NICS) da UNICAMP. O desenvolvimento de controladores gestuais e protocolos de comunicação com as GUIs foi realizado no laboratório IDMIL (*Input Devices for Music Interaction Laboratory*) da *McGill University*, Canadá.

Por outro lado, é notória a tendência atual das artes de se apropriarem das facilidades que a tecnologia digital provê. O termo "Arte Digital" na Internet recupera milhões de páginas. Uma parte significativa da composição musical moderna está baseada em ambientes de performance multimídia, trilhas de filmes, *soundscapes* e outras artes digitais integradas. Além disto, um farto material pode ser encontrado na Internet, mostrando a vitalidade destas atuais formas de expressão artística. De maneira que é importante, tanto do ponto de vista cultural como tecnológico, a pesquisa de novos paradigmas de geração e controle de estruturas audiovisuais integradas. Tendo em mente o acima exposto, neste trabalho queremos contribuir com algumas idéias para o desenvolvimento de novos meios de expressão artística com suporte tecnológico para músicos e artistas digitais que trabalham na interface entre música e imagem.

Inspirados na arte audiovisual de Norman McLaren [9], a qual combina animação gráfica com música derivada ou inspirada no conteúdo informacional das imagens, propomos o desenvolvimento não só de um instrumento musical digital gerando, em tempo real, complexos sons formados com dezenas ou mesmo centenas de estruturas granulares, mas também de um sistema dual totalmente integrado nos domínios do micro-som e do micro seqüenciamento de imagens gráficas, obtendo assim processos automáticos e sincronizados de imagens gráficas com o fluxo sonoro. Tanto quanto sabemos, nossa proposta está baseada em uma idéia original, mas bastante simples, a saber, que tanto a síntese granular como a seqüência de quadros de um vídeo estão no domínio do micro-tempo (dezenas de milésimos de segundo) permitindo assim que o fluxo de sons granulares possa ser gerado e controlado através da evolução temporal de imagens gráficas. Denominamos estes processos de Sonorização Dinâmica de Imagens, ou seja, temos geração e controle de Texturas Sonoras Granulares através do conteúdo informacional e do micro seqüenciamento temporal de imagens de vídeo. Inversamente, imagens gráficas podem ser geradas automaticamente pelo conteúdo informacional do fluxo granular, construindo-se assim animações gráficas em tempo real. Denominamos este processo inverso do anterior de Representação Gráfica Dinâmica de Processos Sonoros. Estas representações têm muito haver com o conceito de Partitura Gráfica já bastante difundida na música contemporânea. Mais recentemente partituras gráficas podem ser lidas dinamicamente como, por exemplo, o trabalho de Rainer Wehinger sobre a peça Artikulation de Ligeti. Estas partituras gráficas podem, em geral, serem lidas, mas não modificadas. Portanto, nosso modelo caracteriza-se por uma representação gráfica dos processos sonoros com interatividade em tempo real e sincronia entre o áudio e o conteúdo visual. Este é o sentido da palavra dual no título deste projeto, os processos sonoros e visuais são interdependentes e complementares em relação ao objetivo final que é a produção de arte audiovisual com suporte de tecnologia digital.

1 METODOLOGIA

1.1 Organização Temporal das Estruturas Sonoras Granulares

Em estilos de composição contemporâneos, uma obra musical é formada por estruturas sonoras que evoluem ou mesmo se transformam ao longo do tempo. Tais transformações podem ser contínuas ou abruptas, incluindo, sem perda de generalidade, como aprendemos com John Cage, o próprio silêncio. A visão desta pesquisa é focada em processos criativos musicais estruturados por métodos formais. Com isto em mente, ao utilizar princípios de organização formal para evolução temporal de estruturas sonoras são necessários três ingredientes:

1) Fonte do material sonoro;
2) Classificação (taxonomia) do material e avaliação da sua complexidade;
3) Regras de organização temporal.

Como uma continuação e aplicação da nossa pesquisa anterior [7-8] as fontes do material sonoro são, principalmente, bancos de sons granulares gerados pelos softwares *EVOGRAIN* e *GranularStreamer*. Para a classificação do material sonoro utilizamos a metodologia de classificação proposta por Curtis Roads e apresentada no seu livro "Microsound" [5]. Alternativamente, dada a diferença de abordagem, pode-se usar também a Espectromorfologia de Dennis Smalley [10]. No caso de síntese granular, uma vez que a possibilidade de controle individual dos grãos é, na maioria das vezes, impraticável, são necessários métodos computacionais para organizar as estruturas granulares efetivamente ao longo do tempo. Sendo este um trabalho em andamento, apresentamos aqui apenas um protótipo desenvolvido cujas aplicações serão baseadas nos softwares desenvolvidos e apresentados em trabalhos anteriores. Assim, nossa proposta, neste caso, é estudar as possíveis correlações temporais entre as várias estruturas sonoras construídas sobre a uma taxonomia. Estas correlações serão construídas a partir de variáveis aleatórias que serão associadas a parâmetros tais como conteúdo espectral, formas de onda e outras características dos grãos sonoros (ou microsons, segundo Roads). De posse destas correlações, podemos proceder ao seqüenciamento local das estruturas, ou seja, do controle micro-textural com Síntese Granular. O seqüenciamento temporal será baseado em Processos Estocásticos e dentre eles os Processos de Markov [11-12].

1.2 Matrizes de Texturas

Formalmente, texturas sonoras granulares podem ser geradas por um conjunto de parâmetros que caracterizam os grãos gerados no fluxo sonoro tais como freqüência, duração, envelope entre outros. Podemos organizar estes parâmetros de fluxos sonoros evoluindo no tempo em matrizes que denominamos Matrizes Texturais. Elas funcionam como modelos de organização sonora tanto da forma (macroestruturas) como do conteúdo (microestruturas formadas por um pequeno conjunto de

sons granulares) e que correspondem ao conteúdo de quadros (frames) de um vídeo. Esta é uma parte bastante importante da nossa pesquisa. Do ponto de vista matemático, esta associação possa ser feita entre dois conjuntos de Matrizes de Texturas, uma descrevendo as texturas de frames e outra descrevendo as texturas sonoras. A idéia é, então, criar um tipo de operador linear [13] que traduz a informação de uma matriz textural na outra e vice-versa. Abaixo apresentamos um modelo simples, mas bem característico da nossa proposta. Definimos inicialmente dois espaços de matrizes: um espaço de matrizes dos possíveis frames ***F*** e um espaço de sons granulares ***G***. Cada elemento (frame) do espaço ***F*** é definido como o conjunto de todos os vetores (ou matrizes colunas 3x1) $\boldsymbol{f}(x,y;\tau) = (R(x,y;\tau), G(x,y;\tau), B(x,y;\tau))$, onde o par (x,y) define a posição do pixel no frame e a tripla (R,G,B) define a sua composição de cores no momento de tempo τ. Vamos supor que (x,y) é um ponto de uma tela retangular usual, isto é, $a \leq x \leq b$ e $c \leq y \leq d$. Num vídeo este conjunto de vetores sofre transformações, deforma-se ao longo do tempo τ que denominamos *tempo de vídeo*. Por outro lado, os sons granulares ***G*** são constituídos, neste exemplo simples, de três senóides com frequências ω_R, ω_G, ω_B com amplitudes A_R A_G e A_B definidas como $A_R = R/(R + G + B)$, $A_G = G/(R + G + B)$ e $A_B = B/(R + G + B)$. Claramente temos uma combinação convexa, pois $A_R + A_G + A_B = 1$. Todas estas funções dependem das variáveis x, y e τ. Então definimos o som granular associado ao frame como:

$$S(x,y;\tau,t) = E(\tau,t)[\, A_R \sin(\omega_R\, t) + A_G \sin(\omega_G\, t) + A_B \sin(\omega_B\, t)];$$

Onde $E(\tau,t)$ é um envelope espectral previamente definido no tempo de vídeo τ e a variável t é o tempo associado ao som e que denominamos *tempo de áudio*. Na definição acima podemos ver que a composição das cores básicas de um pixel controla a composição das freqüências do som a ele associado.

A **matriz de *textura*** associada ao frame no tempo de vídeo τ é definida como o conjunto de todas as triplas amplitudes (A_R, A_G, A_B), isto é

$\boldsymbol{T}(\tau) = \{(A_R, A_G, A_B),\ com\ a \leq x \leq b \text{ e } c \leq y \leq d\}$.

Em outras palavras, cada entrada (i, k) da matriz de textura $T(\tau)$ é associada um vetor de amplitudes das cores básicas formando o pixel na posição (x_i, y_k).

O modelo simples acima associa um som granular a cada pixel de um frame. Para associarmos sons a texturas temos que levar em conta que se usarmos uma simples síntese de sons aditiva para todos os milhares de pixels da textura de um frame vamos obter um ruído muito próximo ao ruído branco. Assim um processamento estatístico da textura da tela associada a cada frame é necessário para obtermos pesos adequados para a síntese sonora dos grãos levando a fluxos sonoros que possam ser controlados em tempo real pelo usuário. Isto está sendo estudado presentemente.

1.3 Composição de Peças Audiovisuais

O uso intensivo da moderna tecnologia digital pode favorecer grandemente a experimentação de modelos formais de organização do material sonoro gerado com Síntese Granular e controlado pelo conteúdo informacional de uma seqüência de quadros de vídeo, levando a uma nova maneira de criação audiovisual. Resumidamente trata-se de uma "sonorização automática" de um vídeo utilizando uma associação adequada entre imagem e som. O modelo descrito acima pode ser implementado em um algoritmo com posterior desenvolvimento de um software o qual deve permitir experimentação com a composição de algumas peças audiovisuais de curta duração (2 a 5 minutos). Claramente isto implica que compositores, músicos, artistas digitais e outras pessoas interessadas em artes audiovisuais podem fazer um uso criativo do software composição e realizar novos experimentos e, com isto, testar a aplicabilidade e versatilidade de nosso modelo. Inversamente queremos estender os resultados obtidos no projeto anterior com o GranularDrawer [8] e compor peças audiovisuais com geração dinâmica de animações gráficas a partir de processos sonoros.

Para integrar estas duas dimensões sensoriais, a audição e a visão, é necessário primeiramente encontrar características inerentes de cada uma (conteúdo informacional) e então definir mapeamentos entre elas, associando características do fluxo sonoro granular com características do fluxo de imagens (vídeo) e vice e versa. Desde os primórdios da pesquisa sobre integrações entre imagem e som é investigada a possibilidade de haver uma correspondência absoluta entre audição e visão, por outro lado se propõe que tal correspondência depende apenas de escolhas estéticas [14].

As associações podem ser realizadas em um *modo interno* onde os parâmetros são obtidos por uma análise das características físicas do som como, por exemplo, amplitude, altura, timbre e duração; e da imagem como brilho, cor, contraste, complexidade entre outras. Associações em *modo externo* são aquelas realizadas entre características consideradas mais semânticas do que físicas. Um exemplo simples seria relacionar a figura de um leão com o som de um rugido. A dificuldade atual em se realizar associações externas é a precariedade de métodos para obtenção destes parâmetros a partir da análise automática do som ou da imagem. Seja qual for o modo da associação, o fator mais importante para que as dimensões sejam percebidas de forma integrada pelos nossos sentidos é a sua sincronicidade [15]. Por sincronicidade entende-se eventos que acontecem no áudio e no vídeo em um mesmo instante de tempo ou com pequeno intervalo de diferença.

2 PROTOTIPAGEM

Conforme indicado por Ciufo [16], por razões de simplicidade, os experimentos iniciais com sistemas de composição podem realizar a análise informacional em uma única direção, isto é, realiza-se a análise do som para então controlar os parâmetros da imagem ou, de maneira oposta, uma seqüência de imagens (frames) é analisada para controlar os parâmetros da síntese sonora. Em nossa primeira implementação de um sistema de composição audiovisual, partimos de informações obtidas a partir da análise de um vídeo para então controlar quatro fluxos independentes (pistas de áudio) de sons granulares sintetizados pelo sistema *GranularStreamer*[8] e emitidos por um sistema estereofônico.

Atualmente estamos analisando apenas duas características físicas da imagem (*modo interno*) utilizando a biblioteca de processamento gráfico Processing 1.0. Pontos (pixels) são amostrados uniformemente de cada quadro de um vídeo e partir deste são obtidos os valores

médios de cor e de brilho da imagem. Para cada quadro, estas características são analisadas de quatro maneiras diferentes. A primeira é proveniente da análise da imagem completa e controla um fluxo sonoro no qual os grão de som são difundidos uniformemente entre o canal esquerdo e o canal direito da saída de áudio dando origem a uma textura de fundo. Na segunda, a imagem é dividida verticalmente em três partes e analisa-se apenas o terço esquerdo da imagem ligando suas características com a síntese sonora onde os grãos são emitidos puramente pelo canal esquerdo. De maneira análoga é feita a análise do terço direito da imagem. Finalmente, o terço central da imagem controla os parâmetros de um fluxo granular centralizado. A Figura 1 descreve a topologia do nosso protótipo.

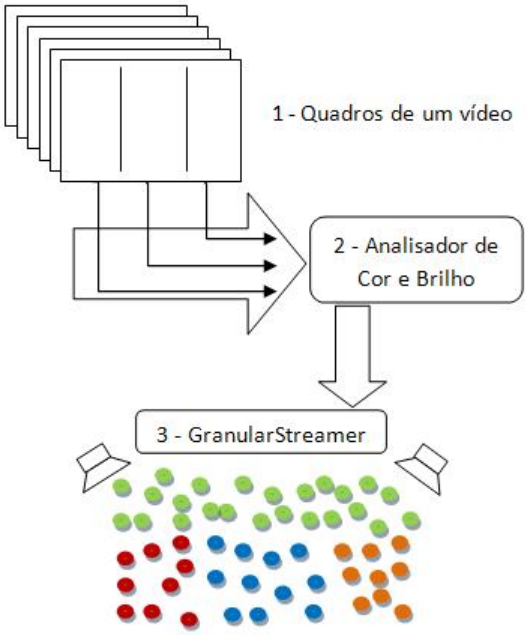


Figura 1 Texturas granulares geradas a partir dos quadros de um vídeo.

Desta forma obtemos uma textura de sons granulares composta por quatro camadas onde o conteúdo informacional do vídeo corresponde temporal e espacialmente ao conteúdo da síntese sonora. Demonstramos também que um mapeamento complexo por ser obtido a partir de diversas ligações simples entre parâmetros físicos da imagem e do som.

3 CONSIDERAÇÕES FINAIS

Neste trabalho propomos uma integração entre áudio e vídeo baseada no fato que as estruturas fundamentais de ambos realizam-se em intervalos de tempo da ordem de algumas dezenas de microssegundos. Assim, dado que os computadores atualmente têm grande capacidade de processamento, podemos construir aplicações que possam carregar o conteúdo informacional de cada um dos domínios no outro.

Podemos antever várias aplicações deste modelo, desde paisagens sonoras dinâmicas que se modificam com a presença de pessoas, assinaturas sonoras em tempo real associadas a imagens já armazenadas, auxílio à composição eletroacústica ou música mista.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Gabor, D., *Acoustical Quanta and the Theory of Hearing*, Nature **159** (4044), pp. 591-594, (1947).

[2] Thomson, P., *Atoms and errors: towards a history and aesthetics of microsound*, Organized Sound, **9** (2), pp. 207-218, (2004).

[3] Truax, B., *Real-time Granular Synthesis with Digital Signal Processing Computer,* Computer Music Journal **12** (2), 14-16, (1988).

[4] Roads, C., *Introduction to Granular Synthesis*, Comp. Mus. Jour., **12** (2), pp. 11-13 (1988).

[5] Roads, C., *Microsound*, MIT Press, Cambridge,MA, (2001).

[6] Roads, C., *Computer Music Tutorial*, MIT Press, Cambridge, MA (1996).

[7] Souza, F. F., Wanderley, M. M., Maia Jr., A., *Evolutive Processes for Granular Synthesis in Extended Gabor Spaces*, Anais do 7º Congresso de Engenharia de Audio, 26 a 28 de Maio de 2009, São Paulo, SP.

[8] Souza, F. F., Wanderley, M. M., Maia Jr., A., *An Input Device to Granular Sound Processes*, Anais do 8º Congresso de Engenharia de Audio, 04 a 06 de Maio de 2010, São Paulo, SP.

[9] Russet, R., *Animated Sound and Beyond*, American Music, *Vol 22(1),* pp. 110-121, 2004.

[10] Smalley, D., *Spectromorphology, Exploring Sound-Shapes*, *Organised Sound, 2(2), pp.107*-126, Cambridge University Press 1997.

[11] Maia Jr. A., Miranda, E.R., *Granular Synthesis of Sounds Through Markov Chains with Fuzzy Control,* Annals of The International Computer Music Conference (ICMC), 2005, Barcelona, Espanha.

[12] Miranda, E. R., Manzolli, J., Maia Jr, A., *Granular Synthesis of Sounds through Fuzzy Set and Markov Chains*, Anais do X SimpósioBrasileiro de Computação Musical (SBCM), Belo Horizonte, 2005. Também apresentado na IX Convenção Nacional da AES - Sociedade de Engenharia de Áudio, 11-13 de abril, São Paulo, SP, 2005.

[13] Halmos, P., *Finite-dimensional Vector Spaces*, Springer, 1999.

[14] Levin, G., Painterly Interfaces for Audiovisual Performance, MIT Master Thesis (2000).

[15] Kapuscinski, J., Basic Theory of Intermedia: Composing with Sounds and Images. Monochord: de music acta, studiaet commentary, Vol.XIX, pg.43-50. Adam Mazek Publications.Torun, 1998.

[16] Ciufu, T.; Real-Time Sound/Image Manipulation and Mapping in a Performance Setting. MAXIS Festival of Sound and Experimental Music Proceedings. Sheffield, UK (2002).

Sociedade de Engenharia de Áudio

Artigo de Congresso

Apresentado no 9º Congresso de Engenharia de Áudio
15ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2011, 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.

Dynamic Linear Downmix from Surround to Stereo Systems

Guilherme Lunhani[1], José Augusto Mannis[1], Jônatas Manzolli[1,2], Adolfo Maia Jr.[2,3],
[1]Departamento de Música, UNICAMP
Campinas, SP, 13083-970, Brazil
[2]Núcleo Interdisciplinar de Comunicação Sonora (NICS), UNICAMP
Campinas, SP, 13083-970, Brazil
[3]Departamento de Matemática Aplicada, IMECC, UNICAMP
Campinas, SP, 13083-970, Brazil
gcravista@gmail.com, jamannis@uol.com.br, jotamanzo@hotmail.com, adolfo@nics.unicamp.br

ABSTRACT

In this work we propose a Linear Downmix from Surround to Stereo Systems, which is capable to keep some psychoacoustics effects present in the Surround System also in the Stereo System from de downmix improvement from 5.1-to two-channel stereophonic system [3,4]. This can be achieved through linear control of signal amplitudes distributed among the surrounding speakers and two low pass filters for the rear speakers. Formally this control can be viewed as a solution of a nonhomogeneous linear equation system of two equations with five parameters. So at least a non-trivial solution always exists. Finally we present an implementation of our model in Pure Data and we discuss the design of the psychoacoustic aspects of our setup with the computer implementation.

0 INTRODUTION: SURROUND 5.1 AND TWO CHANNEL STEREOPHONIC SYSTEMS AND DOWNMIX

Nowadays with the high-speed digital technology together with sophisticate hardware at disposal Sound spatialization turned out a very important aspect of concerts, happenings, shows or simply hearing music at home [1,2]. Surround systems are now available to many people. Nevertheless these systems haven't been easily available to most part of common people. Since the Stereo System is the common one it is interesting to downmix the surround sound material to a stereo version. Solutions for this downmix have been presented for manufactories as well in academic Works[1,2]. These solutions however present a flaw, which we consider important, namely, they loose important qualities of the surround sound patterns. We propose here a solution to this problem with a mathematical model and a computer implementation and we discuss the potential of our methodology.

We present in sections 1 to 3 the structure of surround systems and the spatial distribution of the apparatus. Also we describe the good aspects of surround sound spatialization, mainly the psychoacoustics ones. In section 4 we present our linear model of downmix and calculate some examples. In section 5 we present a computer implementation of the model in Pd. In section 6 we discus the possible experiments with our setup and section 7 we conclude with some observations.

It worth to mention that our approach here is yet a theoretical one based on a linear model. The experimental setup will be performed in near future in order to find the perspectives and limitations of our model.

1 SURROUND 5.1 AND TWO CHANNEL LISTENING ARRANGEMENT

The surround 5.1 (L/C/R/LS/RS) and two channels (L/R) listening present in ITU-R-BS.1116-1 and specified in ITU-R-BS.775-2 is show in Figure 1. The L and R positions in both situations are the same.

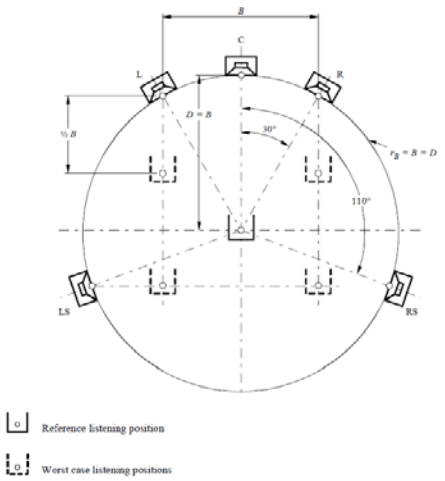


Figure 1 - Multichannel 5.1and 2.0 stereophonic reproduction (ITU R BS.1116-1)

The downmix procedures must pursuit the best result for sound images in L/R reduction including sharpness, deepness, clarity and continuity of sound movements presents in 5.1.

2 REFERENCE DOWNMIX PRODECURE

The reference downmix procedure is done by ITU-R-BS.775-2 and replayed in ITU-R-BS.116-1 in the follow equations formulation:

$$L0 = 1.00\ L + 0.71\ C + 0.71\ Ls \quad (1a)$$

$$R0 = 1.00\ R + 0.71\ C + 0.71\ Rs \quad (1b)$$

3 IMPROVEMENTS

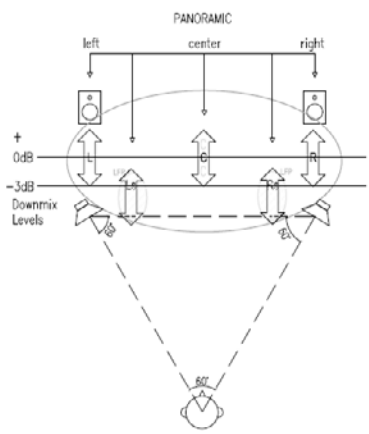


Figure 2 – Sketches to downmix 5.1 to 2.0 improvements [3,4].

One of us [3,4] have got some improvements including in the downmix process a Low Pass Filter and a attenuation corresponding the acoustic shadow effect caused by human head for sounds coming from speakers L_s and R_s, as well to provide a particular position in the downmix space. The values are still in study. They may vary depending on the type of sound material, and the sonorous environmental context of the mixing. This paper works specifically as a support to investigate the behavior of listening in function of parameters of levels and filtering.

4 MATHEMATICAL MODEL

Consider firstly a 5-surround system as in the Fig. 1 above. Let L_0, R_0, C_0, L_{0s}, R_{0s} the maxima power amplitudes chosen by the user for the correspondent speakers of the system. These can be, of course, the maxima amplitudes of the system itself. So, the maximum power that can be used is

$$L_0 + R_0 + C_0 + L_{0s} + R_{0s} = E \quad (2)$$

Defining now the normalized adimensional quantities

$$L = \frac{L_0}{E}, = \frac{R_0}{E},\ C = \frac{C_0}{E}, L_s = \frac{L_{0s}}{E}, R_s = \frac{R_{0s}}{E}, \text{ we have}$$

$$L + R + C + L_s + R_s = 1 \quad (3)$$

We propose here a linear model. The downmix from a Surround System to a Stereo one can be obtained through the linear control of the normalized signals amplitudes, L (left front), R (right front), C (central), L_s (Left Surround) and R_r (Right Surround) of the 5-surround system.
Formally we propose the control to be realized by a non-homogeneous linear system of 2 equations with 5 parameters (5-dimensional parameter space)

$$\begin{aligned} L_{st} &= a_1 L + a_2 R + a_3 C + a_4 L_s + a_5 R_s \\ \mathrm{R_{st}} &= \mathrm{a_2 L + a_1 R + a_3 C + a_5 L_s + a_4 R_s} \end{aligned} \quad (4)$$

where L_{st} and R_{st} are the left and right output amplitudes of the stereo system and $a_k \geq 0$, for $k = 1,2,\ldots,5$. Since here we are using normalized quantities the Principle of Energy Conservation applied to the connected downmix System 5-surround–Stereo give us another constraint equation

$$L_{st} + R_{st} = 1 \quad (5)$$

Now observe that we can have the particular (and extreme) case where $R = C = L_s = R_s = 0$ and $L = 1$. Together with this we also we can choose for the Stereo output $R_{st} = 0$. Then Eqs. (5) and (4) imply $a_1 = 1$. With the same argument for the other coefficients we can prove that the possible range for all the coefficients is $0 \leq a_k \leq 1$. Even with the constraint equations (3) and (5) the linear system (4) has 3 degrees of freedom. So we can write two of them in terms of the other three. Nevertheless some solutions are not allowed. For example, an equal distribution for all five channels, say $a_1 = a_2 = a_3 = a_4 = a_5 = a\ is$ not possible if $a \neq 1/2$ since from Eqs. (3), (4) and (5) we get $a = 1/2$.
In order to get the low frequency from rear speakers the signals sent to them pass through a low band filter, so the amplitudes L_r and R_r are already filtered. Observe that our system has internal symmetries, which are easily seen in the structure of the two above equations, namely, the coefficients of the front speakers as well the rear ones are intertwined. Symmetry implies simplicity at expenses to generality. Nevertheless the above linear system presents a

symmetry, which turns out to be very common to our aural perception.

Now we can proceed in two ways. The first one is, of course, free experimentation by simply by varying the amplitudes *ad libitum* of the surround system and evaluate the result on the stereo output. In this way hearing is the fitness function of the expected results. Nevertheless this approach can take a long time to the user experiment all the effects he/she wish to get from the downmix process. A more effective way is just to fix the expected pattern of the stereo output distribution on time and get the solution set in the parameter space. We can solve the linear system for any output stereo amplitudes L_{st} and R_{st} fixed by user. Since our system is a 2x5 one we have 3 degrees of freedom and so we get a 3-dimensional space of solutions for each L_{st} and R_{st} fixed.

For our downmix setup this means that the user must have in mind the wanted effect along time in terms of the left-right distribution of the stereo signal. In other words the user can prefix the dynamics of the stereo spatialization of the sounds solving the linear system (3) for each pair of outputs (L_{st}, R_{st}) given in time. The solutions above can be used in a computer program to control dynamically the downmix.

5 COMPUTER IMPLEMENTATION

Our computational implementation of the linear model was developed in Pure Data programming language [5][6]. The basic programming structure of PD is called patch and it encapsulates a series of processing routine described by a visual representation. This encapsulation mechanism permits the building of patches that may be reused any number of times. One or more routines (or object boxes) in a Pd patch may be subpatches that are defined in two types: one-off subpatches and abstractions. In either case the subpatch appears as an object box in another patch, called the parent [7]. An "abstraction" can be seen a processing unit that is repeated in several parts of the code. Taken advantage of the modularity of PD, we also implemented the downmix model with a modular architecture. We developed a set of independent abstractions with form downmix.* and filter.*; asteristic indicates specific abstractions in these classes of modules. The pa*rent abstraction (or parent module)* **<downmix.matrix.5.1~>** encapsulate two fundamental PD abstractions: **<downmix.materix.divisor~10>** (see figure 4) and **<filter.lowpass~3000 50>** (see figure 5); The first one is a low-pass filter that simulates the acoustic shadow effect caused by human head, described in section 1. The second abstraction is audio signal divisor that splits the signal in two outputs. After splitting the signal, the DIV~ output is multiplied by the coefficients of linear model (see figure 3). The numbers in object boxes indicate another PD feature, similar to programming languages : the user can provide initialization arguments to object box (in first, each of the parameters describes the splitted signal amplitudes, and in second the *cutoff frequency* and *bandwidth* of filter; a bandwidth of 100 is equivalent to one octave.[1]); in this way we can test different parameters in future test sections to research the best parameters. Others modules only process routing or sum of signals, therefore do not need extensive explanations.

[1] Help file (lowpass-help.pd) of **<lowpass>** object used in **<fiter.lowpass~3000 50>** Guenter Geiger 2000;

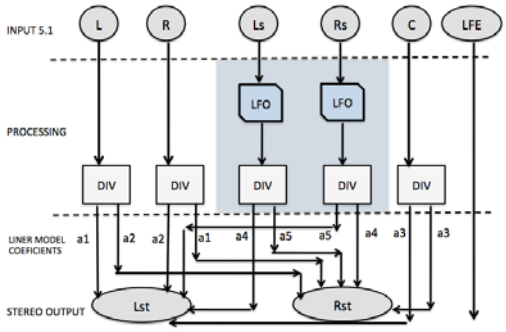


Figure 3 – Linear Downmix implemented in Pure Data (PD). The abstractions implemented in PD are displayed in the processing area for simplicity they are named as LFO and DIV.

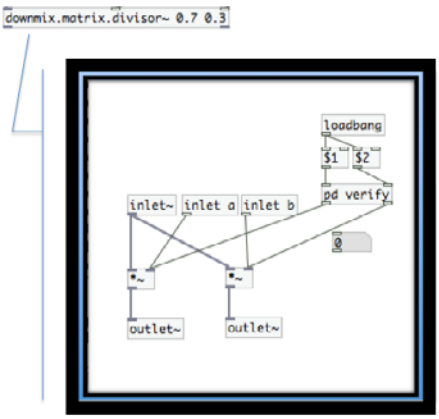


Figure 4 – PD abstraction that splits the signal and multiply by two coefficients. It performs the task of catching a signal and divide into two signals, according to the amplitudes given by arguments [a,b]. Above we have an object box that represents the routine, and the used modules inside the box.

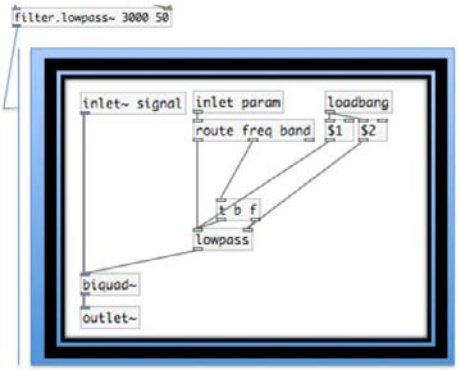


Figure 5 – PD abstraction that performs the LFO of the signal to produce the acoustic shadow. It is a simple 2-pole 2-zero low-pass filter. Our implementation simplifies the use of two PD primitives [biquad~] and [lowpass]. The arguments are the cutoff frequency (freq) and the filter bandwidth (band).

6 DISCUSSION

In our experiments we start from a model of downmix type Left only / Right only (Lo / Ro) with surround channels mixed in phase (in our math model and the figures 3 and 7, they are named L_{st} and R_{st}) respectively. This is similar

type of procedure recommended by ITU-R-1116-2 and BS.775 -1, but with some improvements. We positioned the original channels L_s and R_s closer to the center of the downmix (see figure 2), separating them from a position coincident with the image of the original channels *L* and *R* just over *Lo* and *Ro* (downmix).

Further, we explore the effect of acoustic shadow corresponding to an angle of 120° (filtering attenuating bands from 2 KHz) for general treatment of each surround signal. In our implementation shown above, filtering was applied simultaneously to signals L_s and R_s sent both to the downmix (*Lo* and *Ro*).

Beyond the first experiment described above we will experience, by new features and implementations, a different filtering corresponding to a more realistic situation of the phenomena of acoustic shadow (see Figure 6), i.e., the signal L_s a stronger filtering for signal from channel Ro (because the left ear receives sounds with less high frequencies from Ro signal than Lo signal) and less pronounced filtering for channel Lo (the left ear receives sounds more directly from Ls). In other words the right ear receive the sound from the left surround channel L_s with high frequency attenuation (acoustic shadow) and vice versa for the left ear.

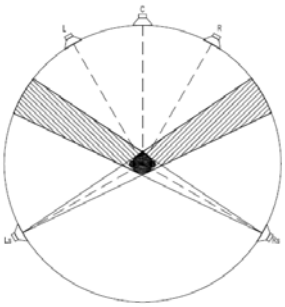


Figure 6 – Acoustic shadow caused by the human head over L_s and R_s sounds

In this way, as shown in figure 7 below, it is possible to make a new arrangement of our linear model to emulate this acoustic shadow on both ears for the stereo apparatus. We adopted also, in addition to filtering process, an attenuation of Ls and Rs, seeking to establish a dynamic plan differentiated frontal sound image to represent the hemisphere behind the listener in your listening space. It is worth to mention that our model, as shown above, allows, theoretically, an infinite number of solutions for each stereo pair *Lo* and *Ro* fixed. So the user has plenty of room (a three dimensional space) to experiment the setup in order to get a specific sound output. Of course, other criteria can be used in order to handle the "navigation" in this parameter space as, for example, simplicity of handling the apparatus.

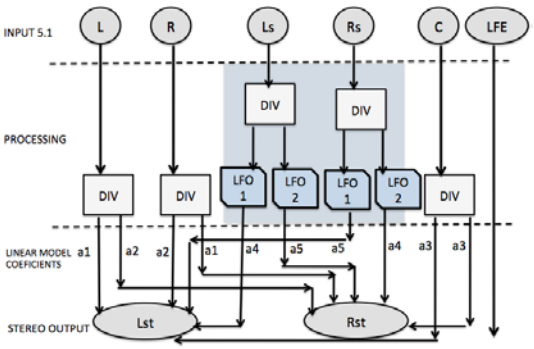


Figure 7 – Diagram of the optional implementation discussed here as improvement of the Linear Downmix.

7 CONCLUSION

We described here a Linear Downmix from Surround to Stereo Systems. Our aim is to keep psychoacoustics effects present in the Surround System also in the Stereo System. We propose a linear control of signal amplitudes distributed among the surrounding speakers and two low pass filters for the rear speakers. We also present a computational implementation of the theoretical linear model in Pure Data. The resources presented here will allow that different sound materials and sound environmental situations are reduced from 5.1 to 2.0 in real time with the impressions and qualitative assessments of listeners in order to correlate downmix parameter settings with the nature of environmental mixing and sound materials involved.

Next step of our work will be to develop psychoacoustic testes to evaluated if the acoustic shadow effect and the distribution of the signal in the stereo field are homogeneous and preserve a good sound balance. Since our model is based on a solution of a nonhomogeneous linear system of two equations with five parameters we think it is interesting to apply an optimization methodology based on an external control as, for example, Genetic Algorithms. The first model will be parameterized with a fixed fitness to drive the evolution. The second model is related to the Interactive Genetic Algorithm (IGA) paradigm and in this case the linear coefficients will obtained by an interactive adjust, including the a dynamical fitness function, taken in to account the interaction with the user during psychoacoustic tests.

BIBLIOGRAPHY

[1] INTERNATIONAL TELECOMMUNICATION UNION. *Multichannel stereophonic sound system with and without accompanying picture*. Recommendation ITU-R BS.775-2. Genève: ITU, 2006. 11p.

[2] INTERNATIONAL TELECOMMUNICATION UNION. *Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems*. Recommendation ITU-R BS.1116-1. Genève: ITU, 1997. 26p.

[3] MANNIS, J. A. Downmix de 5.1 para 2.0 de Criações sonoras em Atelier de composição. Festival de Londrina, PR. Julho de 2007.

[4] MANNIS, J. A. Downmix de 5.1 para 2.0 de Criações sonoras em Atelier de composição para veiculação radifônica e reprodução em equipamentos 2.0. Disciplina MU771 Depto. de Mùsica Unicamp. Campinas, SP, Junho de 2008.

[5] Puckette, M. 1997. "Pure Data." In *Proceedings of the International Computer Music Conference*. San Francisco: International Computer Music Association. 224- 227.

[6] Puckette, M. 2002. "Using Pd as a score language." In*Proceedings of the International Computer Music Conference*. San Francisco: International Computer Music Association. 184-187.

[7] Pd documentation in http://crca.ucsd.edu/~msp/techniques/v0.11/book-html/node67.html, acessed in 20/03/2011

Sociedade de Engenharia de Áudio

Artigo de Congresso

Apresentado no 9º Congresso de Engenharia de Áudio
15ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2011, 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.

Desenvolvimento de um Módulo de Resposta Vocal para a Plataforma Embarcada Arduino

Júlia Cristina Vieira[1] e Fernando Santana Pacheco[1]
[1]Departamento Acadêmico de Eletrônica
Campus Florianópolis
Instituto Federal de Santa Catarina
Florianópolis, SC, 88020-300, Brasil
juliav91@gmail.com, fspacheco@ifsc.edu.br

RESUMO

Este artigo apresenta o desenvolvimento de um módulo de resposta vocal de domínio restrito para a plataforma microcontrolada Arduino. O módulo desenvolvido constitui-se de uma placa eletrônica para leitura do áudio armazenado em um cartão secure digital (SD), conversão D/A e amplificação do sinal de fala; e um programa embarcado, responsável pela seleção dos arquivos e controle da reprodução. Com a escolha de um domínio específico e tomando cuidado durante o processo de segmentação para evitar cortes abruptos, a qualidade das respostas vocalizadas mostra-se adequada.

0 INTRODUÇÃO

Com a disseminação de dispositivos eletrônicos, é de fundamental importância que a comunicação humano-máquina seja fácil e clara [1]. A fala é uma das principais formas de comunicação entre humanos; entretanto, ainda é pouco explorada na interação com sistemas eletrônicos.

A comunicação falada envolve dois sentidos: o ouvir e o falar propriamente dito. Na área de processamento de fala, essas duas vias de comunicação são implementadas por sistemas de resposta vocal (ou seja, quando o dispositivo eletrônico "fala") e de reconhecimento de fala (quando o dispositivo "ouve"). Já existem diversos sistemas *text-to-speech* (TTS) de alta qualidade, baseados em síntese concatenativa, mesmo para a língua portuguesa [2]. Entretanto, como esses sistemas são, em geral, projetados para aplicações em computadores *desktop* ou servidores, não há preocupação em reduzir a ocupação de memória necessária para armazenar os trechos de fala base para a síntese. De forma diversa, sistemas baseados em segmentos fonéticos de curta duração ou em síntese por formantes realizam, com baixo custo computacional, a síntese de qualquer texto, mas com qualidade pobre. Uma solução intermediária é o desenvolvimento de síntese de fala para domínios restritos [3, 4]. Desse modo, considerando a necessidade de incorporar voz a plataformas eletrônicas embarcadas, o objetivo deste trabalho é o desenvolvimento de um módulo de resposta vocal de domínio restrito para a

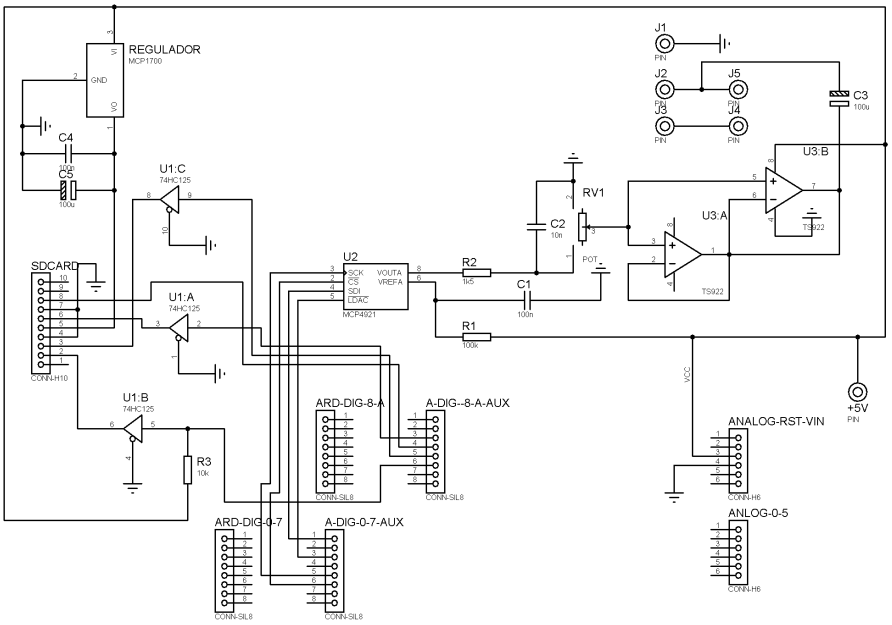


Figura 1 Diagrama esquemático da placa eletrônica desenvolvida

plataforma Arduino. Este artigo apresenta tanto o desenvolvimento do hardware quanto do software embarcado. A aplicação principal vislumbrada refere-se à vocalização de números, direções e informações meteorológicas, constituindo o domínio restrito de síntese.

1 A PLATAFORMA ARDUINO

Em relação à plataforma-alvo, Arduino é uma plataforma microcontrolada de uso geral, em que tanto a especificação do *hardware* quanto o ambiente de desenvolvimento de *software* estão disponíveis gratuitamente [5,6]. Há algumas versões de placas-base, com maior ou menor número de entradas e saídas, mas todas são baseadas nos microcontroladores AVR de 8 bits da família ATmega. Neste trabalho, foi empregada a placa versão Duemilanove, com tamanho de aproximadamente 5x7 cm. Essa placa utiliza o microcontrolador ATmega168, com 16 kB de memória *flash,* e possui 14 portas de entrada e saída digitais, comunicação por interface USB e frequência de trabalho (*clock*) de 16 MHz. À placa-base Arduino, podem ser conectadas placas auxiliares (*shields*, na terminologia dos desenvolvedores), para as mais variadas funções [6]. Quanto ao *software*, uma linguagem própria, também chamada Arduino, foi desenvolvida pelos criadores da plataforma. Essa linguagem é baseada em C/C++. Além da linguagem Arduino, é possível codificar diretamente em C ou *Assembly*.

2 MÓDULO DESENVOLVIDO

O módulo desenvolvido constitui-se de uma placa (*shield*) para leitura do áudio armazenado, conversão D/A e amplificação do sinal de fala e do *firmware*, responsável pela seleção dos arquivos e controle da reprodução. O áudio é armazenado em um cartão *secure digital* (SD), no formato wav, 16 kHz, 16 bits, sem compressão. Justifica-se o uso de um cartão SD pelo baixo custo e facilidade de comunicação com o microcontrolador através da *Serial Peripheral Interface* (SPI).

Devido à dificuldade de aquisição de componentes no mercado brasileiro, optou-se por adaptar um *hardware* já testado, desenvolvido por [7]. Os arquivos de áudio são lidos do cartão (formato digital descrito anteriormente) e convertidos em um sinal analógico, para posterior reprodução, através do conversor D/A Microchip MCP4921. Esse é um conversor digital-analógico com 12 bits de resolução e comunicação SPI, com clock máximo de 20 MHz. Note que os 4 bits menos significativos do sinal de áudio são descartados, mas sem acarretar perceptível perda de qualidade. Embora tenha sido avaliado o uso de um conversor D/A de 16 bits, o custo tornou-se proibitivo, pois não há muitos conversores com interface SPI.

Como o cartão SD opera em 3,3 V e a tensão disponível na placa Arduino é de 5 V, usa-se o regulador de tensão MCP1700 para alimentação do cartão. Ainda em relação ao cartão, são empregados três *buffers* em algumas das linhas, através do uso do CI 74HC125.

Já na saída do sinal de áudio (analógico), é dado um ganho, permitindo que o mesmo seja ouvido diretamente em um fone de ouvido ou um pequeno alto-falante. Emprega-se, aqui, um amplificador operacional, STMicroelectronics TS922, com baixa distorção e capacidade de corrente de saída de até 80 mA. O ganho de tensão do amplificador, que regula o volume do áudio na saída, é controlado por um potenciômetro de 10 kΩ. A conexão ao transdutor sonoro é realizada através de um *jack* padrão P2 (3,5 mm).

O diagrama esquemático da placa eletrônica é mostrado na Figura 1 e a lista de componentes, com os custos de aquisição, é apresentada na Tabela 1.

A placa de circuito impresso foi projetada no software Proteus/Ares. É de face simples, com dimensões 7,1 x 5,6 cm, sendo mostrada na Figura 2. A placa foi confeccionada pelo processo de fresagem, montada e testada nas dependências do Instituto Federal de Santa Catarina.

Tabela 1 Lista de componentes com estimativa de custo

Qtd.	Componente	Preço (R$)
1	Regulador 3,3 V MCP1700	0,85
1	Conversor D/A MCP4921	4,10
1	Amplificador operacional TS922	3,30
1	Conector SD	5,00
1	Jack estéreo 3,5 mm	2,15
1	Buffer 74HC125	0,70
2	Capacitor eletrolítico 100 µF/16 V	0,24
2	Capacitor cerâmico 100 nF/16 V	0,18
1	Capacitor cerâmico 10 nF/16 V	0,09
1	Resistor 1,5 kΩ	0,05
1	Resistor 10 kΩ	0,05
1	Resistor 100 kΩ	0,05
1	Potenciômetro 10 kΩ	1,20
1	Barra de pinos (40) 180º 15 mm	0,69
	TOTAL	18,65

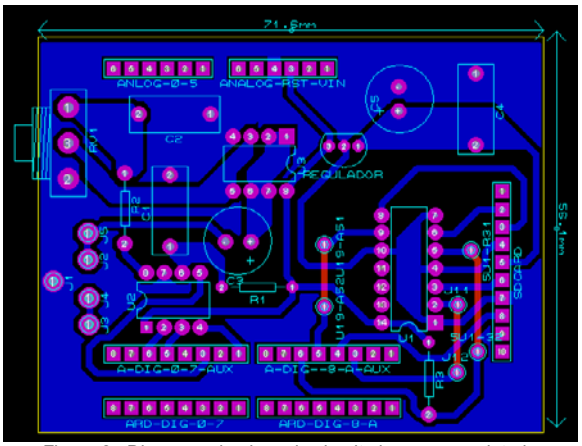


Figura 2 Diagrama da placa de circuito impresso projetada

2.1 Firmware

O *firmware* do módulo foi desenvolvido em linguagem C, com uma pequena parte na linguagem própria do Arduino. Um diagrama de blocos dos processos envolvidos é apresentado na Figura 3. A inicialização consiste na leitura de um arquivo de configuração, gravado no cartão SD, que contém o mapeamento entre o texto e os arquivos wav. Assim, cria-se uma tabela com os nomes dos arquivos wav e os textos correspondentes em cada um deles.

Na etapa de pré-processamento, o texto de entrada é analisado e são expandidas as abreviações e também os números, tendo sido implementada uma função para conversão de números em uma representação por extenso. A seleção da sequência de arquivos a reproduzir constitui-se no núcleo principal do programa, sendo a etapa em que é feita a escolha dos arquivos correspondentes ao texto que deve ser vocalizado. Se não é encontrado áudio referente a um determinado trecho do texto de entrada, há duas opções configuráveis: ou é reproduzido um beep, para marcar que faltou o trecho, ou simplesmente ignora-se aquela parte.

Na última etapa, a partir da lista de arquivos de áudio, é efetuada a leitura dos arquivos do cartão SD e controlada a reprodução. As rotinas de leitura do cartão SD através da SPI são provenientes da biblioteca WaveHC [7].

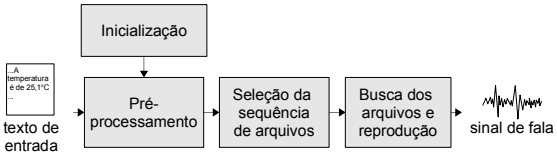


Figura 3 Diagrama em blocos simplificado do firmware desenvolvido

2.2 Gravação e edição do áudio

O áudio utilizado no módulo foi gravado especificamente para essa aplicação. A partir do domínio de frases pretendido (números, meteorologia, direções), foi criado um conjunto de frases (177 no total), lido por uma locutora (autora deste trabalho), de forma natural e pausada, em um estúdio de gravação. Um exemplo de frases utilizadas na gravação é apresentado na Tabela 2. Utilizando o software Praat [8], foi feita a marcação manual dos trechos de interesse, gerando-se o conjunto de arquivos wav e o mapeamento citado anteriormente. Por fim, esses arquivos foram gravados no cartão SD.

Tabela 2 Alguns exemplos de frases gravadas

A corrente é de 0,16 amperes.
A tensão é de um milivolt.
A tensão é de 30,31 volts.
O valor é de 15 pontos.
A temperatura é de 16 graus Celsius.
A capacitância é de 39 picofarads.
A indutância é de 23 mH.
A velocidade é de 28 metros por segundo.
A potência é de 121 watts.
A resistência é de 82 kiloohms.
Umidade relativa do ar em torno de 90%.
A direção do vento é Sul.
O céu encontra-se nublado.
Tempo ensolarado com vento.
São duas horas e 15 minutos.
É meio dia.
Latitude: 27 graus Sul. Longitude: 60 graus Leste.

3 RESULTADOS

Em um primeiro teste, foi verificado o funcionamento do módulo através da leitura de uma sequência de arquivos gravados no SD e posterior reprodução. Mesmo com a limitação do conversor D/A ser de 12 bits, a qualidade do áudio não foi comprometida. Não foi realizado um teste formal de avaliação de qualidade, mas, informalmente, percebe-se que a reprodução é fiel. Com a escolha de um domínio específico e tomando cuidado durante o processo de segmentação para evitar cortes abruptos, a qualidade das respostas vocalizadas mostra-se muito boa. A Figura 4 mostra o resultado da concatenação dos trechos de áudio para gerar uma frase, com a forma de onda, espectrograma e marcação dos trechos selecionados para reprodução.

4 CONCLUSÕES

Este trabalho apresentou o desenvolvimento de um módulo de resposta vocal, de domínio restrito, para uso em aplicações embarcadas com a plataforma Arduino. O

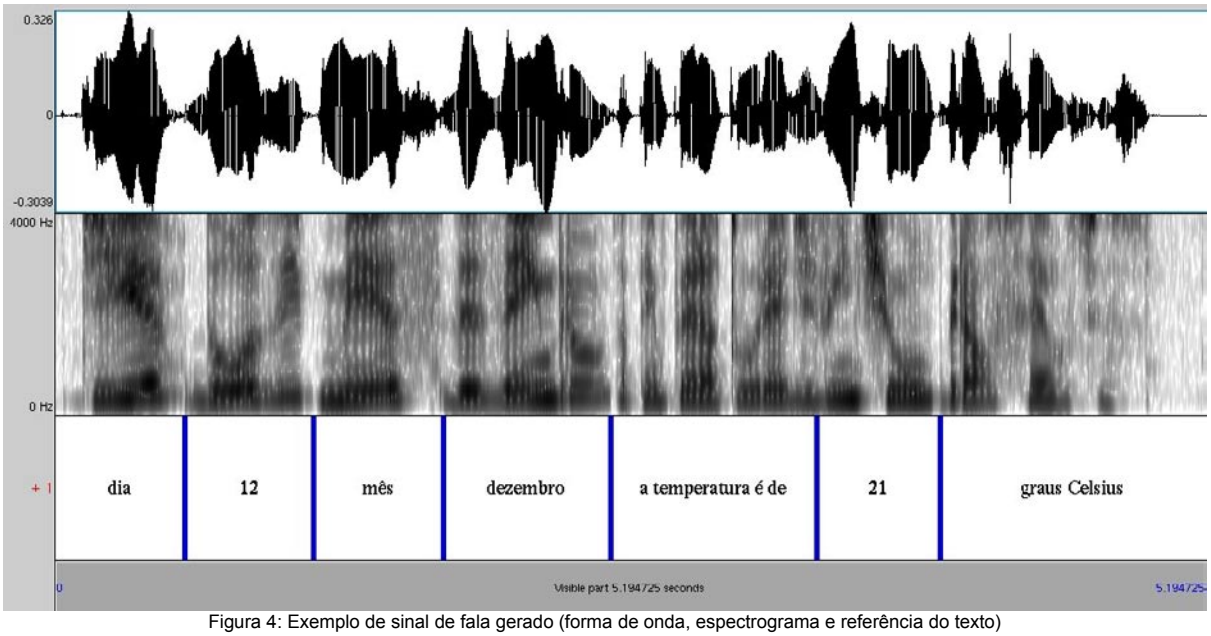


Figura 4: Exemplo de sinal de fala gerado (forma de onda, espectrograma e referência do texto)

módulo, composto por placa eletrônica e o *software* para geração da fala, permite o desenvolvimento de aplicações como relógio falante, auxílio falado para deficientes visuais e informação audível de temperatura e outras variáveis.

AGRADECIMENTOS

Os autores agradecem o apoio financeiro do CNPq, efetivado pela concessão de bolsa de Iniciação Tecnológica à estudante coautora deste trabalho.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] RABINER, L. The Power of Speech. *Science*, v. 301, n. 5639, p. 1494-1495, 12 Sep. 2003.

[2] PACHECO, F. S. *Técnicas de Processamento de Sinais para Alteração de Parâmetros Prosódicos Aplicadas a um Sistema de Conversão TextoFala para Língua Portuguesa Falada no Brasil*. 2001. 100 f. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal de Santa Catarina, Florianópolis. 2001.

[3] BLACK, A. W.; LENZO, K. A. Limited domain synthesis. In: INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING, 2000, Pequim. *Proceedings*... Pequim, 2000. v. 2, p. 411-414.

[4] UTAMA, R. J.; SYRDAL, A. K.; CONKIE, A. Six approaches to limited domain concatenative speech synthesis. In: INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING, 2006, Pittsburgh. *Proceedings*... Pittsburgh, 2006.

[5] BANZI, M. *Getting Started with Arduino*. Sebastopol: O'Reilly, 2008. 128 p.

[6] ARDUINO. Arduino – home page. Disponível em: <http://www.arduino.cc/>. Acesso em: 15 mar. 2010.

[7] WAVE SHIELD. Audio shield for Arduino. Disponível em: <http://www.ladyada.net/make/waveshield/index.html>. Acesso em: 15 mar. 2010.

[8] BOERSMA, P.; WEENINK, D. *Praat*: doing phonetics by computer (Version 5.1.34). Disponível em: <http://www.praat.org>. Acesso em: 10 jun. 2010.

Sociedade de Engenharia de Áudio
Artigo de Congresso
Apresentado no 9º Congresso de Engenharia de Áudio
15ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2011, 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.

Experimentos de Teleconferência Usando o Codec de Voz Speex

Douglas Galetti Ribeiro[1] e Celso Setsuo Kurashima[1]
[1] Universidade Federal do ABC
Santo Andre, SP, 09210-170, Brasil
douggribeiro@gmail.com, celso.kurashima@ufabc.edu.br

RESUMO
Este trabalho apresenta resultados experimentais para teleconferências de voz utilizando-se um codificador de código aberto livre, denominado Speex, desenvolvido por uma comunidade acadêmica sob responsabilidade da Fundação Xiph. Teleconferência de voz é uma parte importante em sistemas de videoconferências ou em outros tipos de aplicações de telecolaboração interativa em tempo real, onde tais aplicações são realizadas em redes de dados ou mesmo pela Internet. Neste contexto, a qualidade decorrente do processo de compressão de voz é um dos requisitos para viabilização do sistema. Visando aplicações avançadas em videoconferência IP, foi investigado o codec de voz Speex e suas peculiaridades, analisando-se sua qualidade objetiva de codificação em relação a outros codecs do gênero, assim como foi realizado um experimento de teleconferência de voz por meio deste codec numa rede local.

0 INTRODUÇÃO

A realização de videoconferência pela Internet vem se tornando uma prática comum entre usuários de computadores, causando impactos positivos na vida das pessoas quanto às suas necessidades de comunicações pessoais. Os sistemas de videoconferência por rede IP ou Internet dependem principalmente de um bom codificador de voz online que ofereça boa compressão da voz capturada sem comprometer sua qualidade [1]. Em sistemas gratuitos, como Skype[1] e MSN Windows Live[2], os codificadores de vídeo são proprietários.

Para tal finalidade, são usados codificadores de voz que realizam diferentes técnicas na compressão de dados. Esses codificadores, no entanto, reduzem o tamanho do arquivo original, mas não é possível que sua qualidade se mantenha a mesma havendo perdas[2]. Tendo por base o codec de voz Speex e seu caráter livre, experimentos foram feitos variando alguns de seus parâmetros como taxa de bits e taxa de amostragem comparando-os e os inserindo em uma teleconferência a fim de estabelecer os valores ótimos de funcionamento do codec.

1 OBJETIVOS

O objetivo deste trabalho é realizar um estudo experimental sobre a qualidade de voz do codec Speex, para investigar a viabilidade de sua utilização em sistemas de videoconferência IP de código aberto com desempenho de boa qualidade e baixo custo. Os experimentos com o codec Speex foram realizados juntamente com outro estudo nosso sobre o codec de vídeo Theora [3], que também vem a ser um codificador de código aberto [4]. Desta forma, uma meta futura é o de implementar experimentalmente um sistema de videoconferência IP através do conjunto Theora e Speex de codificadores de vídeo e voz, respectivamente, visando aplicações acadêmicas em videoconferências. O presente trabalho tem como objetivo

[1] http://www.skype.com/intl/pt/home
[2] http://www.windowslive.com.br

dar suporte a esta meta, pois as comunicações por meio da fala é uma parte importante do sistema.

2 METODOLOGIA E MATERIAIS

Este trabalho foi realizado em diversas etapas para realizar a avaliação do codec Speex, que são as seguintes: (i) levantamento de características dos codificadores de voz; (ii) obtenção de amostras representativas de voz em arquivos adequados; (iii) simulação da codificação e da decodificação dos sinais de voz através de diversos tipos de codec de voz além do Speex; (iv) avaliação comparativa dos arquivos simulados através do método PESQ; (v) realização de teleconferência de voz através de rede local, e sua avaliação informal e subjetiva. Os demais itens desta seção apresentam os detalhes de cada uma destas etapas.

2.1 Levantamento de Características

Inicialmente foram levantados dados característicos de alguns codecs de voz como licença, taxas de amostragem possíveis, atraso algorítmico (*delay*) e tratamentos como supressão de ruído, de eco entre outros, conforme apresentado na Tabela 1.

Tabela 1: tabela comparativa entre os principais codecs de voz [5].

Codec	Frequência de amostragem (kHz)	Taxa de bits (kbps)	Atraso de quadro + *Lookahead* (ms)	Licença
Speex	8, 16, 32	2.15-24.6 (NB) 4-44.2 (WB)	20+10 (NB) 20+14 (WB)	Código Aberto Software Livre
iLBC	8	15.2 ou 13.3	20+5 ou 30+10	Não é Livre
AMR-NB	8	4.75 - 12.2	20+5?	Proprietário
G.722.1	16	(16) 24,32	20+20	Não é Livre
G.729	8	8	10+5	Proprietário
GSM-FR	8	13	20+?	Proprietário
G.723.1	8	5.3 ou 6.3	37.5	Proprietário
G.728	8	16	0.625	Proprietário

Após a busca na literatura foram feitas comparações, para isso foi necessário o uso de um computador com sistema operacional Windows (XP ou mais novo) e algumas ferramentas.

2.2 Obtenção de Amostras

Neste trabalho foram usadas diversas amostras de voz masculina e de voz feminina no formato WAV sem perdas para submissão ao processo de codificação e decodificação de voz. Estas amostras foram obtidas em diferentes taxas de bits e taxas de amostragem, disponibilizadas pela Xiph.org [5], conforme apresentado na Tabela 2. Utilizamos neste experimento apenas as amostras com taxa de amostragem de 8 kHz.

2.3 Simulação da Codificação e Decodificação dos Sinais de Voz

Através da ferramenta *Multimidia UI: Speech and Audio Compression*, disponibilizada por Hwang [6], foi realizada uma simulação da codificação e decodificação destes sinais obtidos (na Tabela 2), para o formato Speex, bem como para os codecs G.723, G.726, G.729, e LPC, que são codificadores conhecidos e padronizados [6], e ainda o codec GSM usado em telefonia móvel.

Tabela 2: arquivos de voz usados nos experimentos de codificação

Item	Arquivo	Gênero	Taxa de Bits (kbps)
1	male.wav	M	128
2	malespeex4.wav	M	4
3	malespeex4vbr.wav	M	4 (VBR)
4	malespeex6.wav	M	6
5	malespeex8.wav	M	8
6	malespeex8vbr.wav	M	8 (VBR)
7	malespeechx11.wav	M	11
8	malespeechx11.wav	M	15
9	female.wav	F	256
10	femalespeex4.wav	F	4
11	femalespeex4vbr.wav	F	4 (VBR)
12	femalespeex6.wav	F	6
13	femalespeex8.wav	F	8
14	femalespeex8vbr.wav	F	8 (VBR)
15	femalespeex11.wav	F	11
16	femalespeex15.wav	F	15

O uso desta ferramenta é fácil e intuitivo, possuindo uma interface simples, e permitindo gravações de voz instantaneamente. A Figura 1 apresenta a interface gráfica de usuário da ferramenta Multimedia UI, onde se pode verificar a possibilidade de escolher o tipo de codec de voz (*Speech Codec*) e sua taxa de bits de operação (*Bit Rate*).

Os sinais passados pelo processo de codificação e decodificação foram salvos em arquivos tipo WAV. Os sinais gerados foram obtidos com mesma taxa de bits do arquivo original.

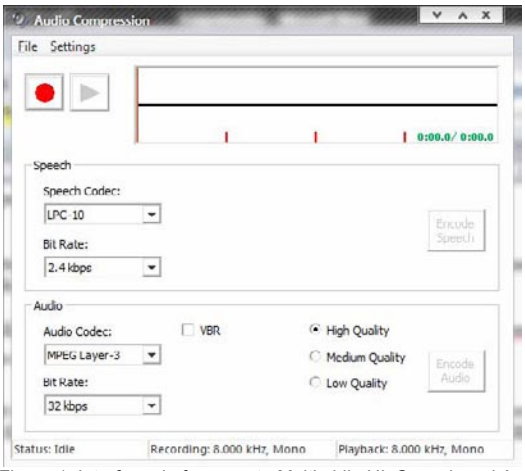


Figura 1: Interface da ferramenta Multimidia UI: Speech and Audio Compression. Com ela é possível mudar diversos parâmetros de qualidade na codificação da voz gravada.

2.4 Avaliação Objetiva

Para fins comparativos, foi usada a ferramenta PESQ (*Percentual Evaluation of Speech Quality*)[7] capaz de avaliar as distorções causadas por codecs diversos, transcodificações, erro de transmissão, perda de pacotes, entrecorte de tempo entre outros. Neste caso, foram avaliadas as distorções obtidas através da codificação dos arquivos de áudio contendo vozes masculinas e femininas.

Foi utilizado um aplicativo do método PESQ que funciona através de comandos de linha (*prompt*) do Windows. É necessário informar através do comando, a taxa de amostragem dos codecs a serem usados na comparação, o nome do arquivo original e sua extensão, o nome do arquivo codificado e sua extensão, conforme mostrado na Figura 2. Além disso, é necessário que tanto os arquivos de voz quanto o aplicativo PESQ estejam na mesma pasta. O resultado é, então, salvo em formato

planilha podendo ser usado para interpretação de dados e tabelas comparativas.

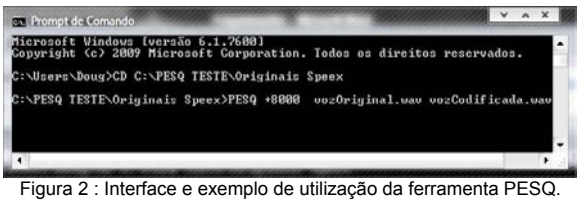


Figura 2 : Interface e exemplo de utilização da ferramenta PESQ.

Em termos comparativos, foi possível gerar gráficos ilustrando a qualidade média (valor médio em PESQ MOS) com relação à taxa de bits. Isso foi feito para os diferentes codecs de voz analisados e as amostras obtidas (Tabela 2).

2.5 Teleconferência Experimental

Uma última etapa do estudo foi a realização de um experimento de teleconferência usando o codec Speex. Isso pôde ser feito através do programa multifuncional conhecido por *VLC Media Player* [8], ou simplesmente VLC. A Figura 3 mostra a tela de configuração do VLC, que permite usar tanto dispositivos de captura de áudio quanto de vídeo e fazer distribuição de feixes de sinal (*streaming*) através dos protocolos UDP ou TCP/IP entre dois ou mais computadores e convertê-los em um formato especificado, como Speex.

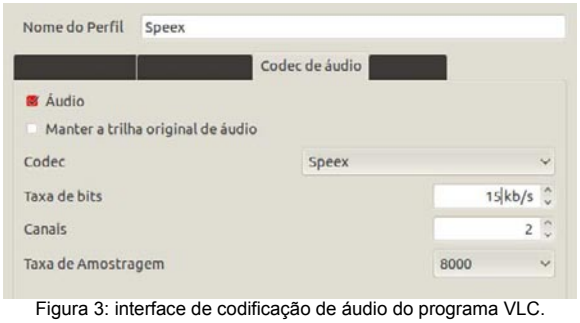


Figura 3: interface de codificação de áudio do programa VLC.

Este experimento foi realizado na rede interna do laboratório da instituição. Também foi realizado através de uma conexão direta de rede entre dois computadores, ligando-os diretamente um ao outro por um cabo cruzado de rede, por onde são enviados e recebidos os sinais de voz através do protocolo UDP. Tentativas foram feitas com redes externas, entretanto, sem sucesso devido a limitação do software VLC cujo *streaming* multimídia através dos protocolos UPD ou RTP são limitadas apenas para redes internas.

3 RESULTADOS

Apresentamos nesta seção os resultados da comparação da qualidade objetiva dos codecs de voz e o experimento de teleconferência com o codec Speex.

3.1 Qualidade através do PESQ

Os resultados valor PESQ-MOS do codec Speex foram comparados com o desempenho dos demais codecs utilizados na avaliação objetiva.

A Figura 4 mostra o gráfico obtido do valor de qualidade objetiva PESQ-MOS em função da taxa de bits utilizado por cada codec de voz na fase de simulação de codificação, para os sinais de voz masculina.

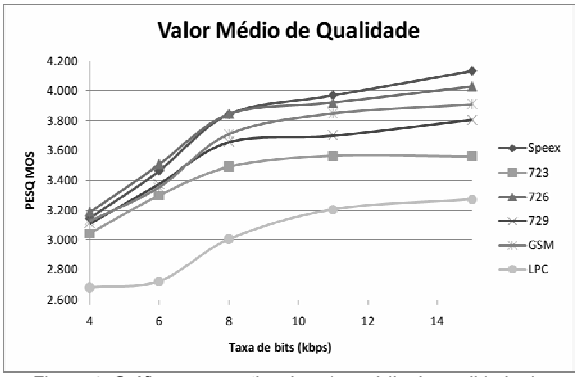


Figura 4: Gráfico comparativo do valor médio de qualidade de diferentes codecs de voz em diferentes taxas de bits com voz masculina.

A Figura 5 mostra o mesmo gráfico, porém para os sinais de voz feminina.

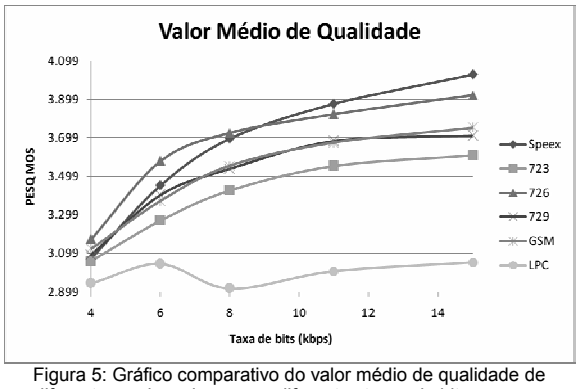


Figura 5: Gráfico comparativo do valor médio de qualidade de diferentes codecs de voz em diferentes taxas de bits com voz feminina.

Através destes gráficos comparativos, é possível perceber que o codec Speex, tanto com vozes femininas quanto masculinas mostrou melhor desempenho de qualidade PESQ do que os demais codecs, principalmente a partir da taxa de bits de 8 kbps. Nota-se que o Speex obteve qualidade superior inclusive ao codec GSM, usado em sistema de telefonia móvel.

Para taxas de bits abaixo de 6 kbps, o codec G.726 mostrou-se superior ao Speex, em termos de qualidade objetiva. Este resultado em termos diferenças relativas de unidades PESQ entre os diversos codificadores são similares aos resultados obtidos na avaliação do codec Speex realizado pela Fundação Xiph.Org [5].

3.2 Teleconferência com Speex

Com o uso do programa VLC foi realizado uma teleconferência simples, que foi útil para verificar a qualidade e desempenho subjetivo do codec Speex ao codificar e decodificar sinais de voz em tempo real numa rede de dados. A Figura 6 mostra o aplicativo VLC executando o envio contínuo de fluxo de voz codificado com o Speex, ou seja, no computador de envio. A Figura 7 mostra o aplicativo VLC executando a recepção do sinal de voz recebido pela conexão de rede, ou seja, no computador cliente. Este experimento foi realizado configurando a rede no modo UDP.

Figura 6: Tela do VLC no lado de envio de sinal de voz em fluxo contínuo (tempo-real) com o codec Speex.

Figura 7: Tela do VLC no lado de recepção (cliente) do sinal de voz com o codec Speex.

Através desta conexão ponto a ponto, o experimento consistiu em enviar sinais de voz de arquivos pré-gravados do lado de envio para o lado cliente. Num segundo experimento, a conexão foi feita utilizando dois microfones e duas instâncias do programa VLC por computador: servidor e cliente. Em cada máquina, uma instância do programa foi responsável pelo envio da voz e a outra, pelo recebimento, tendo cada pessoa falando para enviar sua própria voz pelo computador de envio e recepção do fluxo de voz. Em ambos os casos, no lado cliente uma pessoa ouviu com fones de ouvido o som decodificado pelo VLC cliente.

Em termos de qualidade subjetiva o codec Speex cumpriu o resultado esperado, tornando possível a realização de uma teleconferência com ótima qualidade de envio e recepção, com uma demora menor que dois segundos e provendo uma voz sem distorção e com perdas não significativas. Esse tempo de transmissão entre os dois computadores é decorrente da própria limitação do software. Segundos os testes de codificação, foi possível verificar que o software VLC quando em processo de codificação, gastava em torno de 0,2 segundos para codificar o som recebido e mais 0,2 segundos para decodificá-lo na outra máquina. Somando-se ao o processo todo, cerca de 2 segundos, pode-se prever que o processo de encaminhamento dos pacotes de voz levam em torno de 1,6 segundos entre um computador e outro.

Isto foi possível verificar em caráter informal de teste, pois ambos os computadores se encontravam no mesmo ambiente de laboratório.

Devido limitação do software VLC não foi possível usar outros codecs de voz, para efetuarmos uma comparação subjetiva. O VLC não possui os codecs nativos ou *patches* oficiais que tornem possível a codificação da voz capturada para outros formatos além do Speex.

3.3 Análise dos Resultados

Em relação aos resultados obtidos, pode-se utilizar como parâmetros os de outros autores sobre o mesmo assunto. Em três deles foi utilizado o método subjetivo MOS. Percebe-se que em redes sem fio o codec Speex era inferior aos codecs GSM, G.726 e G.722 em qualidade [9], entretanto, em redes cabeadas atingia valores superiores aos mesmos codecs [10]. Em outro estudo, Speex atingia qualidade superior ao GSM FR, GSM HR e LPC para taxas de bits semelhantes e conseguia bons valores acima de 11 kbps [11]. Em um último estudo, através do chamado Modelo E, Speex atingia valores superiores ao GSM FR e G.726 [12].

Nos experimentos realizados no âmbito deste projeto, percebe-se que os valores obtidos estão de acordo com os resultados experimentais de outros autores. Entretanto, Speex foi superior ao codec G.729 e mais eficaz na compressão de voz em taxas acima de 8 kbps.

4 CONCLUSÕES

O codec de voz Speex mostrou-se apto a realização de uma teleconferência, conseguindo obter desempenho de qualidade objetiva melhor que alguns padrões usados em telefonia e telefonia móvel quando acima de taxas de 8 kbps.

No teste experimental entre dois computadores, a teleconferência apresentou apenas um pequeno atraso de sinal e com qualidade subjetiva de voz muito boa. Não apresentou problemas na recepção ao usuário final, sendo esta avaliação ainda em caráter informal.

Para futuros trabalhos, planeja-se realizar mais testes de teleconferência de voz medindo-se a qualidade de voz subjetiva com maior formalidade, além de submeter a teleconferência de voz através da rede local com TCP/IP e pela Internet. Além disso, planeja-se fazer uma análise do codec Speex em condições de videoconferência, ou seja, trafegando voz e vídeo ao mesmo tempo.

AGRADECIMENTOS

À UFABC pelo apoio e recursos à pesquisa de iniciação cientifica através do programa PDPD.

Ao Prof. Francisco José Fraga da Silva da UFABC, pelo suporte nos experimentos com o método PESQ.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Hannel, K.; *Contribuições ao Processo de Comunicação na Internet Baseado em Videoconferência e Streaming de Áudio e Voz*, 2005. Dissertação - Instituto de Física e Matemática, UFPel.

[2] Haykin, S.; *Sistemas de Comunicação: Analógicos e Digitais*, pag. 25, Bookman 2005, Ed. 4.

[3] Ribeiro, D. G.; Kurashima, C. S.; *Experimentos de Codificação de Video com o Codec Theora.* Boletim Técnico da Faculdade de Tecnologia de São Paulo, v. BT/29, p. 25-25, 2010.

[4] Xiph.org Foundation. Theora Video Compression. Disponível em <http://www.theora.org>. Acesso em 17 mar.2011.

[5] Xiph.org Foundation. Speex Speech Compression. Disponível em <http://www.speex.org>. Acesso em 17 mar.2011.

[6] Hwang J.; *Multimedia Networking – From Theory to Practice*, pag. 467, Cambridge 2009.

[7] PESQ. British Telecommuncations plc. Disponível em <http://www.pesq.org>. Acesso em: 17 mar.2011.

[8] VideoLan Organization. VLC Media Player. Disponível em <http://videolan.org/vlc/>. Acesso em 17 mar.2011.

[9] AIsmail, M.N.; *Best VoIP Codecs Selection for VoIP Conversation over Wireless Carriers Network.* Annals. Computer Science Series, Vol.9, 2011.

[10] Ismail, M.N.; *Implementation of Secure real Time Transport Protocol on VoIP over Wired in Campus.* International Journal of Innovation, Management and Technology, Vol.1 No.3, 2010.

[11] Rämö, A., Toukomaa, H.; *On Comparasing Speech Quality of Various Narrow - And Wideband Speech Codecs*. Proceedings of the Eighth International Symposium on Signal Processing and Its Applications. Sidney, Australia, August 28-31, 2005, pp. 603-606.

[12] Bezerra, T.B. *Avaliação da Qualidade de Codecs de Código Aberto para Transmissão VoIP*. Trabalho de Conclusão de Curso, Engenharia da Computação, Universidade Federal do Amazonas, 2009.

Sociedade de Engenharia de Áudio
Artigo de Congresso

Apresentado no 9º Congresso de Engenharia de Áudio
15ª Convenção Nacional da AES Brasil
17 a 19 de Maio de 2011, 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.

Sistema de Ocultação de Dados em Áudio através de Técnicas de Compactação e Espalhamento Espectral

Luiz Diego Aquino[1], Aurélio Faustino Hoppe[1] e Paulo Roberto Brandt[1]
[1]FURB – Universidade Regional de Blumenau
Blumenau, Santa Catarina, 89012-900, Brasil
{ldaquino, aurelio.hoppe, prbrandt.furb}@gmail.com

RESUMO

Este artigo propõe um sistema robusto de ocultação de dados em áudio baseado na técnica de espalhamento espectral por sequência direta. O sistema insere marcas d'água em frequências aleatórias no áudio de entrada, tornando-a imperceptível para o sistema auditivo humano. Os resultados preliminares apontam a eficiência do sistema quanto à capacidade de inserção e transparência dos dados.

0 INTRODUÇÃO

A crescente necessidade de proteção de direitos autorais em materiais multimídia, assim como a possibilidade de agregação de conteúdo, está cada vez mais dependente de avanços tecnológicos que possibilitem a adição de tais características ao objeto alvo sem degradação de seu teor original. Segundo [1], esta não é uma tarefa trivial, pois a capacidade e a transparência da ocultação das informações estão diretamente relacionadas, sendo inútil aumentar a quantidade de informações que pode-se ocultar, se com isso houver degradação da qualidade do sinal hospedeiro.

Amplamente pesquisadas com propósitos militares, as técnicas de ocultação de informações baseiam-se principalmente em conceitos de criptografia e suas ramificações. Uma delas, a esteganografia, utiliza métodos para inserir mensagens secundárias dentro de mensagens primárias. O termo, originado do alfabeto grego, significa "escrita escondida", pois através desta técnica procura-se camuflar a existência de informações ocultas em mensagens aparentemente inofensivas.

Neste artigo, propomos um sistema de adição e extração de marcas d'água baseado na técnica derivada da comunicação digital conhecida como espalhamento espectral por sequência direta (em inglês, *Direct Sequence Spread Spectrum* ou DSSS). O método desenvolvido processa arquivos no formato *Windows Audio-Visual* (WAV) e utiliza um modelo psicoacústico [1] compatível com áudios de dois canais e frequência de amostragem de 44100Hz e, diferentemente de técnicas de esteganografia como o *Least Significant Bit* (LSB), que distribui a informação que deseja-se ocultar através dos bits menos significativos do arquivo de som [2], o DSSS utiliza o próprio conteúdo sonoro durante o processo.

1 INSERÇÃO DA MARCA D'ÁGUA

O processo de inserção inicia com a transformação do arquivo em amostras *Pulse Code Modulation* (PCM) e o cálculo da quantidade de blocos que serão utilizados para ocultar a informação, chamados de "janelas". Como o tamanho de cada janela foi fixado em 11.14s para manter compatibilidade com o modelo psicoacústico, a quantidade de dados que pode ser ocultada através do sistema é diretamente dependente do tamanho em segundos do áudio.

A mensagem secreta é compactada através de processos heurísticos que substituem preposições de uso comum e combinações de letras. Este método foi aplicado, já que formas de compactação amplamente utilizadas, como o

Código de Huffman [1], não ofereceram resultados satisfatórios para textos de tamanho reduzido.

Em seguida, é realizada a análise espectral do sinal, passando-o do domínio tempo para o domínio da frequência. Para isso, ao invés de utilizar a transformada mais conhecida e mais utilizada em processamento digital, a transformada de Fourier [1], foi empregada a *Modulated Complex Lapped Transform* (MCLT) [3], por possuir propriedades de reconstrução do sinal que a tornam atrativa para o processamento de áudio [4]. Este processo é gradativo no algoritmo, sendo aplicado na iteração principal que analisa os blocos de amostragem do áudio.

A marca d'água, neste caso tratada como um texto a ser mascarado no conteúdo sonoro, é definida como uma sequência de espalhamento espectral w, ou seja, um vetor gerado de forma pseudo-randômica, onde $w \in \{\pm 1\}^N$. Cada elemento w_i é chamado de "chip" [4]. Na prática, cada caractere da mensagem é modulado em chips e escondido no bloco x que está sendo processado, em diversas amplitudes δ. Assim, o sinal marcado gerado y pode ser definido como $y = x + \delta w$.

A Figura 1 ilustra as principais operações do algoritmo de inserção da mensagem no áudio.

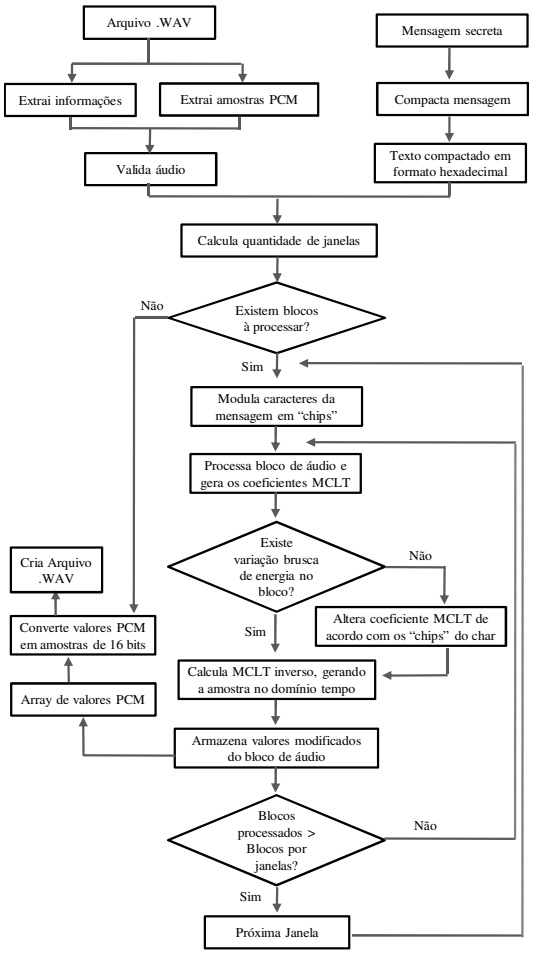


Figura 1 – Algoritmo de inserção da marca d'água

A ocultação dos mesmos chips em diversas faixas de frequência aumenta a segurança do sistema e a taxa de detecção do texto oculto, porém pode ocasionar audibilidade de ruídos indesejáveis, provenientes da adição da marca d'água, em faixas mais silenciosas do áudio. Para evitar esse problema, a amostra processada é submetida a um teste de variação de energia que indica se o ruído adicionado será audível ou não. Se for positivo, os chips não são esteganografados no bloco em análise.

Ao término de cada iteração, o vetor y e a amostra original do bloco são combinados e transformados em um sinal pertencente ao domínio tempo, através da MCLT inversa [3].

Por fim, os valores PCM obtidos pelo sistema são reunidos para a criação de um arquivo WAV com a marca d'água.

2 EXTRAÇÃO DA MARCA D'ÁGUA

Inicialmente, o arquivo WAV de entrada é convertido em um vetor com valores PCM e suas características são validadas para garantir que o áudio seja compatível com o modelo psicoacústico adotado.

Antes do processamento, o sistema estima a quantidade de blocos com caracteres ocultos existente no áudio, para que seja possível definir quando a busca pela mensagem deve ser interrompida. As frequências audíveis do som foram fixadas como o alvo do algoritmo de detecção, pois as porções inaudíveis são mais suscetíveis a ataques de ruído, já que ocupam grande parte do espectro do sinal [4]. É importante mencionar que a extração dos caracteres ocultos é *Blind Detector* [5], por não necessitar do áudio original durante o processo de extração da marca d'água.

Para realizar a extração, são gerados coeficientes de decodificação em diversas escalas de tempo e freqüência. Esta operação tornou-se necessária, pois a utilização do MCLT no processo de inserção espalha os chips de cada caractere por toda a subbanda da janela que mascarará o conteúdo, gerando a redundância de informação que aumenta a segurança da técnica.

A busca pela marca d'água inicia-se pela análise de um bloco do vetor que armazena a informação sonora e em seguida, a geração de seus respectivos coeficientes MCLT. Estes valores são necessários para que sejam realizados testes de audibilidade dos limiares de detecção. Então, com o objetivo de reduzir possíveis ruídos da portadora do sinal que atrapalhariam o procedimento, é aplicado um Filtro de Cepstrum (FC) [4] ao resultado gerado pela combinação dos dados PCM com os coeficientes MCLT. Paralelamente são identificadas as frequências inaudíveis presentes no buffer analisado, para que estas sejam ignoradas durante o procedimento de extração dos caracteres.

Para verificar a existência de algum dado oculto na porção que está sendo processada da janela, são realizados testes de correlação em todas as escalas de tempo e frequência utilizadas durante a inserção da informação. Resultados elevados nestes testes denunciam a presença de marcas d'água [6], conforme pode-se observar na Figura 2.

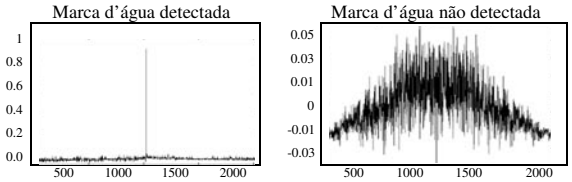


Figura 2 – Resultados da correlação normalizada

Quando a soma das últimas três correlações realizadas supera o valor de um limiar de detecção τ, o sistema

entende que um caractere da marca d'água foi encontrado. A ordem de eventos e processos realizados pelo algoritmo de extração do texto ocultado é apresentada na Figura 3.

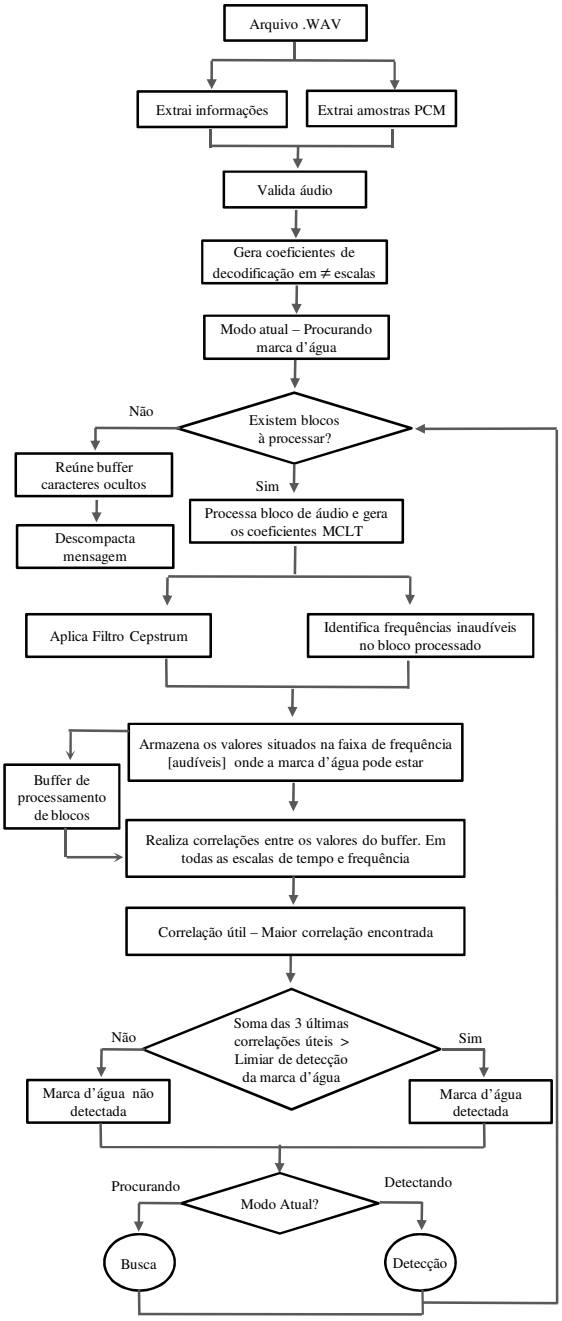


Figura 3 – Algoritmo de extração da marca d'água

A seguir, são detalhadas duas rotinas que se completam, porém não são executadas simultaneamente. Uma delas, chamada de *busca*, onde se gera um coeficiente baseado nos últimos resultados de correlação obtidos e a outra, intitulada de *detecção*, que analisa este coeficiente e armazena o caractere representado pelo seu valor.

Ao final deste procedimento, os caracteres armazenados são reunidos, formando a mensagem oculta. Antes de apresentá-la, o sistema aplica o algoritmo de descompactação compatível com o método utilizado na inserção, para revelar a marca d'água que foi originalmente adicionada ao áudio.

2.1 Busca

Ao encontrar indícios da marca d'água no espectro sonoro, o sistema inicia o monitoramento dos valores das correlações, com o objetivo de encontrar um ponto de estabilização de resultados.

Quando este ponto é alcançado, os últimos dados obtidos são combinados, gerando o coeficiente para detecção do caractere oculto e, consecutivamente iniciando a detecção do texto. A Figura 4 exibe a sequência de passos executados para encontrar a marca d'água espalhada pelo áudio.

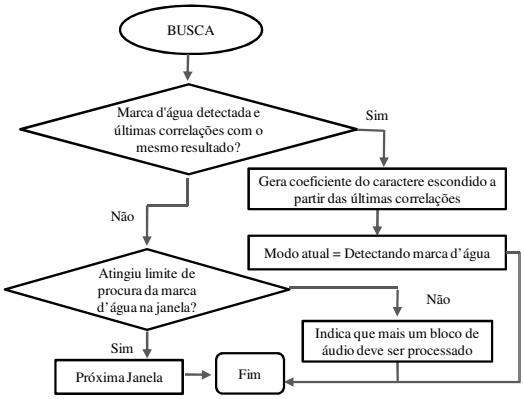


Figura 4 – Rotina de busca da marca d'água

Foi adotado um limite de busca de 5s por janela. Se ele for atingido, a busca é interrompida, pois nenhum texto oculto foi percebido. Quando não são encontrados indícios de mensagens esteganografadas no áudio após a análise de uma parte significativa do bloco, este ponto de parada é importante para que o sistema não permaneça realizando cálculos desnecessários.

2.2 Detecção

A etapa de detecção, apresentada na Figura 5, é responsável por verificar quando ocorre dessincronização nos resultados das correlações. Este cenário indica que o caractere deve ser extraído, através do coeficiente armazenado pelo sistema durante a análise da janela atual, e mantido em um *buffer*.

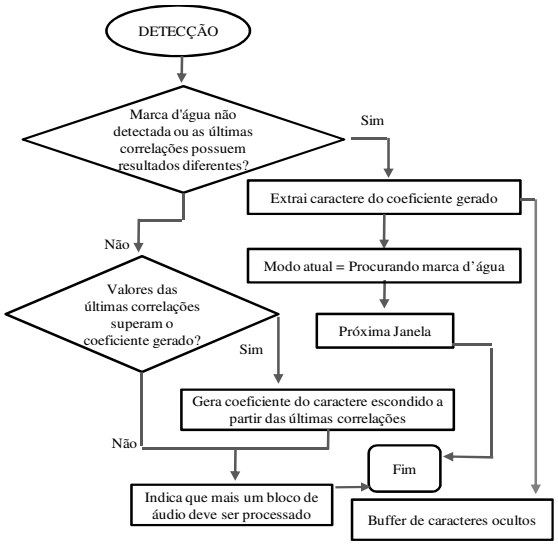


Figura 5 – Rotina de detecção da marca d'água

Enquanto este estado não for atingido (dessincronização), a rotina continua atualizando o coeficiente, caso as últimas correlações apresentem resultados mais favoráveis do que aqueles capturados anteriormente.

3 RESULTADOS

Para validar o sistema, foram realizados dez testes inserindo a *string* "Teste de esteganografia" nas amostras de áudio. Para mensurar a qualidade perceptual utilizou-se a ferramenta PQevalAudio [7]. Nela, variáveis de saída do modelo (em inglês, *Model Output Variables* ou MOVs) são ponderadas e combinadas, gerando uma nota única para o nível de degradação dos sinais de entrada, chamada de *Objective Difference Grade* (ODG). A faixa de valores desta nota varia de –4.0, em casos de distorção muito desagradável, até 0.0, representando ausência total de distorções. Para serem considerados aceitáveis, os resultados devem apresentar ODGs maiores do que –1.5. Valores maiores que –1.0 são considerados muito bons [1].

Os resultados obtidos através desta ferramenta estão listados na Tabela 1.

Tabela 1 Testes de qualidade perceptual e fidelidade da marca d'água extraída

Áudio	*Autor / Música*	*% de fidelidade da marca extraída*	*ODG*
1	Bob Marley / Buffalo Soldier	100%	-1.195
2	Bob Marley / Could You Be Loved	100%	-1.459
3	Bob Marley / Is This Love	100%	-0.870
4	Bob Marley / No Woman No Cry	100%	-0.032
5	Oasis / Wonderwall	100%	-0.13
6	Oasis / Don't Look Back in Anger	100%	-0.617
7	Scorpions / Still Loving You	50%	-0.296
8	Scorpions / Wind of Change	30%	-0.430
9	The Beatles / If I fell	95%	0
10	The Beatles / Strawberry fields	90%	-0.119

Na maioria dos testes, a taxa de acerto na extração da marca d'água foi satisfatória, com níveis de ruído aceitáveis ou muito bons.

Nos casos 7 e 8, as amostras de som tratam-se de gravações ao vivo, com volume médio e pouca presença de percussão. Nestes casos, os resultados foram abaixo do esperado. Isso certamente ocorreu devido às características particulares dessas faixas. Na primeira, o sinal de áudio possui energia muito baixa e na segunda, possui vários períodos de silêncio ao longo do trecho utilizado.

No caso 9, não foi notado nenhum tipo de ruído após o mascaramento do texto, porém a compreensão da mensagem extraída foi comprometida pelo tamanho limitado do áudio, que por conseqüência limita a quantidade de caracteres que o sistema consegue ocultar.

4 CONCLUSÃO

Neste artigo apresentamos um sistema de ocultação de dados em áudio através de técnicas de compactação e espalhamento espectral, onde nosso objetivo era inserir o maior número de informação no sinal hospedeiro, sem que isso gerasse degradação da qualidade perceptual.

Mesmo estando em fase de desenvolvimento, o sistema proposto mostrou-se eficiente quanto à capacidade de inserção de dados e quanto à transparência desses dados. Os resultados obtidos tornam sua aplicação interessante em qualquer nicho de mercado em que a venda de conteúdo de áudio possa ser enriquecida por material adicional, sem a necessidade de utilizar espaço extra de armazenamento.

Como trabalhos futuros, pretende-se realizar novos testes usando métodos mais precisos e com usuário reais para medir a qualidade do áudio. Outras possibilidades de evolução são a implementação de técnicas de criptografia para prover maior segurança às informações ocultadas no sinal de áudio e a criação de um canal de comunicação utilizando ondas sonoras de rádio ou a geração de selos que comprovem a legitimidade das músicas adquiridas via sistemas virtuais.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Schütz, C. A; *Sistema de esteganografia em áudio digital que utiliza técnicas eficientes de inserção de dados*, 97 f. Dissertação (Mestrado em Engenharia Elétrica) - Curso de Pós-Graduação em Engenharia Elétrica, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2009.

[2] Kobuszewski, A., *Protótipo de software para ocultar textos compactados em arquivos de áudio utilizando esteganografia*, 51 f. Trabalho de Conclusão de Curso (Bacharelado em Ciências da Computação) - Centro de Ciências Exatas e Naturais, Universidade Regional de Blumenau, Blumenau, 2004.

[3] Malvar, H. S., *Fast Algorithm for the Modulated Complex Lapped Transform*, IEEE Signal Processing Letters, vol. 10, No. 1, 2003.

[4] Kirovski, D.; Malvar, H. S., *Spread-Spectrum Watermarking of Audio Signals*, *IEEE Trans. Signal Processing*, Vol. 51, No. 4, pp. 1020-1033, 2003.

[5] Alsalami, M. A. T.; Al-Akaidi, M. M., *Digital audio watermarking: survey*, 17th European Simulation Multiconferece, ESM, 2003.

[6] Malik, H.; Khokhar, A.; Ansari, R. *Robust Audio Watermarking using Frequency Selective Spread Spectrum Theory*, Proc. ICASSP'04, 2004.

[7] Kabal, P., A*n Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality*. Montreal, 2002.

Antunes Júnior, I.: 95
Aquino, L. D.: 128

Bertoli, S.: 45
Brandão, E.: 39
Brandt, P.: 128

Cancela, P.: 80
Carrasco, C.: 112
Cordeiro Jr, M.: 74

da Silva, J. V.L.: 45
Dalcastagnê, A. L.: 12
de Oliveira, H. M.: 74
Donoso-Garcia, P.: 61

Fraga, F. J.: 108

Gontijo, A. W..: 12

Hoppe, A. F.: 128

Kurashima, C. S.: 124

Lenzi, A.: 39
Lunhani, G.: 116

Maia Jr., A.: 112, 116
Mannis, J. A.: 45, 102, 116
Manzolli, J.: 88, 116
Mateus, C. A.: 108
Monteiro, A.: 88
Munhoz, R.: 108

Noceti Filho, S.: 12, 53

Oliveira, T. R.: 61
Oliveira, F.: 45

Pacheco, F. S.: 120
Pupo, R.: 45

Ratton, M.: 69
Ribeiro, D. G.: 124
Rocamora, M.: 80

Santos, C.: 45
Souza, F. F.: 112

Vanzella, P.: 108
Vieira, J.: 120

Lista de Revisores / *Reviewers list*

Alexandre Leizor Szczupak
André Dalcastagnê
André Luís Dalcastagnê
Anibal Ferreira
Carlos Ynoguti
Celso Kurashima
Christian Herrera
Diego Haddad
Fabricio Ourique
Fernando Pacheco
Flávio Avila
Francisco Fraga
José Mannis
José Vieira
Julio Lucchi
Leonardo Nunes
Luiz Biscainho
Marcelo Queiroz
Marcio Brandao
Mario Minami
Paulo Antonio Esquef
Phillip Burt
Regis Rossi A. Faria
Sidnei Noceti Filho
Walter Gontijo

AUDIO ENGINEERING SOCIETY

BRAZIL SECTION

Patrocínio - AES Brasil Expo

Roland Systems Group

FZ feel it
www.fzaudio.com.br

AUDIO AES
Audio Engineering Society BRASIL

Realização - Brasil Radio Show

EMBRASEC EDITORA & EVENTOS 15

Promoção

Francal Feiras Desde 1969
Tel (11) 2226 3100
Fax (11) 2226 3200
www.francal.com.br
Compromisso de gerar oportunidades de negócios

Apoio - AES Brasil Expo

CAPES

SBC Sociedade Brasileira de Computação

Montadora Oficial

Estrutural
Tel.: (11) 3934-0015 • Fax: (11) 2831-8357
www.estrutural.com.br • adm@estrutural.com.br

Agência de Viagem

TAM

Local de Realização

anhembi parque
www.anhembi.com.br

Apoio

São Paulo turismo
www.spturis.com