Anais do
III Congresso Brasileiro de Engenharia de Áudio da AES-Brasil
apresentado na
IX Convenção Nacional da AES-Brasil

Centro de Convenções Rebouças
São Paulo, SP
11 a 13 de abril de 2005

Apresentação

Sejam todos bem vindos à IX Convenção Nacional da Sociedade de Engenharia de Áudio – AES Brasil. A AES é uma sociedade mundial, sem fins lucrativos, cujo objetivo é estimular o estudo e o desenvolvimento da engenharia de áudio. A AES está dividida em regiões, e essas regiões estão subdivididas em seções. A Seção Brasil da AES foi criada em 1996 e pertencem a essa seção todos os membros da AES residentes no Brasil. Com mais de 50 anos desde sua fundação, a AES é a única sociedade profissional dedicada exclusivamente à tecnologia de áudio.

A AES serve aos seus membros, à indústria e ao público em geral, estimulando e facilitando os avanços no campo dinâmico da engenharia de áudio. Ela apóia e dissemina novos desenvolvimentos através de reuniões técnicas, de exposições de equipamentos profissionais e do seu conceituado periódico Journal of the Audio Engineering Society.

As Convenções da AES Brasil consistem de exposições e palestras convidadas e, desde o ano de 2003, incluem um Congresso Científico. Este ano, estamos realizando o III Congresso Brasileiro de Engenharia de Áudio. Esse Congresso visa a dar à Convenção também um caráter científico/acadêmico, apresentando trabalhos técnicos relacionados aos avanços da área de engenharia de áudio. A organização técnica do Congresso contou com a participação ativa de três importantes universidades brasileiras (UFRJ, UFSC e USP). Os trabalhos foram selecionados por um corpo de revisores formado basicamente por professores/pesquisadores atuantes na área de engenharia de áudio e suas áreas correlatas.

Esperamos que esta Convenção seja mais um marco na história da Sociedade de Engenharia de Áudio – AES Brasil. Todos nós, Comissão Organizadora, Comissão Técnica, revisores e autores dos artigos técnicos submetidos ao Congresso, fizemos o máximo para alcançar essa meta.

João Américo (Coordenador Geral)

Luiz Wagner Pereira Biscainho (Coordenador Técnico)

Autores

Berger, P. A.
Codificação Perceptiva de Sinais de Voz de Banda Larga.

Dalcastagnê, André L.
Sistema MIDI para Auxílio de Músicos Portadores de Necessidades Especiais.

Faria, Regis R. A.
Simulação de som 3D em um ambiente de realidade virtual imersiva utilizando HRTF.

Fornari, José
Análise da Sonoridade Pianística através do Critério de Adequação da Síntese Evolutiva.

Freitas, Luiz C. de
Amplificador Classe D de 1000W RMS com Realimentação.

Freitas, Luiz C. G. de
Amplificador Classe D de 1000W RMS com Realimentação.

Garcia, Pedro F. D.
Qualidade da Energia em Sistemas de Sonorização - Harmônicos na Rede.

Goldemberg, Ricardo.
Estudo Experimental da Sonoridade "Chalumeau" da Clarineta através de Projeto Fatorial.

Guimarães, Gustavo P.
Qualidade da Energia em Sistemas de Sonorização - Harmônicos na Rede.

Herrera, Christian G.
Qualidade da Energia em Sistemas de Sonorização - Harmônicos na Rede.

Lastch, Vagner L.
Obtenção de Marcas de Pitch em Sinais de Voz para Síntese por concatenação Temporal.

Lima, L. M.
Codificação Perceptiva de Sinais de Voz de Banda Larga.

Livero, Iracele
Análise da Sonoridade Pianística através do Critério de Adequação da Síntese Evolutiva.

Luz, Flávio B. da
Sistema MIDI para Auxílio de Músicos Portadores de Necessidades Especiais.

Maia Jr., Adolfo
Granular Synthesis of Sound through Fuzzyfied Markov Chains.

Manzolli, Jônatas
Análise da Sonoridade Pianística através do Critério de Adequação da Síntese Evolutiva.
Estudo Experimental da Sonoridade "Chalumeau" da Clarineta através de Projeto Fatorial.
Granular Synthesis of Sound through Fuzzyfied Markov Chains.

Miranda, Eduardo R.
Granular Synthesis of Sound through Fuzzyfied Markov Chains.

Moreira Jr., Normandes J.
Amplificador Classe D de 1000W RMS com Realimentação.

Nascimento, F. A. O.
Codificação Perceptiva de Sinais de Voz de Banda Larga.

Netto, Sergio L.
Obtenção de Marcas de Pitch em Sinais de Voz para Síntese por concatenação Temporal.

Noceti Filho, Sidnei
Sistema MIDI para Auxílio de Músicos Portadores de Necessidades Especiais.

Oliveira, Luiz C. de
Estudo Experimental da Sonoridade "Chalumeau" da Clarineta através de Projeto Fatorial.

Silva, Fábio V. R. da.
Amplificador Classe D de 1000W RMS com Realimentação.

Thomaz, Leandro F.
Simulação de som 3D em um ambiente de realidade virtual imersiva utilizando HRTF.

Vieira Jr., João B.
Amplificador Classe D de 1000W RMS com Realimentação.

Zuffo, João A.
Simulação de som 3D em um ambiente de realidade virtual imersiva utilizando HRTF.

Zuffo, Marcelo K.
Simulação de som 3D em um ambiente de realidade virtual imersiva utilizando HRTF.

AUDIO ES

Trabalhos

Amplificador Classe D de 1000W RMS com Realimentação.
Fábio Vincenzi Romualdo da Silva; Luiz Carlos Gomes de Freitas; Normandes José Moreira Júnior; João Batista Vieira Júnior e Luiz Carlos de Freitas.

Análise da Sonoridade Pianística através do Critério de Adequação da Síntese Evolutiva.
José Fornari; Iracele Livero e Jônatas Manzolli.

Codificação Perceptiva de Sinais de Voz de Banda Larga.
P. A. Berger; F. A. O. Nascimento e L. M. Lima.

Estudo Experimental da Sonoridade "Chalumeau" da Clarineta Através de Projeto Fatorial.
Luis Carlos de Oliveira; Ricardo Goldemberg e Jônatas Manzolli.

Granular Synthesis of Sound through Fuzzyfied Markov Chains.
Eduardo Reck Miranda; Jonatas Manzolli e Adolfo Maia Jr.

Obtenção de Marcas de Pitch em Sinais de Voz para Síntese por concatenação Temporal.
Vagner L. Lastch e Sergio Lima Netto.

Qualidade da Energia em Sistemas de Sonorização - Harmônicos na Rede.
Christian Gonçalves Herrera; Pedro F. Donoso-Garcia e Gustavo Paulinelli Guimarães

Simulação de som 3D em um ambiente de realidade virtual imersiva utilizando HRTF.
Leandro Ferrari Thomaz, Marcelo K. Zuffo, João Antonio Zuffo, Regis Rossi A. Faria.

Sistema MIDI para Auxílio de Músicos Portadores de Necessidades Especiais.
Flávio Bressan da Luz; Sidnei Noceti Filho e André Luís Dalcastagnê

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na IX Convenção Nacional
11 – 13 de Abril de 2005, São Paulo, SP

ESTUDO EXPERIMENTAL DA SONORIDADE "CHALUMEAU" DA CLARINETA ATRAVÉS DE PROJETO FATORIAL

Luís Carlos de Oliveira, Ricardo Goldemberg, Jônatas Manzolli
Instituto de Artes (IA) e Núcleo Interdisciplinar de Comunicação Sonora (NICS)
Universidade Estadual de Campinas (UNICAMP)
CEP: 13083-970, Campinas, SP, Brasil
luis@nics.unicamp.br, rgoldem@iar.unicamp.br, jonatas@nics.unicamp.br

RESUMO

A sonoridade *chalumeau* da clarineta é investigada empiricamente através de um método de otimização de experimentos conhecido como *Projeto Fatorial de Experimentos*. Como resultado deste procedimento determinamos as variáveis que efetivamente interferem na sonoridade e descobrimos se existe interação entre elas. Obtemos também modelos lineares que descrevem a influência das variáveis na sonoridade na região estudada.

1. INTRODUÇÃO

O processo sistemático de investigação científica a respeito dos instrumentos de sopro teve seu início no final do século XIX. Desde então o timbre desses instrumentos é uma preocupação constante. Dentre os autores representativos, destacam-se H. L. F. HELMHOLTZ [1], J. RAYLEIGH, H. BOUASSE, J. BACKUS [2, 3, 4, 5, 6] e mais recentemente, H. BENADE [7, 8, 9, 10, 11].

Ainda que de maneira limitada, as contribuições desses e outros investigadores permitem afirmar que, em nossos dias, existe um corpo de conhecimento teórico e experimental que possibilita descrever e simular razoavelmente o comportamento dos instrumentos musicais de sopro.

Entretanto, devido à complexidade do problema em estudo, várias simplificações são impostas. Deste modo, os resultados obtidos divergem consideravelmente das condições reais, tanto do ponto de vista teórico quanto do experimental.

Com relação aos trabalhos empíricos, notamos um elevado grau de preocupação com o sistema oscilador composto pelo conjunto formado pela boquilha e palheta.

Nossa proposta consiste primeiro, em eliminar a variável subjetiva do músico, segundo, trabalhar simultaneamente com um conjunto maior de variáveis havendo a possibilidade de estabelecer se há interação entre elas.

Para tal objetivo tomaremos como ponto de partida uma montagem experimental instalada no estúdio do NICS (Núcleo Interdisciplinar de Comunicação Sonora). Em seguida, vamos analisar o efeito das variáveis na sonoridade *chalumeau* (região grave) da clarineta.

Para otimizar a análise será utilizado um projeto de experimentos que procura obter o máximo de informação com um mínimo de esforço operacional.

Ao utilizarmos este projeto de experimentos obtemos também um modelo linear que descreve o efeito das variáveis na sonoridade na região proposta.

2. APARATO EXPERIMENTAL

Neste trabalho experimental tivemos a preocupação de não utilizar um músico para a coleta de dados por considerarmos que sua experiência musical poderia interferir nos resultados. Montamos, então, no estúdio do NICS um aparato que consiste basicamente de cinco unidades: 1) compressor cujo modelo é utilizado em clínica dentária; 2) tanque "pulmão" que simula o reservatório de ar no corpo humano; 3) unidade de contato com a palheta, que daqui por diante denominaremos por "mordedura"; 4) unidade formada pela clarineta e 5) unidade de captação de dados. Com exceção da unidade 5, as demais podem ser visualizadas nas figuras 1 a 3. Instrumentos de medida como um rotâmetro (medidor da vazão volumétrica de ar que passa pela clarineta) e dois manômetros (para medir a pressão na entrada do tanque e no interior do tanque) complementam o sistema.

O compressor tem um motor de 0,75 kW, monofásico aberto, com 2 pistões, um único estágio de compressão e

um reservatório de 0,03 m^3. O tanque pulmão é de material acrílico, com parede de 0,011 m de espessura, 0,300 m de comprimento por 0,170 m de largura e 0,178 m de altura. O seu tampo possui ainda a característica de ser deslizante.

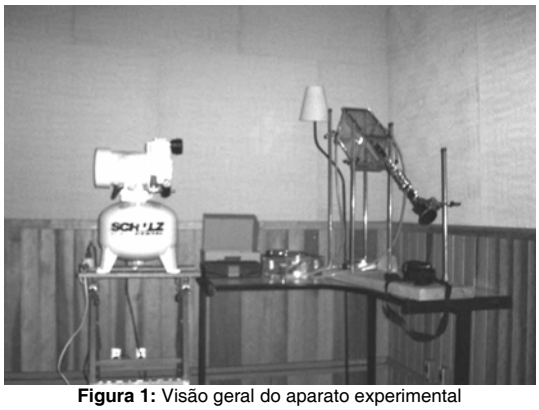

Figura 1: Visão geral do aparato experimental

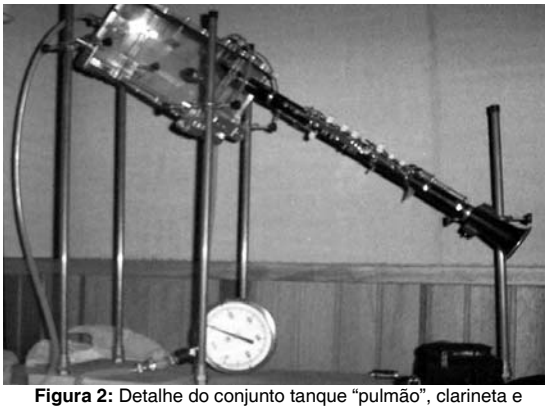

Figura 2: Detalhe do conjunto tanque "pulmão", clarineta e "mordedura".

Figura 3: Detalhe da unidade de contato com a palheta: "mordedura".

O tanque está conectado ao compressor, em uma das faces, por um tubo flexível de 0,120 m de diâmetro através de um sistema de "engate rápido", veja figura 2.

A unidade de captação de dados consiste em um microfone colocado próximo à clarineta e conectado a um computador onde está instalado o software SOUND FORGE 4.5.

O microfone é unidirecional dinâmico (cardióide) e tem resposta de freqüência de 50 a 15000 Hz. A taxa de amostragem utilizada foi de 44 kHz e a configuração utilizada foi de 16 bits. A placa de som tem entradas analógicas de 18 bits (-10 dBV, conexões RCA desbalanceadas), relação sinal/ruído (A/D/A): 88 dB não ponderada e 90 dB "A" ponderada e limite de banda de 22 a 22000 Hz.

A clarineta é um instrumento cilíndrico com uma única palheta em sua boquilha. É o membro de uma família de 17 instrumentos similares os quais, essencialmente, diferenciam basicamente em tamanho [12]. O representante líder desta família é a usual clarineta afinada em Si bemol[1] e será utilizada neste projeto. Veja figura 2.

A mordedura consiste de um parafuso que se desloca na direção ortogonal à clarineta e na sua base está fixada uma pequena borracha que efetivamente faz o contato com a palheta. A borracha terá sua área variável para efeito de estudo. Veja figura 3.

3. PROCEDIMENTO EXPERIMENTAL

A seguir, será descrito o procedimento utilizado para a realização dos experimentos.

Primeiro enche-se o compressor. As chaves da clarineta são fechadas com pequenas borrachas que podem alterar a nota desejada. No tanque pulmão o contato com a palheta é mantido fechado, impossibilitando a passagem de ar através da palheta.

Abre-se lentamente a válvula de saída do compressor de modo que os manômetros indiquem o aumento gradativo da pressão. Quando os manômetros indicarem a pressão de fundo de escala (9808 Pa) abre-se muito lentamente o contato com a palheta de modo que uma pequena quantidade de ar é injetada no interior da clarineta. As pressões indicadas pelos manômetros começam a diminuir como conseqüência.

Deste momento em diante a clarineta está em estado de emissão sonora. Uma vez emitido o som a gravação é acionada no momento que se percebe que sua intensidade é máxima e constante. A gravação é efetuada por volta de três minutos. Desta gravação apenas 15 segundos são selecionados para a análise.

As pressões na entrada da boquilha (interna) e externa são anotadas. A vazão volumétrica de ar correspondente também é anotada.

Depois de uma bateria de testes, estabelecemos que as variáveis que poderiam estabelecer certa influência na sonoridade da clarineta foram: i) volume do tanque pulmão; ii) dureza da palheta; iii) posição de contato da mordedura na palheta; iv) ângulo de abertura da boquilha; v) Tipo de mordedura (área de contato com a palheta) vi) quantidade de material absorvente sonoro (estopa) no tanque pulmão.

Uma vez que o volume do tanque foi variado colocando-se uma placa de isopor em seu interior para separar uma região de circulação de ar e outra isenta de circulação, pensamos em colocar um material dentro da caixa para verificar se teria alguma influência. Daí a justificativa da sexta variável.

Nesta concepção experimental, a vazão volumétrica de ar e a pressão na entrada da boquilha não são variáveis independentes. Observamos também que a produção sonora existe apenas para uma faixa de valores de pressão. Uma visão esquemática do aparato experimental está apresentada na figura 4.

Para o estudo destas variáveis utilizamos um método estatístico conhecido por *Projeto Fatorial de Experimentos*. Ele possibilita determinar quais destas variáveis efetivamente têm influência no estudo e ainda se existe interação entre elas. Esta metodologia utiliza

[1] A clarineta afinada em Si bemol significa que suas notas soam um tom abaixo daquelas descritas pelo piano.

pequenas perturbações das variáveis independentes e reduzido número de experimentos além de fornecer um modelo linear nas variáveis envolvidas.

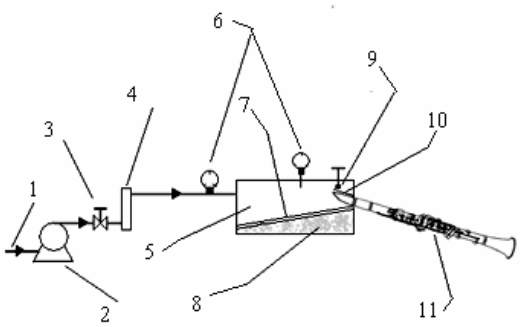

Figura 4: Visão esquemática do aparato experimental: 1) Entrada de ar; 2) Compressor; 3) Válvula; 4) Rotâmetro; 5) Volume Vazio; 6) Manômetros; 7) Placa de isopor; 8) Material para absorção sonora (estopa); 9) Mordedura; 10) Palheta; 11) Clarineta

A figura 5 é esclarecedora sobre a variável correspondente à posição da mordedura na palheta.

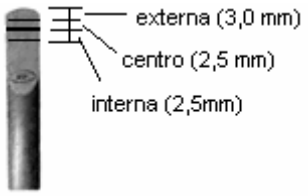

Figura 5: Posição da mordedura na palheta

3.1 Projeto Fatorial de Experimentos

Preliminarmente, *fator* é sinônimo de *variável*. Neste projeto os fatores variam entre dois níveis (**–1** e **+1**), além de um ponto central (**0**). Grosseiramente, um projeto fatorial é uma expansão por Taylor, no nosso caso linear, a partir do ponto central nos diversos fatores (variáveis). Como respostas, (**Y**), analisaremos as intensidades da nota fundamental emitida ($\mathbf{D_3}$) e seus harmônicos.

A elaboração de um Projeto Fatorial de Experimentos está minuciosamente detalhada em BOX [13]. Aqui mencionamos apenas que foi utilizado um **Projeto Fatorial Fracionado** com **Resolução III**. Ser fracionado significa que realizaremos parte (1/8) de um Projeto Fatorial Completo, que no nosso caso para 6 fatores variando em dois níveis seriam 2^6=64 experimentos. Então, realizaremos (1/8)*64=2^{6-3}=8 experimentos. A resolução indica que ao realizarmos parte do projeto completo nossa "perda" de informação está em fundir o efeito de um fator principal (I) com interações entre dois fatores (II), I+II=III.

Na tabela 1 estão descritos os níveis dos fatores utilizados neste experimento. A boquilha **A** tem a menor abertura enquanto a **C**, a maior.

Neste projeto em particular, o padrão de fusão (*confounding patterns*) é **4=12**, **5=13** e **6=23**. Isto significa que o efeito do fator **4** será confundido com a interação **12** (entre os fatores **1** e **2**), o efeito do fator **5** será confundido com a interação **13** (entre os fatores **1** e **3**), etc.

Na região grave da clarineta, também denominada *chalumeau*, escolhemos a nota com a clarineta totalmente fechada, correspondendo à nota $\mathbf{D_3}$. Esta nota refere-se à do piano, portanto para a clarineta corresponde à nota $\mathbf{E_3}$.

No ponto central serão realizados mais três experimentos que terão a finalidade de determinar o erro experimental. Teremos, então, apenas 11 experimentos. Os experimentos foram sorteados para determinar a ordem de execução. A tabela 2 resume o que foi mencionado anteriormente.

FATORES	-1	0	+1
1)Volume Vazio do Tanque Pulmão(%)	60	65	70
2)Dureza da Palheta (Nº)	2	2,5	3
3)Posição da Mordedura na Palheta	Interna	Centro	Externa
4)Boquilha	A	B	C
5)Área de Contato com palheta (m^2)	$3x10^{-5}$	$1x10^{-4}$	$1,4x10^{-4}$
6)Quantidade de Estopa (kg)	0	0,015	0,030

Tabela 1: Níveis dos fatores envolvidos na experiência.

A título de exemplo, a tabela 2 indica que o primeiro ensaio realizado foi o ensaio nº 5. Sendo seguido pelo ensaio nº 7, nº 6 e assim sucessivamente. No ensaio nº 5 o fator **1** apresenta nível **–1**, e consultando a tabela 1 sabemos que o volume vazio do tanque pulmão corresponde a 60% do volume total. Para o fator **2** o nível também é **–1** e a tabela 1 indica que a palheta utilizada foi de nº 2. Analogamente, a posição da mordedura na palheta é a mais externa; a boquilha utilizada foi a **C**; a área de contato da mordedura com a palheta foi de 0,3 cm^2 e não foi utilizada estopa para absorção sonora.

Na análise espectral feita através do SOUND FORGE 4.5 foram testados vários "janelamentos" e o de Blackmann-Harris mostrou ser o mais apropriado. Em todas as análises os dados foram normalizados em –3,0 dB.

A tabela 3 apresenta os valores de pressão e vazão utilizadas para os diversos ensaios.

ENSAIO Nº	FATORES						Ordem
	1	2	3	4	5	6	Y_G
1	-1	-1	-1	+1	+1	+1	7
2	+1	-1	-1	-1	-1	+1	8
3	-1	+1	-1	-1	+1	-1	11
4	+1	+1	-1	+1	-1	-1	4
5	-1	-1	+1	+1	-1	-1	1
6	+1	-1	+1	-1	+1	-1	3
7	-1	+1	+1	-1	-1	+1	2
8	+1	+1	+1	+1	+1	+1	10
9	0	0	0	0	0	0	5
10	0	0	0	0	0	0	9
11	0	0	0	0	0	0	6

Tabela 2: Níveis dos fatores envolvidos na experiência.

Na tabela 4 estão indicados, em **dB**, os valores das intensidades da nota fundamental ($\mathbf{D_3}$) e de seus harmônicos, indicados pela letra **H** e o respectivo índice do segundo ao décimo segundo harmônico.

Na análise espectral feita através do SOUND FORGE 4.5 foram testados vários "janelamentos" e o de Blackmann-Harris mostrou ser o mais adequado pois apresentou maior nitidez de visualização dos picos dos

componentes harmônicos na análise espectral. Em todas as análises os dados foram normalizados em –3,0 dB.

ENSAIO Nº	Região Grave		
	P_{ext} (Pa)	P_{int} (Pa)	Vazão ($10^{-4}m^3/s$)
1	7453	6865	1,97
2	6276	5492	2,36
3	5492	4119	1,97
4	7257	6668	1,97
5	7257	6767	2,28
6	6080	5296	3,30
7	7649	7355	1,89
8	7649	7453	3,15
9	6276	5296	2,20
10	6472	5688	2,20
11	6472	5688	2,36

Tabela 3: Valores da pressão externa, interna e vazão de todos ensaios.

Na tabela 4 estão indicados, em **dB**, os valores das intensidades da nota fundamental (**D_3**) e de seus harmônicos, indicados pela letra **H** e o respectivo índice para o segundo, terceiro, etc. harmônico.

4. ANÁLISE DOS RESULTADOS

O projeto fatorial permite determinar o efeito de cada variável (fator) sobre as intensidades e ainda verificar se existe interação entre elas.

As observações descritas a seguir estão agrupadas na tabela 5. Cada parágrafo corresponde a uma variável (ou fator). Assim, à medida que se lê um parágrafo pode-se acompanhar na tabela a respectiva variável para uma melhor compreensão.

Simulação da nota D_3												
ENSAIO	D_3	H_2	H_3	H_4	H_5	H_6	H_7	H_8	H_9	H_{10}	H_{11}	H_{12}
1	-29	-71	-28	-55	-32	-35	-35	-30	-34	-30	-42	-29
2	-21	-52	-22	-45	-29	-31	-31	-30	-30	-32	-36	-34
3	-22	-65	-23	-47	-29	-31	-33	-28	-37	-33	-54	-30
4	-22	-57	-22	-42	-40	-31	-37	-29	-41	-35	-38	-33
5	-20	-54	-23	-45	-39	-40	-35	-24	-40	-39	-45	-43
6	-29	-61	-29	-52	-32	-37	-32	-37	-32	-32	-39	-31
7	-18	-50	-23	-45	-39	-32	-36	-30	-36	-41	-40	-41
8	-26	-67	-26	-49	-32	-34	-34	-29	-36	-32	-48	-28
9	-21	-60	-23	-48	-31	-35	-34	-29	-39	-37	-40	-36
10	-21	-58	-22	-44	-34	-33	-37	-27	-50	-37	-37	-37
11	-22	-59	-23	-49	-32	-33	-33	-27	-38	-35	-47	-35

Tabela 4: Intensidade, em dB, da fundamental e de seus harmônicos para a região "chalumeau".

A segunda linha existente na tabela indica o valor médio das freqüências obtidas experimentalmente. A figura 6 mostra a concordância dos valores obtidos experimentalmente com a série de Fourier.

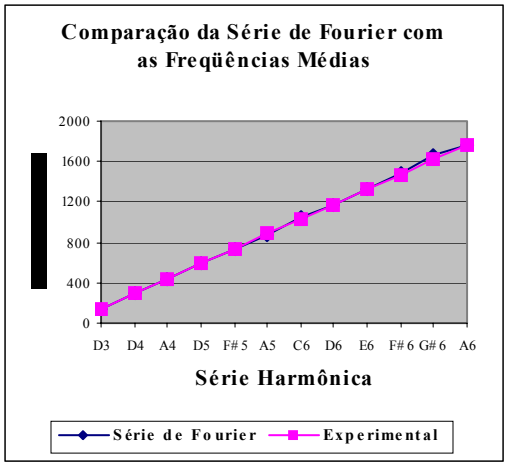

Figura 6: Comparação das Freqüências Experimentais com a Série de Fourier.

Uma análise preliminar mostra que a área de contato da mordedura com a palheta *(variável 5)* é o fator mais importante na determinação da intensidade da fundamental como da maioria dos harmônicos (exceção para o 6º, 7º e 9º). Esta variável atuou sobre grande quantidade de harmônico e as magnitudes são as mais elevadas. O aumento da área de contato de 0,3 cm^2 para 1,4 cm^2 *diminui* a intensidade da fundamental em 6,3 dB. Notamos que nesta região o 2º e o 4º harmônicos têm intensidade relativamente mais baixa que os demais harmônicos e o aumento da área de contato provoca um *decréscimo* da intensidade de 12,8 dB e 6,5 dB dos respectivos harmônicos. Os demais harmônicos pares apresentam intensidades médias equivalentes aos harmônicos ímpares. O nosso resultado está de acordo com o apresentado por HALL [14]. Este aumento de área de contato provoca uma *queda* de 4,0 dB no 3º harmônico, de 2,8 dB no 8º e 6,0 dB no 11º. O aumento da área de contato com a palheta provoca uma diminuição da área de vibração da palheta proporcionando assim a queda de suas intensidades. No entanto para o 5º, 10º e 12º harmônicos o efeito do aumento da área de contato é inverso, provocando um *aumento* de suas intensidades em 5,5 dB, 5,0 dB e 8,3 dB respectivamente. Para o 6º, 7º e 9º harmônicos o aumento da área de contato praticamente não altera suas respectivas intensidades.

A dureza da palheta atua significativamente em oito dos doze harmônicos. O volume da caixa e a posição da mordedura na palheta atuam igualmente em sete dos doze harmônicos. Logo em seguida, a abertura da boquilha atua em seis.

O sexto fator, quantidade de estopa dentro do volume da caixa, praticamente não tem efeito sobre a intensidade da fundamental e seus harmônicos.

Quando aumentamos a dureza da palheta *(variável 2)* de ´soft´ (Nº 2) para ´medium´ (Nº 3) o efeito observado é o *aumento* da intensidade da fundamental, 3º, 4º, 6º, 8º e 12º harmônicos, respectivamente, em 2,8 dB, 2,0 dB, 3,5 dB, 3,8 dB, 1,3 dB e 1,3 dB. Efeito contrário é notado para o 5º e 10º harmônicos, isto é, o aumento da dureza da palheta provoca uma *queda* da intensidade destes harmônicos igualmente em 2,0 dB. Para o 2º, 7º, 9º e 11º harmônicos a dureza da palheta praticamente não apresentou qualquer efeito.

Série de Fourier		(D3)	(D4)	(A4)	(D5)	(F#5)	(A5)	(C6)	(D6)	(E6)	(F#6)	(G#6)	(A6)
Freqüência Experimental (Hz)		*146 ±2*	*295 ±4*	*440 ±4*	*588 ±4*	*735 ±2*	*885 ±4*	*1031 ±5*	*1178 ±5*	*1325 ±4*	*1473 ±6*	*1620 ±5*	*1770 ±7*
FATORES, INTERAÇÃO e desvio padrão	**1**	**-2,3**	0,8	-0,5	1,0	**1,5**	**1,3**	1,3	**-3,3**	2,0	**3,0**	5,0	**4,3**
	2	**2,8**	-0,3	**2,0**	**3,5**	**-2,0**	**3,8**	-1,8	**1,3**	-3,5	**-2,0**	-4,5	**1,3**
	3	0,3	**3,3**	**-1,5**	-0,5	**-3,0**	**-3,8**	-0,3	-0,8	-0,5	**-3,5**	-0,5	**-4,3**
	4	**-1,8**	**-5,3**	-0,5	-0,5	**-3,5**	**-2,3**	**-2,3**	**3,3**	-4,0	0,5	-1,0	0,8
	5	**-6,3**	**-12,8**	**-4,0**	**-6,5**	**5,5**	-0,8	1,3	**-2,8**	2,0	**5,0**	**-6,0**	**8,3**
	6	-0,3	-0,8	-0,5	-2,0	**2,0**	**1,8**	0,3	-0,3	3,5	1,0	2,5	**1,3**
	16+25+34	2,3	0,3	2,0	2,0	3,5	-0,3	1,8	3,8	0	0,5	-6,0	-0,3
	σ	**0,6**	**1,0**	**0,6**	**2,6**	**1,5**	**1,2**	**2,1**	**1,2**	**6,7**	**1,2**	**5,1**	**1,0**

Tabela 5: Efeito das variáveis e interação entre elas para a região grave.

O volume da caixa *(variável 1)* tem relativa importância na sonoridade da clarineta, como veremos adiante. Ao aumentarmos o volume da caixa de 60% para 70% de seu volume total o efeito verificado é a *queda* da intensidade da fundamental e do 8º harmônico em 2,3 dB e 3,3 dB respectivamente. O *aumento* da intensidade é notado para o 5º, 6º, 10º e 12º harmônicos, respectivamente, em 1,5 dB, 1,3 dB, 3,0 dB e 4,3 dB. A variação deste fator mostrou-se indiferente para o 2º, 3º, 4º, 7º, 9º e 11º harmônicos.

O efeito da variação da posição da mordedura na palheta *(variável 3)*, que corresponde à posição do lábio inferior na palheta mais para dentro ou mais para fora, não foi observado para a intensidade da fundamental nem para as intensidades do 4º, 7º, 8º, 9º e 11º harmônicos. No entanto, quando passamos a mordedura de uma posição mais interna para uma mais externa na palheta observou-se um *aumento* da intensidade do 2º harmônico em 3,3 dB e um *decréscimo* da intensidade do 3º, 5º, 6º, 10º e 12º harmônicos, respectivamente, em 1,5 dB, 3,0 dB, 3,8 dB, 3,5 dB e 4,3 dB.

Outra variável de relativo destaque é abertura da boquilha *(variável 4)*. Quando passamos de uma boquilha mais fechada (A) para uma mais aberta (C) observamos a *diminuição* da intensidade da fundamental e do 2º, 5º, 6º e 7º harmônicos em 1,8 dB, 5,3 dB, 3,5 dB, 2,3 dB e 2,3 dB, respectivamente. Efeito contrário foi constatado para o 8º harmônico aumentando-o em 3,3 dB. Nenhum efeito desta variável foi observado para o 3º, 4º e do 9º ao 12º harmônicos.

Após a análise das variáveis resta mencionar o efeito da interação entre elas. Esta interação, na realidade, é a soma do efeito das interações **16**+**25**+**34**, isto é, corresponde à *soma* do efeito das interações entre as variáveis **1** e **6**, **2** e **5** e finalmente **3** e **4**. Sendo a soma de três interações, em qualquer experimento desta região, dividindo os valores por 3, estes serão sempre menores que o desvio padrão. Este resultado é adequado, pois não se espera uma interação muito grande entre as variáveis.

O projeto fatorial assim elaborado permite obter um modelo linear nas variáveis envolvidas na região de estudo. Aqui serão apresentados *apenas* os modelos para a fundamental (com o sobrescrito $\mathbf{D_3}$), ver equação (1), para o segundo harmônico (com o sobrescrito $\mathbf{D_4}$), ver equação (2), terceiro ($\mathbf{A_4}$), equação (3) e quarto harmônicos ($\mathbf{D_5}$), equação (4). Porém o procedimento segue-se analogamente para os demais harmônicos.

O modelo linear é obtido a partir dos efeitos das variáveis sobre a respectiva nota, pois estes são proporcionais aos coeficientes lineares da série de Taylor. Ao passar do nível **–1** para **+1** caminhamos duas unidades, assim basta dividir por 2 os valores da tabela 5. O termo independente da equação corresponde ao valor médio dos oito experimentos do projeto, desconsiderando o ponto central. Os fatores $\mathbf{f_1}$, $\mathbf{f_2}$, etc são aqui denotados pelas variáveis independentes $\mathbf{x_1}$, $\mathbf{x_2}$, etc respectivamente.

Modelo para $\mathbf{D_3}$:

$$\mathbf{Y}^{\mathbf{D}_3} = -23{,}4 - 1{,}65\mathbf{x}_1 + 1{,}40\mathbf{x}_2 + 0{,}15\mathbf{x}_3 - 0{,}90\mathbf{x}_4 - 3{,}15\mathbf{x}_5 - 0{,}15\mathbf{x}_6 \quad (1)$$

Modelo para $\mathbf{D_4}$:

$$\mathbf{Y}^{\mathbf{D}_4} = -59{,}6 + 0{,}40\mathbf{x}_1 - 0{,}15\mathbf{x}_2 + 1{,}65\mathbf{x}_3 - 2{,}65\mathbf{x}_4 - 6{,}40\mathbf{x}_5 - 0{,}40\mathbf{x}_6 \quad (2)$$

Modelo para $\mathbf{A_4}$:

$$\mathbf{Y}^{\mathbf{A}_4} = -24{,}5 - 0{,}25\mathbf{x}_1 + 1{,}00\mathbf{x}_2 - 0{,}75\mathbf{x}_3 - 0{,}25\mathbf{x}_4 - 2{,}00\mathbf{x}_5 - 0{,}25\mathbf{x}_6 \quad (3)$$

Modelo para $\mathbf{D_5}$:

$$\mathbf{Y}^{\mathbf{D}_5} = -47{,}5 + 0{,}50\mathbf{x}_1 + 1{,}75\mathbf{x}_2 - 0{,}25\mathbf{x}_3 - 0{,}25\mathbf{x}_4 - 3{,}25\mathbf{x}_5 - 1{,}00\mathbf{x}_6 \quad (4)$$

5. CONCLUSÃO

Neste estudo inicial avaliamos a influência de algumas variáveis sobre a sonoridade "chalumeau" da clarineta. Observamos que a área de contato com a palheta mostrou ser a variável de maior influência. No entanto, neste experimento a variação da área de contato (mais de 450%) é muito grande com relação às demais variáveis. Como conseqüência, a magnitude de seu efeito é bastante superior. Por isso não devemos descartar a influência das outras variáveis. Vale mencionar que o material para absorção sonora não teve muita relevância na sonoridade.

Notamos que as variáveis agem de modo homogêneo ao longo dos harmônicos. Observamos também a grande

aproximação dos valores obtidos experimentalmente para as freqüências com a série de Fourier.

Nesta região o 2º e o 4º harmônicos têm intensidade bastante baixas, porém, para os harmônicos superiores, tanto os pares como os ímpares possuem intensidades da mesma ordem de grandeza, aproximadamente.

Assim, na região grave, o 2º e o 4º harmônicos têm intensidades bastante reduzidas quando aumentamos a área de contato com a palheta.

O aumento do volume vazio, de 60% para 70% do volume total provoca uma diminuição da intensidade da fundamental e de seu 8º harmônico, ver tabela 5. Porém ele provoca o aumento da intensidade do 5º e 10º e do 6º e 12º harmônicos. Os valores envolvidos apresentam magnitudes relativamente altas para a pequena variação de apenas 10% do volume total.

Quando passamos a mordedura de uma posição mais interna para uma mais externa notamos a tendência em enriquecer o 2º harmônico e empobrecer o 4º. No entanto consultando a tabela 5 notamos que os harmônicos superiores também são empobrecidos quando a posição da mordedura é mais externa.

O aumento da abertura da boquilha provoca um enfraquecimento de praticamente quase todos os harmônicos, porém notamos que esta queda de intensidade é mais notabilizada na fundamental, 2º e do 5º ao 8º harmônicos.

Certamente uma nova seqüência de experimentos deve ser elaborada. Porém nesta nova etapa devemos diminuir a variação da área de contato e aumentarmos a variação do volume do tanque pulmão, além de eliminar o uso de material para absorção sonora. Este novo conjunto de experimentos poderá ser efetuado em uma outra região que poderá ser encontrada através do "steepest ascent" na direção de otimização de algum dos harmônicos.

6. REFERÊNCIAS BIBLIOGRÁFICAS

[1] HELMHOLTZ, H.L.F. (1877). *On the Sensations of Tone as a Physiological Basis for the Theory of Music.* 4th ed., trad. ELLIS, A.J. (Dover, New York, 1954)

[2] BACKUS, J. (1985). *The effect of the player´s vocal tract on the woodwind instrument tone.* J. Acoust. Soc. Am. **78**, 17-20.

[3] BACKUS, J. (1978). *Multiphonic tones in the woodwind instruments.* J. Acoust. Soc. Am. **63**, 591-599.

[4] BACKUS, J. (1974). *Input impedance curves for the reed woodwind instruments.* J. Acoust. Soc. Am. **56**, 1266-1279.

[5] BACKUS, J. (1963). *Small-vibration theory of the clarinet.* J. Acoust. Soc. Am. **35**, 305-313; erratum (1977) **61**, 1381-1383.

[6] BACKUS, J. (1961) *Vibrations od the reed and air column in the clarinet.* J. Acoust. Soc. Am. **33**, 806-809.

[7] BENADE, A.H. e KOUZOUPIS, S.N. (1988). *The clarinet spectrum: Theory and experiment.* J. Acoust. Soc. Am. **83**, 292-304.

[8] BENADE, A.H. e LARSON, C.O. (1985). *Requirements and Techniques for measuring the musical spectrum of the clarinet.* J. Acoust. Soc. Am. **78**, 1475-1498.

[9] BENADE, A.H. (1976). *Fundamentals of Musical Acoustics.* Oxford University Press, New York.

[10] BENADE, A.H. e GANS, D.J. (1968). *Sound Production in wind instruments.* Ann. N.Y. Acad. Sci. **155**, 247-263.

[11] BENADE, A.H. (1966). *Relation of air-column resonances to sound spectra produced by wind instruments.* J. Acoust. Soc. Am. **40**, 247-249.

[12] NEDERVEEN, C.J. (1998). *Acoustical Aspects of Woodwind Instruments.* Northern Illinois University Press, DeKalb, Illinois.

[13] BOX, G.E.P.; HUNTER, W.G.; HUNTER, J.S. (1978). *Statistics for Experimenters – An Introduction to Design, Data Analysis and Model Building.* John Wiley & Sons, NY.

[14] HALL, D.E. (1990). *Musical Acoustics.* Pacific Grove,CA: Brooks/Cole Publishing Montery:Brooks/Coole.

[15] FLETCHER, N.H.; ROSSING, T.D. (1998). *The Physics of Musical Instrument.* Springer-Verlag, 2nd ed., NY.

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na IX Convenção Nacional
11 – 13 de Abril de 2005, São Paulo, SP

Este artigo foi reproduzido do original entregue pelo autor, sem edições, correções e considerações feitas pelo comitê técnico deste evento. 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 reservados. Não é permitida a reprodução total ou parcial deste artigo sem autorização expressa da AES Brasil.

Análise da Sonoridade Pianística Através do Critério de Adequação da Síntese Evolutiva

José Fornari[1], Iracele Livero[2], Jônatas Manzolli[1,2]
1. Núcleo Interdisciplinar de Comunicação Sonora (NICS)
2. Instituto de Artes (IA)
Universidade de Campinas, (UNICAMP)
Cidade Universitária "Zeferino Vaz" Campinas, SP, Brasil
[fornari, jonatas]@nics.unicamp.br; iracelelivero@uol.com.br

RESUMO

Apresentamos nesse trabalho um método de análise de sonoridade pianística com base no critério de adequação da síntese evolutiva [6]. O estudo foi feito através da gravação de 12 pianistas interpretando, no mesmo piano, 4 obras que representam respectivamente 4 critérios escolhidos de sonoridade, a saber: intensidade, *legato*, *staccato* e pulsação rítmica. O critério de adequação é baseado na distância euclidiana das curvas psicoacústicas dos segmentos sonoros gravados. Espera-se desse modo investigar uma possível utilização prática desse trabalho para a avaliação automática e impessoal da qualidade da sonoridade na interpretação pianística.

INTRODUÇÃO

Em meados do século XVIII, o aparecimento de um sistema de marteleiras acionado por um teclado que percutia cordas metálicas possibilitou a criação do instrumento chamado *Pianoforte* (posteriormente chamado apenas de *Piano*) e fez dele o sucessor histórico do Cravo, diferenciando-se do seu predecessor por ser capaz de oferecer recursos de expressividade e ampla variação na intensidade sonora (razão do nome *Pianoforte*, que em Italiano significa Fraco e Forte).

Estes novos recursos proporcionaram aos artistas uma nova gama de possibilidades musicais, estendendo o horizonte composicional e trazendo um grande impacto para a linguagem musical, como pode-se observar através das obras de diversos compositores, como: Mozart, Beethoven, Chopin entre outros.

Entretanto, o aparecimento desta nova ferramenta passou a exigir um maior preparo técnico na execução, tanto pela possibilidade de variar a intensidade do som, como pelo domínio do repertório. Desde então, estudos foram desenvolvidos com a finalidade de melhorar o desenvolvimento técnico do músico pianista. Entende-se por técnica pianística o conjunto dos movimentos gesticulares empregados na execução do instrumento e suas características, cujo objetivo fundamental é a precisão e o domínio da sonoridade gerada pelo instrumento. A técnica pianística deverá ser então encarada como um meio empregado para se obter uma interpretação de boa qualidade.

Um dos assuntos controversos, questionados entre pianistas e físicos é aquele que diz respeito à possibilidade da técnica pianística poder ou não variar o timbre de uma mesma nota ou acorde (grupo de notas), independente da variação de sua

intensidade. Muitos cientistas, ao analisarem o mecanismo do instrumento acreditavam que a única variação possível corresponderia necessariamente a uma variação da intensidade desse som [1].

No entanto, minuciosos experimentos comprovaram que duas notas emitidas com a mesma intensidade e com toques diferenciados – por exemplo, percussivos ou não-percussivos – possuem uma diferença na composição dos harmônicos que constituem o timbre daquele som [2].

A sonoridade consiste, portanto, na qualidade do timbre dos sons produzidos pelo pianista, que a partir do seu conhecimento, domínio técnico e ouvido musical será capaz de obter os melhores resultados. Timbre é formalmente definido pela ASA (*American Standart Association*) como " aquele atributo do sentido auditivo em termos do qual o ouvinte pode julgar que dois sons apresentados como similares, com a mesma intensidade e altura, não são similares.". Sabe-se que qualquer variação de intensidade ou de freqüência corresponda a uma variação do timbre. Uma determinada nota tocada *mezzoforte* terá um timbre diferente quando tocada *forte*, independente do aumento da sua intensidade [3].

Constitui tarefa complexa conseguir amplo domínio sobre os movimentos gesticulares necessários à boa execução do piano, ainda mais tendo que possuir os conhecimentos adequados à compreensão de uma obra musical, e assim desenvolvê-los e vivenciá-los de acordo com a interpretação musical desejada. A interpretação musical não dispensa estudos intelectuais refinados, como as características dos diferentes estilos de cada compositor, maneiras de se fazer um fraseado, ornamentações e estilo da época da composição.

Para isto, as versões finais devem ser enviadas em arquivos postscript (*.ps) ou em PDF (*.pdf), segundo este formato.

Itens principais (veja acima) são em letras maiúsculas, fonte Helvetica, tamanho 8, estilo negrito, alinhado à esquerda. O texto, propriamente dito, é em fonte Times Roman, tamanho 9 e justificado (como este).

A Arte da Interpretação

O grande momento de mudança para a interpretação moderna ao piano se deu nos anos vinte. Em reação aos excessos um tanto arbitrários da geração anterior, os pianistas proclamaram a necessidade da "fidelidade ao texto musical", demonstrando uma grande personalização da interpretação pianística [4].

A carreira de pianista hoje está muito mais rápida e internacionalizada e o repertório cada vez mais amplo. Isso faz com que se aumentem as exigências técnicas para o pianista que, em tempos passados, podia ter um repertório mais restrito e especializado. Como se pode então qualificar um pianista? Existem certamente as regras estéticas tais como: ritmo, estilo e forma, a época em que o compositor viveu, características do instrumento, entre outros, que o pianista deve conhecer e sentir, a fim de que possa alcançar uma melhor interpretação. Torna-se essencial, portanto, dominar todas as características técnicas de uma composição, ao lado de um perfeito domínio técnico, o que constitui a base necessária para uma boa interpretação.

Interpretar, portanto, quer dizer submeter-se a uma experiência de auto-análise. Estudar o repertório, encontrar seus limites, identificá-los e tentar outros meios de contorná-los. O mundo da interpretação não é então determinístico e se torna inconsistente querer apresentar sempre interpretações idênticas ao público ouvinte.

Neste trabalho para análise de sonoridade pianística baseado no critério de adequação da síntese evolutiva, foram considerados os seguintes parâmetros para a análise de uma interpretação: pulsação rítmica, *legato*, *staccato*, e variação da intensidade. Apesar de não serem os únicos, podem ser critérios básicos.

Duas escolhas diferentes se colocam para nós na questão do tempo: a decisão da pulsação como base de um movimento e a decisão da maneira de conduzi-lo. Por "tempo" entende-se o conjunto de regras que regulam o movimento ordenado dos sons e exige a presença de um valor mensurável, uma métrica constante. O "ritmo" é o elemento fundamental do tempo e desempenha um papel em muitos outros aspectos da música, sendo um importante elemento na melodia, afeta a progressão da harmonia e desempenha papéis em questões como "textura musical", timbre e ornamentações. A pulsação musical pode ser identificada quando os sons são apoiados em intervalos regulares de tempo. Ciclos repetidos de pulsos regulares estabelecem uma métrica.

Outra maneira de qualificar uma interpretação além dos critérios já mencionados é a maneira como o pianista descreve o seu discurso musical, especificamente falando de fraseados e suas articulações. "Frasear" quer dizer pronunciar. O pianista projeta um texto para o público comparável a um ator em cena, ou à leitura de uma poesia.

A articulação diz respeito ao modo propriamente dito, a qualidade do toque e suas transformações ao longo do tempo. Nesse contexto, o toque *legato* se define como uma execução de sons sem interrupção perceptível e sem ênfase especial. Verifica-se melhor na voz humana ou em instrumentos que permitem manter um som sem o seu respectivo decréscimo. É evidente que no piano, onde cada som se apresenta invariavelmente com intensidade decrescente, uma perfeita conexão entre as notas não é acusticamente possível. Porém, executá-las com toques percussivos, mesmo prolongando cada som até o início do próximo, diminuirá a sensação de *legato*. Pode-se dizer que a qualidade do timbrística do som gerado pelo pianista influencia na sensação do legato.

Em oposição a este, o toque *staccato* se define exclusivamente pela duração do som, independente de que intensidade ou tipo de toque esse som seja emitido. Porém, não se justifica associar o staccato a um som de má qualidade tímbrica proveniente de um toque percussivo. É perfeitamente possível executá-lo com qualidade tímbrica usando os mecanismos técnicos necessários para isso.

A amplitude do som se refere à dinâmica. Encontra-se em compositores como Beethoven, oposições combinadas de registro e de dinâmica, por exemplo nota aguda em *pianíssimo* (*ppp*) com uma nota grave marcada *fortíssimo* (*fff*) [18]. Sabendo como os agudos da época eram transparentes, deve-se evidentemente reforçar esse efeito para dar um melhor entendimento no piano atual.

É legítima a possibilidade de graduar do "mais docemente possível" ao mais "forte possível", ou seja do *ppp* ao *fff*. Esta variação de intensidade pode às vezes vir ligada ao tempo, quando se verificará um acelerando ligado ao crescendo, ou *ritardando* ao *decrescendo*. Cabe ao intérprete tomar a decisão mais coerente. Isto leva a essencial noção de "cor".

Um outro ponto importante na interpretação é o emprego correto do pedal do instrumento e seus efeitos sobre o legato,

o staccato entre outros [18]. O pedal, a princípio, é a amplificação da ressonância pela sustentação do som, portanto deve ser usado com parcimônia. Todas as graduações são possíveis com um emprego bem controlado do pedal, e seu efeito está diretamente relacionado à questão de estilo.

Como se pode observar, o estudo da sonoridade pianística se baseia em conceitos estéticos que, como tais, são fortemente influenciados pelo gosto pessoal do ouvinte. Nesse sentido, a avaliação da sonoridade de um pianista fica sempre atrelada ao critério pessoal de cada membro de uma banca julgadora. Por outro lado, não é possível criar algoritmos determinísticos que sejam adequados para a avaliação da sonoridade de uma performance pianística uma vez que este problema pertence a categoria dos problemas não-determinística, ou seja, problemas que não apresentam uma solução determinada, ou fixa. Nesse sentido a utilização de algoritmos evolutivos para a solução desse problema parece ser bastante pertinente. Em nosso trabalho utilizamos o processo de seleção da síntese evolutiva [5] como critério de adequação para a busca da gravação digital com melhor sonoridade, baseada num conjunto de gravações previamente selecionadas como possuindo diferentes aspectos de uma ótima sonoridade.

A SÍNTESE EVOLUTIVA

O método ESSynth utiliza conceitos da computação evolutiva para a geração de segmentos sonoros cujo comportamento sônico tem características dinâmicas (i.e. entre outros parâmetros, o comportamento psicoacústico da intensidade, altura e espectro se modificam ao longo do tempo). O ESSynth gera seqüências sonoras que, como nos processos biológicos, tem o potencial de evoluir ao longo do tempo, segundo conceitos inspirados na teoria da evolução de Darwin. Mais especificamente, toma-se emprestado o conceito de seleção biológica das espécies onde os segmentos sonoros são representados como indivíduos de uma população sobre a qual agem dois processos evolutivos: Seleção e Reprodução. A Reprodução é feita por algoritmos genéticos (GAs) e a Seleção é feita por uma função de adequação (função de *fitness*) utilizando-se a medida de distância denominada distância de *Hausdorff* que avalia o conteúdo perceptual dos segmentos sonoros gerados.

O método completo da ESSynth pode ser visto em [5]. O trabalho apresentado aqui projeto propõe o desenvolvimento de um aplicativo que deverá ser capaz de processar o método ESSynth em tempo real e controlar os parâmetros da síntese dinamicamente. A implementação em software da síntese evolutiva será também chamada de ESSynth (*Evolutionary Sound Synthesizer*) e virá a ser, ao que sabemos, um dos primeiros sintetizadores em software desenvolvidos com tecnologia evolutiva.

Na síntese evolutiva os indivíduos são amostras discretas (digitais) de segmentos sonoros com uma dada taxa de amostragem e resolução. O conjunto de todos os indivíduos compõe a população, onde ocorre a evolução. O curso da evolução da população é condicionado através de uma distância propriamente definida pelas características perceptuais sonoras dos indivíduos de outro conjunto, o conjunto alvo. A evolução da população ocorre em estágios, cada qual sendo chamado de geração.

A evolução da população é feita por dois processos: a reprodução e a seleção. Em cada geração a reprodução gera novos indivíduos e a seleção escolhe o melhor indivíduo da população, ou seja, o mais adaptado aos critérios dados pelo conjunto alvo. É chamado de genótipo do indivíduo o conjunto de características psicoacústicas que o caracterizam percentualmente. Consideramos como genótipo três curvas psicoacústicas: a curva de *loudness* (percepção da intensidade sonora), *pitch* (percepção da freqüência do harmônico fundamental do som) e espectro (a composição harmônica do segmento sonoro). Os processos de reprodução e seleção agem sobre o genótipo de cada indivíduo.

No processo de reprodução agem dois operadores genéticos: *crossover* e mutação. O *crossover* permuta características sonoras dos indivíduos em reprodução (os progenitores). A mutação insere modificações aleatórias nessas características, aumentando assim a diversidade da população. Chamamos assim de genótipo do indivíduo o conjunto de características perceptuais sonoras que o compõem, ou seja, suas grandezas psicoacústicas. O processo de evolução atua sobre os genótipos dos indivíduos.

No processo de seleção, cada indivíduo pertencente à população é comparado com um conjunto de indivíduos chamado alvo, através de uma função de adequação (*fitness*), que mede a distância entre o genótipo do indivíduo e todos os genótipos do conjunto alvo. A medida de distância utilizada neste trabalho é a distância euclidiana das curvas psicoacústicas que compõem o genótipo de cada indivíduo.

O resultado sonoro deste método de síntese é o segmento sonoro escolhido pelo processo de seleção como melhor indivíduo (indivíduo na população com menor distância ao conjunto alvo). A cada geração da população o processo de seleção busca pelo melhor indivíduo da população. Ao longo das gerações tem-se uma sucessão de melhores indivíduos que, como segmentos sonoros, tende a convergir para um único segmento sonoro, isso considerando que o conjunto alvo permaneça inalterado.

AS CURVAS PSICOACÚSTICAS DOS SEGMENTOS SONOROS

A medida da sonoridade pianística foi aqui realizada a partir da gravação digital de 12 pianistas, cada qual tocando 4 peças musicais que expressam qualidades distintas da sonoridade pianística. Estas são: *legato* (tocar as notas o mais ligadas possível), *staccato* (tocar as notas o mais destacadas possível), pulsação (tocar as notas o mais ritmado possível) e intensidade (explorar as variações de intensidade sonora entre as notas). Para isso utilizou-se o método de extração do genótipo sonoro [6], para se obter as 3 curvas psicoacústicas que compõem o genótipo sonoro de cada gravação digital. O conjunto das gravações dos 12 pianistas é o nosso conjunto População, onde cada gravação digital é chamada de individuo.

Foi feita a análise da sonoridade pianística de cada indivíduo através da medida de distâncias entre seus genótipos, e os genótipos de um conjunto de segmentos sonoros da gravação de 3 pianistas que, dentro da população consideramos como apresentando a melhor sonoridade, que corresponde a nosso conjunto Alvo. Os gráficos a seguir mostram a distância entre os genótipos de cada indivíduo em relação ao conjunto Alvo. Note que os indivíduos que

apresentam distância nula são aqueles que pertencem ao próprio conjunto Alvo (no caso da figura a seguir, os indivíduos 1, 4 e 11).

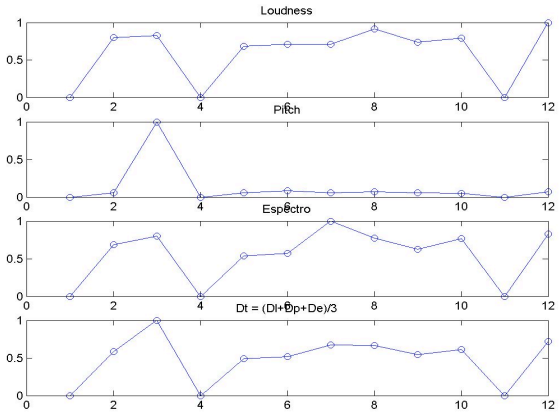

Figura 1: Análise da distância do critério PULSAÇÃO. Dl=distância de loudness, Dp=Distância de pitch, De=distância de espectro, Dt=média das três distâncias anteriores

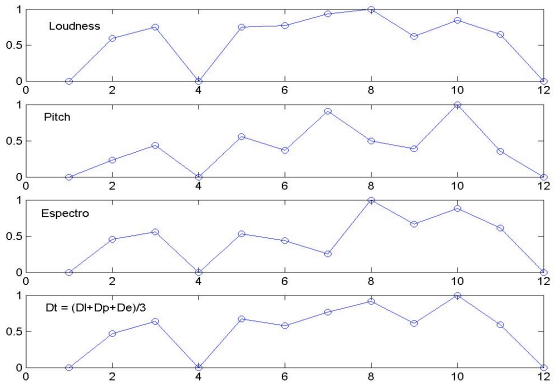

Figura 2: Análise da distância do critério LEGATO. Dl=distância de loudness, Dp=Distância de pitch, De=distância de espectro, Dt=média das três distâncias anteriores.

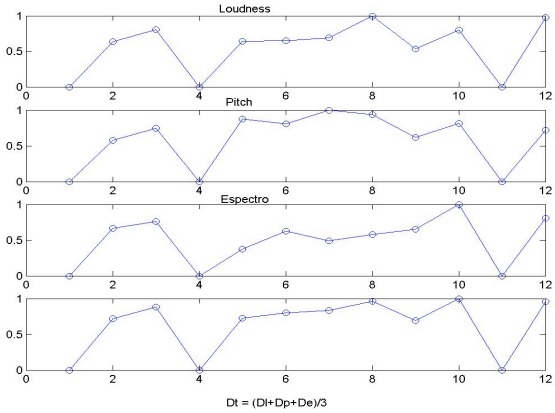

Figura 3: Análise da distância do critério STACCATO. Dl=distância de loudness, Dp=Distância de pitch, De=distância de espectro, Dt=média das três distâncias anteriores.

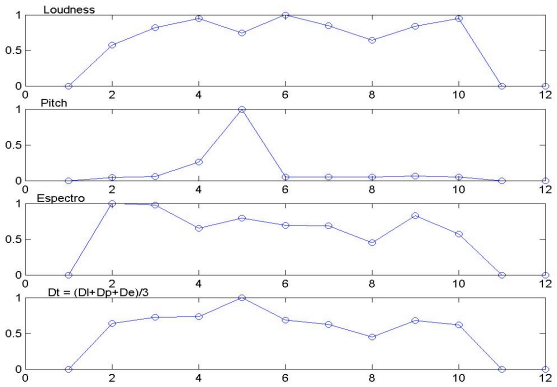

Figura 4: Análise da distância do critério INTENSIDADE. Dl=distância de loudness, Dp=Distância de pitch, De=distância de espectro, Dt=média das três distâncias anteriores.

Definições usadas nos experimentos

Indivíduo: segmento sonoro, gravação da performance de um pianista em áudio digital: 44.1KHz, 16bits, mono.
Melhor indivíduo: indivíduo mais próximo ao conjunto Alvo, baseado na medida de distância da função de adequação.
Conjunto de População: indivíduos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 e 12
Conjunto Alvo: indivíduos considerado com melhor sonoridade em cada categoria.
Dl: distância para curva psicoacústica de loudness
Dp:distância para curva psicoacústica de pitch
De: distância para curva psicoacústica de espectro
D: média aritmética das distâncias das 3 curvas psicoacústicas

RESULTADOS:

Pulsação	Legato	Staccato	Intensidade
Conj. Alvo:1,4,11	Conj.Alvo: 1,4,12	Conj. Alvo: 1,4,11	Conj. Alvo:1,11,12
Dl = 5	Dl: = 2	Dl = 9	Dl = 2
Dp =10	Dp = 2	Dp = 2	De = 2
De = 5	De = 7	De = 5	De = 8
D = 5	D = 2	D = 2	D = 8

Figura 5- Resultados obtidos nas análises, pelo critério de adequação. Entre são os indivíduos, entre os 12 indivíduos da população, que apresentaram as menores distâncias com seu respectivo conjunto alvo, de 3 indivíduos.

PULSAÇÃO				LEGATO			
Dt	Dl	Dp	De	Dt	Dl	Dp	De
0	0	0	0	0	0	0	0
0.5858	0.7998	0.0568	0.6826	0.4690	0.5920	0.2334	0.4543
1.0000	0.8243	1.0000	0.8028	0.6393	0.7487	0.4359	0.5599
0	0	0	0	0	0	0	0
0.4871	0.6865	0.0557	0.5374	0.6742	0.7548	0.5565	0.5282
0.5190	0.7075	0.0823	0.5737	0.5752	0.7691	0.3651	0.4353
0.6734	0.7102	0.0588	1.0000	0.7692	0.9377	0.9071	0.2542
0.6667	0.9130	0.0681	0.7705	0.9153	1.0000	0.4974	1.0000
0.5417	0.7412	0.0578	0.6242	0.6136	0.6238	0.3853	0.6652
0.6141	0.7937	0.0536	0.7660	1.0000	0.8475	1.0000	0.8811
0	0	0	0	0.5918	0.6474	0.3541	0.6134
0.7212	1.0000	0.0699	0.8248		0	0	0

Figura 6: Distâncias entre cada indivíduo da população e o conjunto alvos da pulsação e do legato.

STACCATO				INTENSIDADE			
Dt	Dl	Dp	De	Dt	Dl	Dp	De
0	0	0	0	0	0	0	0
0.5762	0.6339	0.5758	0.6622	1.0000	0.5766	0.0446	1.0000
0.7484	0.8036	0.7478	0.7575	0.9781	0.8191	0.0548	0.9781
0	0	0	0	0.6516	0.9496	0.2611	0.6516
0.8734	0.6407	0.8733	0.3728	0.7935	0.7476	1.0000	0.7935
0.8095	0.6477	0.8093	0.6239	0.6910	1.0000	0.0532	0.6910
10.000	0.6906	10.000	0.4920	0.6868	0.8483	0.0478	0.6868
0.9360	10.000	0.9352	0.5778	0.4518	0.6413	0.0510	0.4518
0.6214	0.5352	0.6212	0.6525	0.8257	0.8428	0.0627	0.8257
0.8116	0.7980	0.8112	10.000	0.5738	0.9501	0.0483	0.5738
0	0	0	0	0	0	0	0
0.7205	0.9744	0.7194	0.8089	0	0	0	0

Figura 7: Distâncias entre cada indivíduo da população e o conjunto alvo do staccato e intensidade.

CONCLUSÃO E FUTURAS PESQUISAS

Observando os resultados da análise da sonoridade com base no critério de adequação da síntese evolutiva de cada parâmetro estabelecido, pode-se chegar a seguinte conclusão:

Quanto à Pulsação: O trecho escolhido para essa análise foi uma peça de Bela Bartok: *Bear Dance.* O indivíduo escolhido não condiz com o critério de análise pulsação, visto que há oscilações do ritmo, não mantendo uma mesma pulsação do início ao fim do trecho. Acredito que uma influência sobre esta escolha tenha sido o fato de que o toque pianístico esteja similar (o que pode ser observado no gráfico da figura 1), bem como também o andamento inicial estabelecido é muito próximo.

Quanto ao *Legato:* O trecho escolhido para essa análise foi um trecho do *Noturno op. 9 n. 2* de Chopin. Observou-se que o indivíduo escolhido teve grande semelhança na qualidade do toque pianístico e do *loudness*. Assim sendo a qualidade tímbristica do som do pianista influenciou diretamente no critério *legato,* ou seja, na execução sem interrupção perceptível dos sons, tornando-os o mais interconectado possível. O pianista escolhido pelo processo da adequação da síntese evolutiva usou sem moderação o pedal de sustentação do instrumento, mesmo que misturando as harmonias, o que criou uma ilusão maior de *legato.*

Quanto ao *Staccato:* O trecho escolhido para essa análise foi um trecho da peça de Bela Bartok: *Jeering Song.* O indivíduo escolhido está de acordo com o critério *staccato* principalmente pela qualidade do toque pianístico, ou seja, a qualidade da sonoridade pianística e a clareza de cada som.

Quanto à Intensidade: O trecho escolhido para essa análise foi a introdução de *Estro Africano 1* da compositora brasileira Eunice Katunda [17]. Neste critério pode-se haver mais de um indivíduo se não levarmos em conta a reverberação dos sons. O processo de adequação da síntese evolutiva considerou o pianista que, além de realizar os *crescendos* e *decrescendos* com acuidade, também foi o que mais utilizou o pedal de sustentação do instrumento. Isto não deprecia um outro pianista, que tenha conduzido bem uma variação da intensidade, com uma boa qualidade sonora, mas que tenha utilizado o pedal com menor freqüência.

Consideramos os resultados experimentais obtidos encorajadores para o prosseguimento da pesquisa. O próximo passo seria a implementação desse algoritmo em um aplicativo de software capaz de avaliar os critérios de sonoridade (staccato, legato, pulsação e intensidade), e mostrá-los em uma graduação numérica enquanto o pianista estiver executando a peça, em um concurso de piano, ou mesmo estudando em casa. Em uma segunda etapa da pesquisa iremos testar o segundo processo da síntese evolutiva [7], a utilização dos operadores genéticos, para a reconstrução da sonoridade de uma peça gravada, de acordo com os critérios de adequação de sonoridade estudados. Isto seria similar à correção de afinação de voz que os estúdios de gravação podem fazer para os cantores, só que a correção nesse caso seria feita nos critérios de sonoridade estudados, de acordo com modelos de sonoridade dados por gravações de pianistas renomados pela boa sonoridade. Também nada impede que este método se estenda para o estudo da sonoridade de outros instrumentos musicais, ou da voz humana. São grandes as possibilidades de aplicação deste método para música e acreditamos que muitas delas irão surgindo à medida que continuarmos o seu desenvolvimento. As gravações dos pianistas utilizadas para a análise da sonoridade estarão disponíveis para serem escutadas no site http://www.nics.unicamp.br/~fornari/epp

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Allen. F. J. 1913. *Pianoforte touch*. Nature 91(2278),424-425.

[2] Askenfelt, A., Galembo, A., Cuddy, L. E. 1998. *On the accoustics and psychology of piano touch and tone*. Journal of Acoustical Society of America. 103(5 Pt. 2), 2873.

[3] Bresin, R., Battel, G. 2000. *Articulation strategies in expressive piano performances*. Journal of New Music Research. 29 (3), 211-224.

[4] Fornari, J. 2003. *A Síntese Evolutiva de Segmentos Sonoros*. PhD Dissertation. Faculty of Electrical Engeering, State University of Campinas (UNICAMP). Brazil.

[5] Levaillant, Denis. Le Piano. J.C. Lattès/Salabert,1986,p.53

[6] Manzolli, J., Maia Jr. A., Fornari, J. & Damiani, F. 2001. *The evolutionary sound synthesis method*. Proceedings of the ninth ACM international conference on Multimedia, Ottawa, Canada, 585 – 587, ISBN:1-58113-394-4.

[7] Moroni, A., Manzolli, J., Von Zuben, F. and Gudwin, R. 2000. *Vox Populi: An Interactive Evolutionary System for Algorithmic Music Composition*, San Francisco, USA: Leonardo Music Journal, - MIT Press, Vol 10, pg 49-54.

[8] Moroni, A, von Zuben, F. and Manzolli, J. 2002. *ArTbitration*, San Francisco, USA: Leonardo Music Journal - MIT Press, 2002, Vol:11-45-55.

[9] Richerme, C. 1996. *A técnica pianística. Uma abordagem científica*. S.João Boa Vista. Air Musical. p.27 and 28.

[10] Repp. B. H. 1993. *Some empirical observations on sound level properties of recorded piano tones*. Journal of the Acoustical Society of America. 93(2),1136-44.

[11] Repp. B. H. 1996. *Patterns of note onset asynchronies in expressive piano performances*. Journal of the Acoustical Society of America. 100(6),3917-3932.

[12] Shaffer, L. H. 1981. *Performances of Chopin, Bach and Bartòk: Studies in motor programming*. Cognitive Psycology 13,326-376.

[13] Tro. J. 1998. *Micro dynamics deviation as a measure of musical quality in piano performances?*. In Proceedings of the 5th International Conference on Music Perception and Cognition (ICPMC5), August, 26-30, edited by S. W. Yi (Western Music Research Institute, Seoul National University, Seoul, Korea).

[14] Wasserman, K.C., Eng, K., Verschure, P.F.M.J., Manzolli, J. 2003. *Live Soundscape Composition Based on Synthetic Emotions*. In: Published by IEEE Multimedia Computer Society, Vol:10:4, p. 82-90.

[15] Widmer, G. 2001. *Using AI and machine learning to study expressive music performance: Project survey and first report.* AI Communications 14(3), 149-162.

[16] Goebl, W. 2003. *The role of timing and intensity in the Production and Perception of Melody in Expressive Piano Performance*. PhD Dissertation, Institut fur Musikwissenschaft, Karl-Franzens University, Germany.

[17] Livero, Iracele. Pesquisa de Doutorado. *A Música Brasileira na Obra Pianística de Eunice Katunda: Uma visão histórica, Crítica e Analítica*. Unicamp, 2004.

[18] Sadie, Stanley. *Dicionário de Música. Edição concisa*. Rio de Janeiro: Zahar, 1994. pp. 708,709.

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na IX Convenção Nacional
11 – 13 de Abril de 2005, São Paulo, SP

Este artigo foi reproduzido do original entregue pelo autor, sem edições, correções e considerações feitas pelo comitê técnico deste evento. 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 reservados. Não é permitida a reprodução total ou parcial deste artigo sem autorização expressa da AES Brasil.

Sistema MIDI para Auxílio de Músicos Portadores de Necessidades Especiais

Flávio Bressan da Luz[1], Sidnei Noceti Filho[2], André Luís Dalcastagnê[2]
1 – Brasil Telecom S.A.
SIA/SUL – ASP, Lote D, Bloco A, Térreo
71215-000, Brasília, DF, Brasil
flavio@bressan.eng.br
2 – LINSE - Laboratório de Circuitos e Processamento de Sinais
Departamento de Engenharia Elétrica, UFSC
Telefone: (48) 331-9504, Fax: (48) 331-9091
88040-900, Florianópolis, SC, Brasil
sidnei@linse.ufsc.br andre@linse.ufsc.br

RESUMO

Este trabalho apresenta um sistema que, acoplado a um teclado ou módulo com entrada MIDI, permite que uma pessoa portadora de algum tipo de deficiência física na mão esquerda possa executar uma quantidade razoável de acordes completos. O sistema é formado por um miniteclado com doze teclas (tônicas) e por uma pedaleira composta por quinze chaves, responsáveis pela seleção do tipo de acorde desejado.

INTRODUÇÃO

Uma música é composta pela harmonia (acordes de três ou mais notas) e pela melodia (solo). Cada acorde é formado por uma nota principal (tônica) e por um conjunto de duas ou mais notas que o caracterizam. Por exemplo, no caso do acorde dó menor com sétima (Cm7), a tônica é o dó e as notas que o caracterizam são o mi bemol, o sol e o si bemol. Em um teclado, o músico normalmente executa o acorde com a mão esquerda e a melodia com a mão direita. Portanto, para executar um acorde Cm7, o músico precisa utilizar quatro dedos da mão esquerda. Sendo assim, uma pessoa portadora de algum tipo de deficiência física na mão esquerda não pode executar tal acorde em um teclado convencional.

O objetivo deste trabalho é desenvolver um sistema que acoplado a qualquer tipo de instrumento musical com entrada MIDI (acrônimo de *musical instruments digital interface*) [1] possibilite a uma pessoa portadora de algum tipo de deficiência física na mão esquerda executar uma quantidade razoável de acordes completos (quinze). O sistema proposto é formado por um miniteclado com doze teclas (tônicas) e por uma pedaleira, que possui um conjunto de quinze chaves que selecionam o tipo de acorde que deve ser executado. Nesse sistema, é necessário utilizar apenas um dedo, ou alguma prótese ligada ao braço esquerdo, para executar uma tônica. As demais notas de um acorde são selecionadas através da utilização da pedaleira. No exemplo do acorde Cm7, o músico acionaria apenas uma tecla com um membro superior e uma das quinze chaves da pedaleira, correspondente ao acorde menor com sétima, com um membro inferior. As informações provenientes do miniteclado e da pedaleira são

então processadas e o código resultante é enviado através do protocolo MIDI para um teclado convencional.

Com as devidas alterações na disposição das chaves, é possível suprir as necessidades de outros tipos de deficiência física. Por exemplo, ao invés de um sistema com um miniteclado e uma pedaleira, pode-se pensar em utilizar duas pedaleiras. Além disso, o princípio exposto neste trabalho pode ser aplicado a outros tipos de instrumentos musicais que possuam entrada MIDI, tais como baterias eletrônicas, *expanders* e *samplers*.

O PROTOCOLO MIDI

Os primeiros sintetizadores comerciais analógicos surgiram no final da década de 1960. Originalmente, esses sistemas eram monofônicos, ou seja, eram capazes de tocar apenas uma nota por vez. Além disso, a programação dos sintetizadores era realizada através de cabos que interligavam diferentes circuitos eletrônicos, tais como osciladores, filtros e geradores de envoltória. Com a evolução da tecnologia, os sintetizadores analógicos foram recebendo novos recursos, como, por exemplo, a capacidade de tocar várias notas simultaneamente e a de memorizar diferentes conexões entre os cabos. O passo seguinte na evolução desses equipamentos foi a interligação de diferentes tipos de instrumentos musicais. Sintetizadores de diversos fabricantes passaram a se comunicar, com o propósito de criar estruturas sonoras complexas.

Com o advento dos processadores digitais, surgiram os instrumentos musicais digitais, tais como sintetizadores e *samplers*. Tal fato trouxe consigo a necessidade de comunicação entre instrumentos digitais de diferentes fabricantes. Para atender a esse requisito, alguns dos mais importantes fabricantes mundiais de instrumentos musicais eletrônicos (Sequential Circuits, Oberheim, Roland, Yamaha, Korg e Kawai) criaram no início da década de 1980 o protocolo de comunicação MIDI. Atualmente, esse protocolo está presente em praticamente todos os instrumentos e equipamentos eletrônicos profissionais e semiprofissionais dedicados à música, tais como baterias eletrônicas, seqüenciadores, *samplers*, sintetizadores, mesas de gravação e computadores.

O protocolo MIDI utiliza uma transmissão serial de dados, do tipo assíncrona, com *frame* de dez *bits* (um *start bit* + oito *bits* + um *stop bit*) e taxa de transmissão de 31250 *bits*/s. Essa transmissão pode ser feita em até dezesseis canais simultaneamente. Existem instruções MIDI destinadas à execução, à sincronização e à programação de instrumentos musicais. Uma mensagem MIDI é composta por um *status byte* e por um ou mais *data bytes*. O *status byte* identifica o tipo de instrução que está sendo enviada e é caracterizado por possuir o *bit* mais significativo igual a "1". O *data byte* fornece o dado necessário a essa instrução e é caracterizado por possuir o *bit* mais significativo igual a "0". A equação abaixo apresenta a típica estrutura de uma mensagem MIDI:

$$1xxxcccc\ \ 0ddddddd\ \ 0ddddddd\ \ 0ddddddd\ \cdots. \quad (1)$$

onde 1xxxcccc é o *status byte*; xxx é o código binário que identifica a instrução; cccc é o código binário que identifica o canal pelo qual essa instrução deve ser transmitida. O termo 0ddddddd representa o *data byte*, onde dddddd é o dado a ser transmitido.

Neste trabalho, somente as instruções MIDI "NOTE ON" e "NOTE OFF" são utilizadas. Elas apresentam as seguintes estruturas:

$$\text{NOTE ON:}\ \ 1001cccc\ \ 0nnnnnnn\ \ 0vvvvvvv\ , \quad (2)$$

$$\text{NOTE OFF:}\ \ 1000cccc\ \ 0nnnnnnn\ \ 0vvvvvvv\ , \quad (3)$$

onde cccc é o código binário que identifica o canal que deve ser utilizado para a transmissão dos dados; nnnnnnn é o código binário da nota a ser executada ou desligada; vvvvvvv é o código binário do volume dessa nota, relativo à velocidade com que a tecla é acionada. Neste trabalho, cccc é mantido em 0000, o que indica que o canal MIDI utilizado é o "1" e vvvvvvv é mantido em 1000000 (64 em decimal), que indica a velocidade média de acionamento de uma tecla.

De acordo com as Equações (2) e (3), é necessário o envio de três *bytes* para ligar ou desligar uma nota. Neste trabalho, os acordes disponíveis possuem quatro notas, o que implica na transmissão de doze *bytes* para cada instrução. Entretanto, pode-se utilizar o recurso MIDI *running status* [1], que permite diminuir o número de *bytes* que devem ser transmitidos caso seja necessário enviar várias instruções iguais mas com dados diferentes. Assim, o número de *bytes* transmitidos para ligar ou desligar uma nota é reduzido de doze para nove.

HARWARE

A Fig. 1 apresenta o diagrama de blocos simplificado do sistema desenvolvido. O sistema é composto por um miniteclado e por uma pedaleira, cada um ligado em um circuito codificador próprio, e por uma CPU, constituída basicamente por um microcontrolador, por uma memória e pela saída MIDI OUT [1].

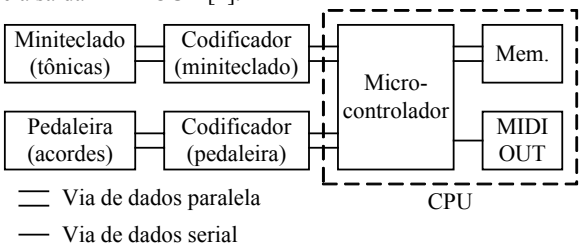

Fig. 1: Digrama de blocos do sistema proposto.

A função do miniteclado é permitir ao usuário executar a tônica de um acorde. Para isso, ele possui doze teclas (doze tônicas correspondentes à escala cromática do piano) ligadas em chaves normalmente abertas. Essas chaves são ligadas no circuito codificador do miniteclado, que gera a palavra binária de quatro *bits* ABCD, como mostra a Fig. 2.

Os acordes são executados através da pedaleira, que possui quinze chaves normalmente abertas (quinze acordes). As chaves da pedaleira também são ligadas em um circuito codificador, responsável por gerar a palavra binária de quatro *bits* EFGH que identifica a chave que foi acionada pelo músico (Fig. 3). A única diferença entre esses dois codificadores é que o da pedaleira gera dezesseis códigos binários (0000 corresponde a nenhum acorde selecionado), ao passo que o do miniteclado gera apenas treze (0000 corresponde a nenhuma tônica selecionada).

A Fig. 4 apresenta o circuito lógico combinacional do codificador da pedaleira, que é composto por quatro portas

lógicas tipo OU de oito entradas e uma saída. A estrutura do codificador do miniteclado (Fig. 2) é idêntica à do codificador da pedaleira, a não ser pelo fato de possuir apenas doze entradas. Neste trabalho, esses dois circuitos foram implementados com o circuito integrado (CI) 4078.

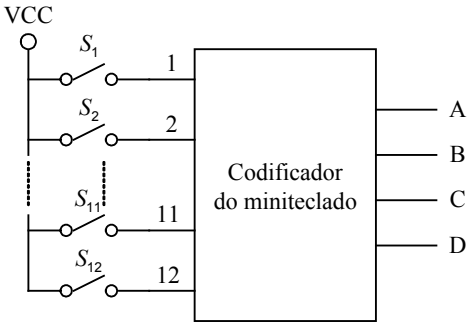

Fig. 2: Codificador do miniteclado.

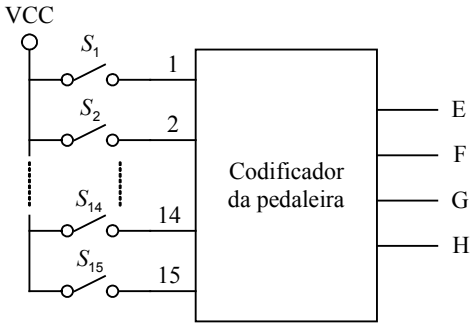

Fig. 3: Codificador da pedaleira.

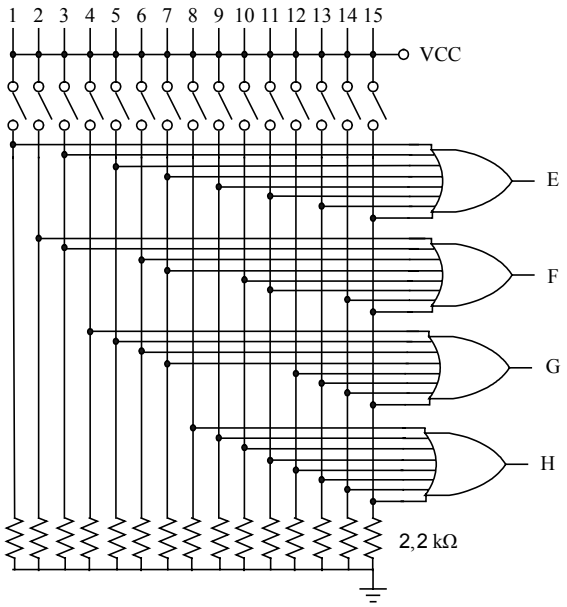

Fig. 4: Circuito lógico combinacional do codificador da pedaleira.

A Fig. 5 apresenta o diagrama esquemático da CPU. Ela é formada basicamente por um microcontrolador (CI 8031AH); por uma memória do tipo EPROM (CI 2716); por um *latch* (CI 74LS373); por um estágio de eliminação de ruído dos sinais oriundos das chaves do miniteclado e da pedaleira, composto por oito filtros passa-baixa de primeira ordem (componentes R_1 e C_1); pelo estágio de saída (MIDI OUT), que segue o padrão MIDI [1].

O microcontrolador utilizado neste projeto é o INTEL 8031AH, da família Intel MCS 51 [2]. Por simplicidade, esse microcontrolador será chamado apenas de 8031 deste ponto em diante. A escolha dentre os microcontroladores que atingiam a taxa de transmissão requerida para o projeto recaiu sobre o 8031 devido à disponibilidade do CI e do seu respectivo sistema de desenvolvimento. O 8031 possui 32 linhas de entrada/saída, memória RAM de 128 *bytes*, transmissor/receptor serial síncrono/assíncrono (UART), dois temporizadores/contadores de dezesseis *bits*, cinco fontes de interrupção com dois níveis de prioridade e um oscilador de *clock* interno. Como o 8031 não possui memória interna do tipo ROM, é necessário utilizar uma memória externa para armazenar o programa do sistema. Neste trabalho, optou-se por utilizar uma memória do tipo EPROM. Outros microcontroladores, eventualmente com menor custo, podem ser usados, especialmente aqueles que não necessitam utilizar externamente um conjunto memória + *latch*.

ACORDES ESCOLHIDOS

Por limitações práticas, é necessário limitar o número de acordes disponíveis no sistema apresentado. Como a porta 1 do 8031 (P1) possui oito *bits* e a palavra binária ABCD do codificador do miniteclado é de quatro *bits*, fica determinado que a palavra binária do codificador da pedaleira deve ser de quatro *bits* (EFGH). Com quatro *bits*, pode-se implementar dezesseis códigos binários EFGH diferentes. O valor 0000 é utilizado para indicar que nenhum acorde foi selecionado (nenhuma chave da Fig. 3 ou 4). Com isso, o número de acordes disponíveis no sistema foi limitado em quinze. Esses quinze acordes foram escolhidos com base na experiência prática. O critério utilizado foi disponibilizar os quinze acordes considerados como os mais utilizados em música popular. A Tabela 1 apresenta os acordes implementados relativos à tônica dó.

Acorde	Notas	Teclas correspondentes à tônica dó (Fig. 6)
1 – C	C2, E2, G2, C3	1, 5, 8, 13
2 – Cm	C2, Eb2, G2, C3	1, 4, 8, 13
3 - C5+	C2, E2, Ab2, C3	1, 5, 9, 13
4 - Cm5-/7	C2, Eb2,Gb2,Bb2	1, 4, 7, 11
5 - C6	C2, E2, G2, A2	1, 5, 8, 10
6 - Cm6	C2, Eb2, G2, A2	1, 4, 8, 10
7 - C7	C2, E2, G2, Bb2	1, 5, 8, 11
8 - Cm7	C2, Eb2, G2, Bb2	1, 4, 8, 11
9 - C7+	C2, E2, G2, B2	1, 5, 8, 12
10 - Cm7+	C2, Eb2, G2, B2	1, 4, 8, 12
11 - C^o	C2, Eb2, Gb2, A2	1, 4, 7, 10
12 - C5+/7	C2, Bb2, E3, Ab3	1, 11, 17, 21
13 - C6/7	C2, Bb2, E3, A3	1, 11, 17, 22
14 - C7/9	C2, Bb2, Db3, E3	1, 11, 14, 17
15 - C7/9/11	C2, Bb2, D3, F3	1, 11, 15, 18

Tabela 1: Acordes relativos à tônica dó.

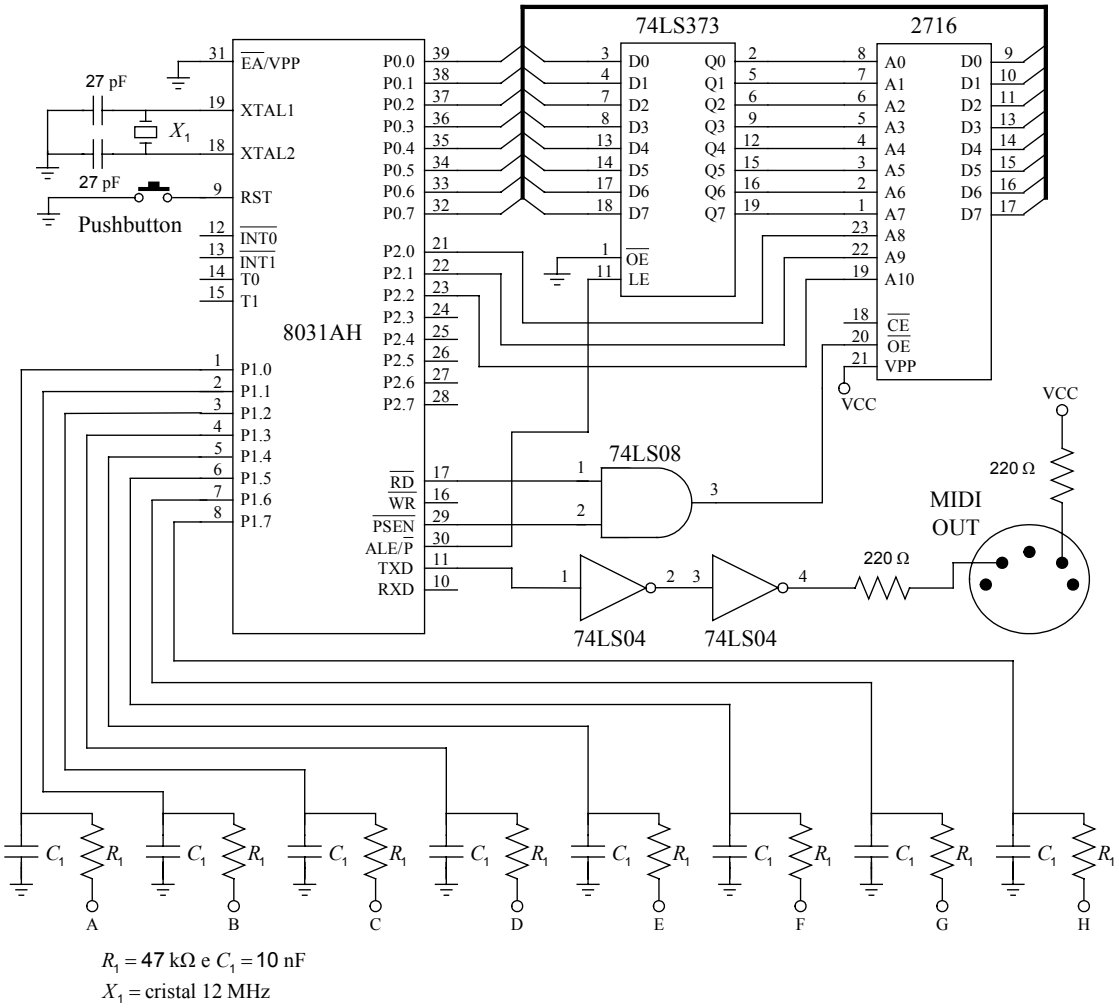

Fig. 5: Diagrama esquemático da CPU.

A Fig. 6 apresenta uma parte do lado esquerdo do teclado que é acionado pelo sistema proposto via código MIDI. A terceira coluna da Tabela 1 está relacionada com as teclas apresentadas nessa figura. Por exemplo, para tocar o primeiro acorde de dó (C), o músico aciona as teclas 1, 5, 8 e 13 do teclado da Fig. 6. A nota C2, cuja freqüência fundamental é de 65,41 Hz, é normalmente a nota relativa à tecla mais à esquerda da grande maioria dos teclados. Note que neste trabalho, nem todas as notas da Fig. 6 foram implementadas.

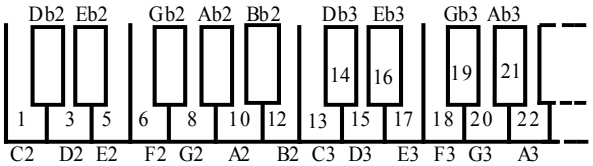

Fig. 6: Representação de uma parte de um teclado acionada pelo sistema proposto.

Cada nota da Tabela 1 possui o seu respectivo código MIDI. A Tabela A1 do Apêndice A apresenta os códigos MIDI (em decimal) das notas da região acionada do teclado da Fig. 6. Através da segunda coluna da Tabela 1, pode-se observar a existência de dezesseis notas musicais diferentes. Isso ocorre para qualquer outra tônica, não apenas para a tônica dó. Como o sistema possui doze tônicas, chega-se a um total de 192 notas, o que significa 192 códigos MIDI diferentes. Esses códigos são gravados na memória EPROM do sistema a partir do endereço *I*, definido pelo programador, como mostra a Fig. 7. O endereço *E* representa a posição de memória na qual está gravado o primeiro código MIDI para uma determinada tônica. A sua função é explicada na próxima seção.

FUNCIONAMENTO

Enquanto o usuário não aciona uma chave do miniteclado ou da pedaleira, todos os *bits* da P1 são iguais a zero e o microcontrolador permanece em estado de espera. O acionamento de uma chave do miniteclado faz com que a parte menos significativa da P1 (ABCD) se torne diferente de zero. O acionamento de uma chave da pedaleira faz com que a parte mais significativa da P1 (EFGH) se torne diferente de zero. A verificação dos valores de ABCD e EFGH é feita através de uma varredura constante da P1. Os códigos binários ABCD e EFGH expressos na base decimal serão

representados daqui em diante por T (tônica) e P (pedaleira), respectivamente. A combinação de T e P identifica qual acorde dentre os 180 implementados foi executado pelo músico. O programa então lê na memória os códigos MIDI das quatro notas do acorde. Para isso, utiliza-se a variável E da Fig. 7, que representa o endereço da posição de memória na qual está gravado o código da primeira nota (tônica T):

$$E = I + 16(T - 1)\,. \tag{4}$$

Os endereços das outras três posições de memória nas quais estão gravados os códigos das demais notas do acorde executado são determinados através da Tabela 2. Essa tabela apresenta as relações entre E e as demais posições de memória para um determinado acorde em função de P. Como exemplo, considere que o acorde executado foi o F7+. Nesse caso, $T = 6$ (tônica fá), $E = I + 80$ e $P = 9$. De acordo com a Tabela 2, os quatro endereços nos quais estão gravados os códigos MIDI das notas do acorde são: E, $E + 2$, $E + 4$ e $E + 8$. Os quatro códigos (*data bytes*) são lidos e cada um recebe o *status byte* e outro *data byte*, conforme (2) e (3), além do *start bit* e do *stop bit*. Cada uma dessas palavras MIDI é então enviada através da porta serial assíncrona do 8031 (pino 11) para a saída MIDI OUT. O equipamento no qual o sistema está conectado recebe a palavra MIDI, interpreta os dados transmitidos e executa a função correspondente.

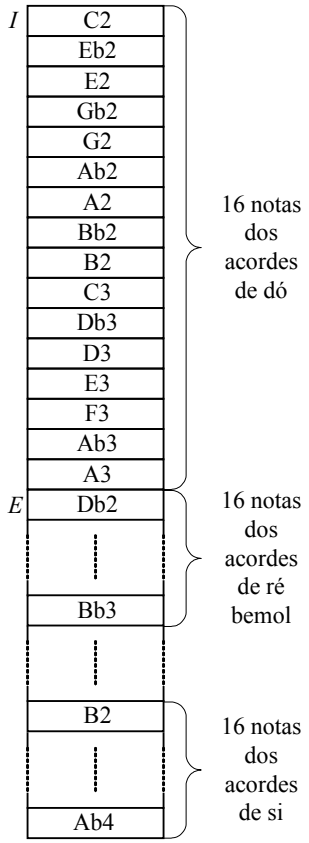

Fig. 7: Disposição dos códigos MIDI na memória EPROM.

Durante a utilização do sistema pelo músico, diferentes combinações de acionamento das chaves do miniteclado e da pedaleira podem ocorrer. Por exemplo, o usuário pode apertar primeiro uma tecla do miniteclado e depois uma chave da pedaleira ou vice-versa. Como o programa faz uma varredura constante da P1, ele precisa conhecer os estados atual e anterior de T e de P, para poder decidir o que deve ser enviado ao instrumento musical via código MIDI. A Tabela 3 apresenta todas as possíveis combinações de acionamentos das chaves do miniteclado e da pedaleira. Algumas conclusões podem ser tiradas dessa tabela. Por exemplo, se no instante atual nenhuma tecla do miniteclado for acionada ($T = 0$), nenhum som deve ser emitido pelo instrumento musical. Já se no instante atual tanto uma tecla do miniteclado quanto uma tecla da pedaleira forem acionadas ($T > 0$ e $P > 0$), o som do acorde selecionado deve ser emitido, não importando o estado anterior de T e de P. Todas as outras possíveis situações podem ser entendidas através da Tabela 3, que serviu de base para o desenvolvimento do programa escrito em linguagem ASSEMBLER.

P	Posições de memória
1	$E, E+2, E+4, E+9$
2	$E, E+1, E+4, E+9$
3	$E, E+2, E+5, E+9$
4	$E, E+1, E+3, E+7$
5	$E, E+2, E+4, E+6$
6	$E, E+1, E+4, E+6$
7	$E, E+2, E+4, E+7$
8	$E, E+1, E+4, E+7$
9	$E, E+2, E+4, E+8$
10	$E, E+1, E+4, E+8$
11	$E, E+1, E+3, E+6$
12	$E, E+7, E+12, E+14$
13	$E, E+7, E+12, E+15$
14	$E, E+7, E+10, E+12$
15	$E, E+7, E+11, E+13$

Tabela 2: Endereços de memória dos códigos MIDI das notas do acorde P de uma tônica qualquer T em função do endereço E que armazena o código MIDI da primeira nota.

Estado		Função desejada do microcontrolador	Som
Atual	Anterior		
T P	T P		
0 0	0 > 0	Continuar em *loop* de espera	Sem som
	> 0 0	Parar de tocar a tônica e entrar em *loop* de espera	
0 > 0	0 0	Continuar em *loop* de espera	
	> 0 > 0	Para de tocar o acorde e entrar em *loop* de espera	
> 0 0	0 0	Tocar a tônica escolhida	Tônica
	> 0 > 0	Continuar em *loop* de espera	Acorde
> 0 > 0	x x	Tocar o acorde escolhido	Acorde

Tabela 3: Possíveis combinações de acionamento das chaves do miniteclado e da pedaleira.

A Fig. 8 apresenta o fluxograma do programa do sistema proposto. Como já dito, P1 é a porta 1 do 8031, *T* é o valor decimal da parte menos significativa da P1 (ABCD) e *P* é o valor decimal da parte mais significativa da P1 (EFGH).

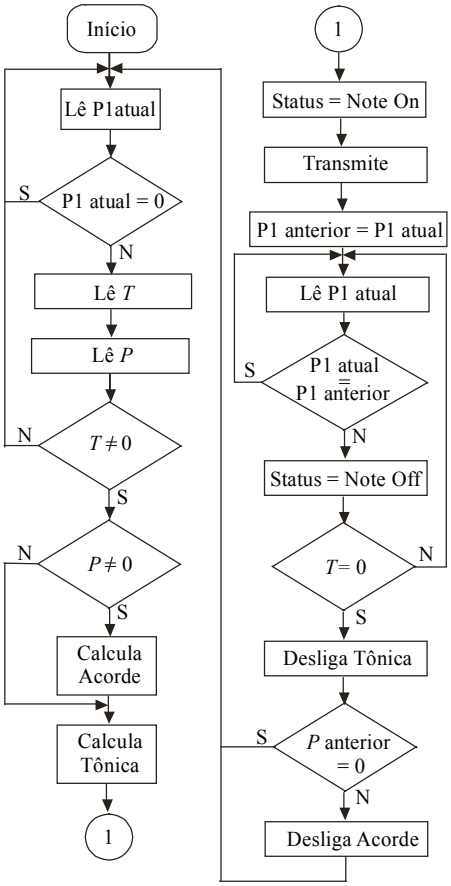

Fig. 8: Fluxograma do programa.

PROTÓTIPO CONSTRUÍDO

A Fig. 9 apresenta uma fotografia do protótipo construído. O miniteclado é mostrado na parte esquerda da figura e a pedaleira encontra-se à direita do miniteclado.

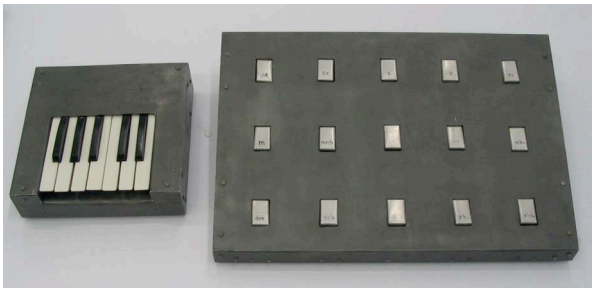

Fig. 9: Protótipo do sistema proposto.

O protótipo montado foi submetido a testes com diferentes modelos de teclados profissionais. Em todos esses experimentos, o sistema apresentou o desempenho esperado. Esses testes também mostraram que o leiaute da pedaleira (Fig. 10) é adequado para ser utilizado pelo músico.

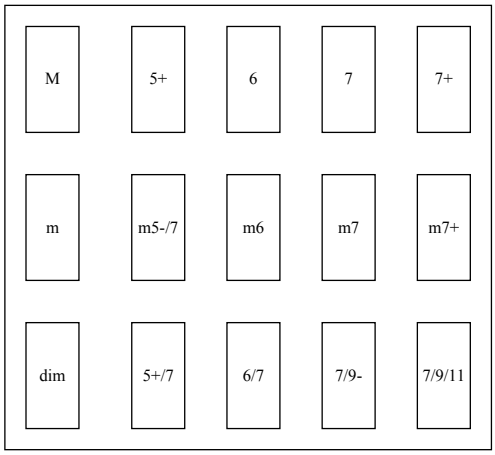

Fig. 10: Leiaute da pedaleira.

CONCLUSÕES

Este trabalho apresentou um sistema que busca auxiliar músicos com algum tipo de deficiência física na mão esquerda a tocarem teclado. O princípio exposto pode ser adaptado para atender outros tipos de deficiência física, através da alteração na disposição das chaves do sistema, e para ser utilizado com outros tipos de instrumentos musicais com entrada MIDI. Um protótipo foi construído e submetido a testes com teclados de diferentes modelos. Os testes mostraram que o protótipo apresenta o desempenho esperado e que o seu leiaute é adequado para ser utilizado pelo músico.

AGRADECIMENTO

Os autores gostariam de agradecer ao CNPq pelo apoio financeiro.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] C. Anderton, *MIDI for Musicians*. Amsco Publications, 1986.

[2] V. Pereira da Silva Jr., *Aplicações Práticas do Microcontrolador 8051*, 10ª ed. Editora Érica, 2002.

APÊNDICE A

C2	Db2	D2	Eb2	E2	F2
36	37	38	39	40	41
Gb2	G2	Ab2	A2	Bb2	B2
42	43	44	45	46	47
C3	Db3	D3	Eb3	E3	F3
48	49	50	51	52	53
Gb3	G3	Ab3	A3	Bb3	B3
54	55	56	57	58	59
C4	Db4	D4	Eb4	E4	F4
60	61	62	63	64	65
Gb4	G4	Ab4	A4	Bb4	B4
66	67	68	69	70	71

Tabela A1: Códigos MIDI (decimal) das notas da região acionada do teclado.

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na IX Convenção Nacional
11 - 13 de Abril de 2005, São Paulo, SP

Este artigo foi reproduzido do original entregue pelo autor, sem edições, correções e considerações feitas pelo comitê técnico deste evento. 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 reservados. Não é permitida a reprodução total ou parcial deste artigo sem autorização expressa da AES Brasil.

Qualidade da Energia em Sistemas de Sonorização - Harmônicos na Rede

Christian Gonçalves Herrera, Pedro F. Donoso-Garcia, Gustavo Paulinelli Guimarães
CEFET-MG UNED Divinópolis, DELT/UFMG, ISOBRASIL
Rua Monte Santo 319, Divinópolis, Minas Gerais, Brasil, CEP 35500-029
herrera_dsp@yahoo.com.br

RESUMO

Um problema que pode comprometer a fidelidade sonora de sistemas de áudio são os ruídos provenientes da rede elétrica que atingem os equipamentos e, conseqüentemente, os sinais de áudio. Estes ruídos podem ser componentes harmônicos da tensão alternada entregue pela rede, geradas a partir do efeito de cargas não lineares no barramento. Este artigo mostra que os conversores CA-CC das fontes de alimentação dos próprios amplificadores de áudio de potência podem introduzir este tipo de ruído na rede.

INTRODUÇÃO

Diversos problemas relacionados à qualidade da energia em sistemas elétricos de potência podem comprometer o funcionamento de equipamentos conectados à mesma rede elétrica. Em sistemas de áudio, problemas relacionados à distorção das formas de onda da tensão e da corrente da rede são críticos, uma vez que podem ser introduzidos ruídos no programa sonoro, o que não é desejável.

Em sistemas de sonorização de médio e grande porte alguns dos próprios equipamentos de áudio são responsáveis pela geração de harmônicos que poluem a rede. Os principais são os amplificadores de potência.

Neste trabalho são abordados os principais conceitos relativos à natureza e à geração de componentes harmônicos, assim como alguns efeitos destes no sistema e as soluções mais empregadas.

O foco de análise são os amplificadores de potência de áudio. Após a caracterização de um sistema de sonorização de médio porte, é evidenciada a relevância destes componentes na demanda total de potência do sistema e a sua contribuição na distorção das formas de onda de corrente e tensão. Resultados de medições são apresentados e discutidos, como comprovação da existência e da importância do problema.

FUNDAMENTOS TEÓRICOS

Freqüências Harmônicas

As freqüências harmônicas são múltiplos inteiros da freqüência fundamental de um sinal periódico. Isto é, para uma fundamental de 60 Hz o terceiro harmônico corresponde a 180 Hz e o quinto a 300 Hz (e assim por diante). A Figura 1 mostra uma forma de onda senoidal com terceiro e quinto harmônicos.

A Figura 2 mostra a forma de onda resultante da adição de uma componente de terceiro harmônico de 70% e uma componente de quinto harmônico de 50% à fundamental. Na prática, formas de onda distorcidas por harmônicos podem

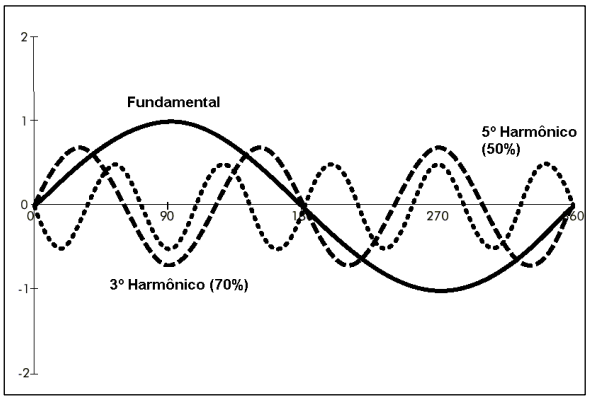

Fig. 1. Formas de onda com freqüências múltiplas da fundamental.

ser muito mais complexas do que neste exemplo, contendo outras componentes e relações de fase diferentes.

Equipamentos convencionais de medição, como multímetros de leitura não *true-RMS*, podem fornecer resultados pouco exatos. Nota-se também que existem agora seis pontos onde a onda corta o valor zero em um ciclo ao invés de somente dois. Assim, equipamentos que utilizam a passagem por zero como referência terão a confiabilidade diminuída.

Correntes com componentes harmônicos já existem no sistema elétrico há várias décadas. Inicialmente elas eram produzidas por retificadores a arco de mercúrio, utilizados na conversão CA-CC para eletrificação de estradas de ferro e em máquinas de corrente contínua com velocidade variável na indústria. Atualmente os tipos e o número de equipamentos causadores de harmônicos tem crescido muito rápido, e continuará aumentando. Dessa forma os projetistas de equipamentos que demandam precisão e linearidade devem considerar os harmônicos e seus efeitos colaterais de maneira cuidadosa.

Equipamentos de áudio de alta fidelidade estão entre os mais susceptíveis ao mau funcionamento quando expostos a interferências eletromagnéticas na forma de ruído dentro da faixa do espectro de freqüências utilizada pelos sinais de áudio (tipicamente entre 20 e 20000 Hz). Estas interferências podem atingir o programa sonoro processado e tornarem-se audíveis, o que é pouco tolerável.

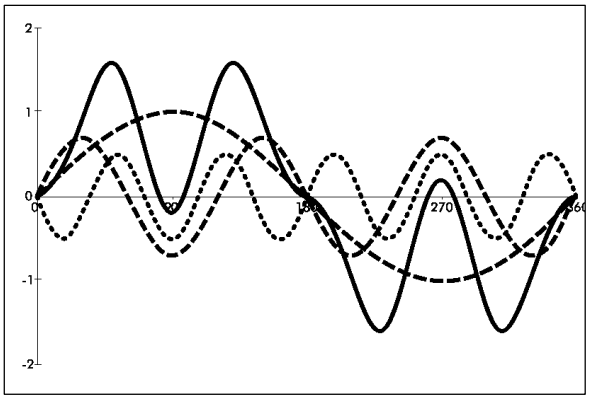

Fig. 2. Forma de onda resultante da soma das três componentes.

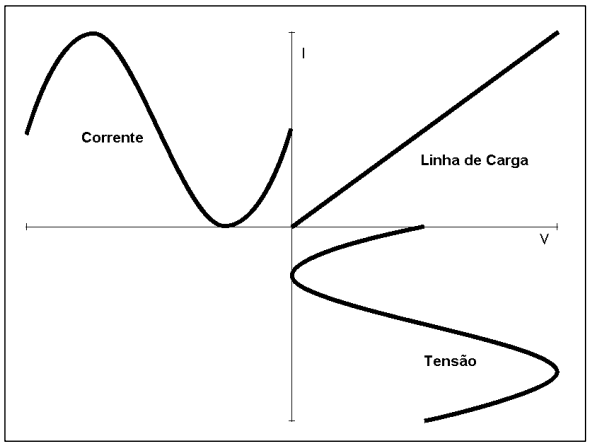

Fig. 3. Relação entre tensão e corrente num circuito formado por elementos lineares.

Como os Harmônicos são Gerados

Num sistema elétrico de potência idealmente "limpo", tanto a forma de onda da corrente como a da tensão são puramente senoidais. Na prática, correntes não-senoidais são geradas por cargas que não se relacionam linearmente com a tensão aplicada sobre elas. Em um circuito simples, que possui apenas elementos lineares - resistores, indutores e capacitores - a corrente que flui é proporcional à tensão aplicada (em uma freqüência particular). Então, se uma tensão senoidal é aplicada, uma corrente senoidal flui, como ilustrado na Figura 3. A linha de carga é o que relaciona a tensão aplicada com a corrente resultante na carga. Note que para cargas compostas por elementos reativos (capacitores e/ou indutores) pode existir um defasamento entre a tensão e a corrente mas o circuito ainda é linear.

A Figura 4 mostra uma situação onde a carga é uma ponte retificadora de onda completa e um capacitor, como num típico conversor CA-CC. Neste caso, a corrente flui somente quando a tensão suprida ultrapassa a tensão armazenada no capacitor, isto é, quando próxima ao valor de pico da tensão senoidal, como mostrado pela forma da linha de carga.

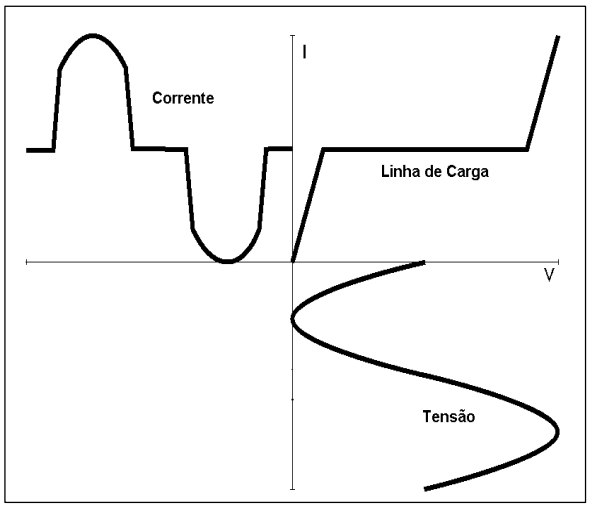

Fig. 4. Forma de onda de corrente distorcida num circuito formado por elementos não lineares.

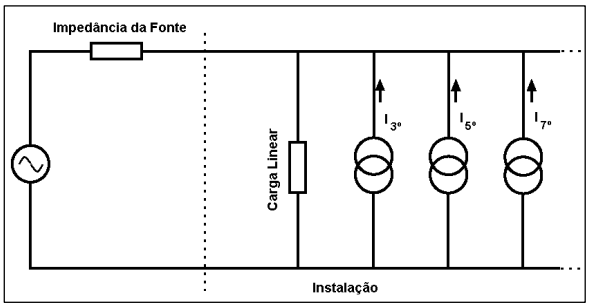

Fig. 5. Circuito equivalente de uma carga não linear.

Na prática, a linha de carga (e dessa maneira, a forma de onda da corrente) vem a ser mais complexa do que a mostrada neste exemplo, pois podem existir alguns pontos de assimetria e histerese. Sendo assim os pontos de variação e as inclinações podem variar com a mudança da corrente suprida à carga.

Qualquer forma de onda periódica pode ser decomposta em senóides com uma freqüência fundamental somada a um certo leque de harmônicos, cada um com amplitude e fase relacionadas com a forma de onda original. Para uma onda simétrica (onde o ciclo negativo e o positivo têm a mesma forma e amplitude) as harmônicas pares não existem. Os harmônicos pares são relativamente raros, mas aparecem em situações onde a retificação de meia onda é utilizada.

O circuito equivalente para uma carga não-linear é mostrado na Figura 5. Pode ser modelado como uma carga linear em paralelo com um certo número de fontes de corrente, sendo cada fonte correspondente a uma freqüência harmônica.

As correntes harmônicas geradas pela carga têm que circular através do circuito pela impedância da fonte e por todos os outros caminhos em paralelo. Como resultado, tensões harmônicas aparecem através da impedância da fonte e ficarão presentes ao longo da instalação elétrica.

A distorção harmônica da tensão resultante de uma corrente harmônica é proporcional à razão entre a impedância de fonte e a impedância da carga.

Problemas Causados Por Correntes Harmônicas

Correntes harmônicas causam problemas tanto para o sistema de distribuição de energia quanto para as cargas instaladas no barramento. Os efeitos e as soluções são bem diferentes e precisam ser tratados separadamente. As medidas apropriadas para controlar os efeitos de harmônicos dentro da instalação podem não necessariamente reduzir a distorção causada na rede elétrica, e vice versa.

Sobreaquecimento do Condutor Neutro

Num sistema trifásico a forma de onda da tensão de cada fase é defasada em 120º com relação ao neutro. Assim, quando as cargas são equilibradas, a soma das correntes no neutro é zero. Mesmo quando as cargas não são equilibradas, a corrente que flui pelo neutro é pequena, fato este explorado pelos projetistas de instalações, que costumam especificar condutores com capacidade menor de condução de corrente. No entanto, embora as componentes fundamentais de corrente se cancelem no neutro, as correntes de freqüências harmônicas podem não se cancelar. De fato, aquelas múltiplas ímpares de três vezes a freqüência fundamental (3º, 9º, 15º harmônicos) se somam no neutro. A Figura 6 ilustra este efeito. Neste diagrama, as correntes de fase, mostradas no topo da figura, são introduzidas com um defasamento de 120º. Os terceiro harmônicos de todas as fases são idênticos, com amplitude de 70% da fundamental, resultando numa corrente de neutro com amplitude mais de duas vezes maior que o valor da amplitude da corrente fundamental.

Uma solução de caráter paleativo para este efeito é o sobredimensionamento do condutor do neutro, o que implica no aumento do custo da instalação.

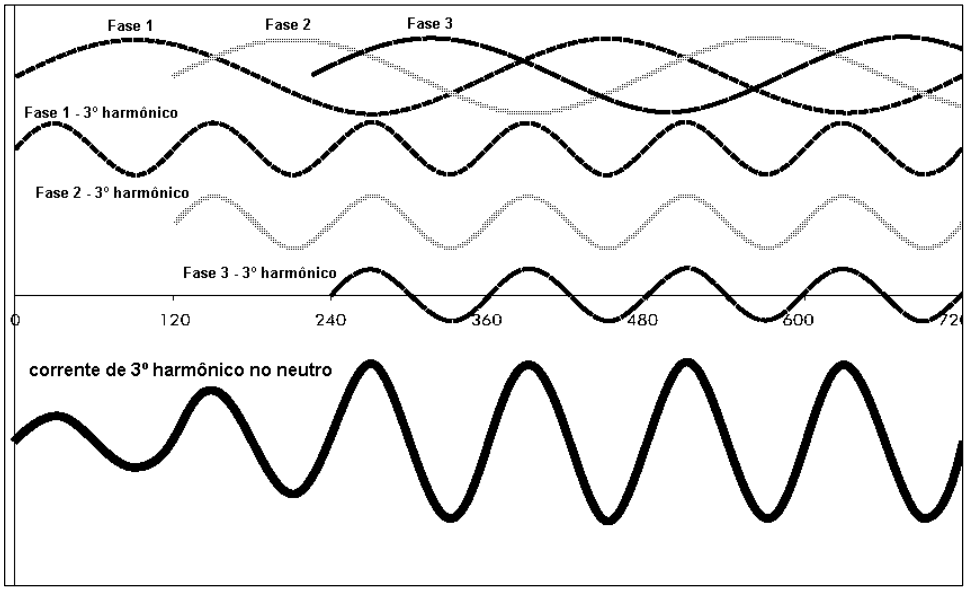

Fig. 6. Correntes de 3º harmônico se somam no neutro.

Efeitos em transformadores

Transformadores são afetados pelas harmônicas no que diz respeito às correntes de magnetização, que representam cerca de 10% das perdas do transformador a carga plena e crescem juntamente com a freqüência da componente harmônica. O resultado é o sobreaquecimento e a perda da vida útil do transformador. Embora nem todos transformadores operem a carga plena, este efeito deve ser levado em conta na especificação destes dispositivos.

Sobrecorrente em capacitores de correção de fator de potência

Capacitores são utilizados para correção do fator de potência em sistemas com cargas indutivas como motores de indução e reatores.

A Figura 7 mostra um circuito equivalente onde um capacitor é empregado para a correção do fator de potência. A impedância destes capacitores diminui com o aumento da freqüência, enquanto que a impedância da fonte, que geralmente é indutiva, aumenta. O capacitor é levado então a conduzir correntes harmônicas bem mais elevadas do que o valor nominal para o qual foi dimensionado, resultando em possíveis danificações nestes componentes.

Um problema mais sério é no caso do capacitor juntamente com a indutância da fonte entrarem em ressonância numa das freqüências harmônicas. Isto pode ser evitado adicionando-se uma indutância em série com o capacitor tal que a combinação seja puramente indutiva na freqüência do harmônico menos significante. Esta solução também limita as correntes harmônicas no capacitor. O tamanho físico do indutor pode ser um problema, especialmente quando harmônicos de baixa ordem estiverem presentes.

Efeito Skin

Correntes alternadas tendem a fluir na superfície externa do condutor. Este é o chamado efeito *skin* e é mais pronunciado em altas freqüências. O efeito *skin* normalmente é ignorado porque tem pouca influência na freqüência da rede elétrica (60 Hz), mas acima de 350 Hz, isto é a partir do 7º harmônico em diante, ele se torna mais significante, causando perdas adicionais e aquecimento dos condutores. Quando correntes harmônicas estão presentes, os projetistas devem se preocupar com o efeito *skin* e redimensionar os cabos apropriadamente.

Problemas Causados Por Tensões Harmônicas

Correntes harmônicas resultam em distorção na forma de onda da tensão quando a impedância da fonte de alimentação não é nula. Existem dois elementos que conformam esta

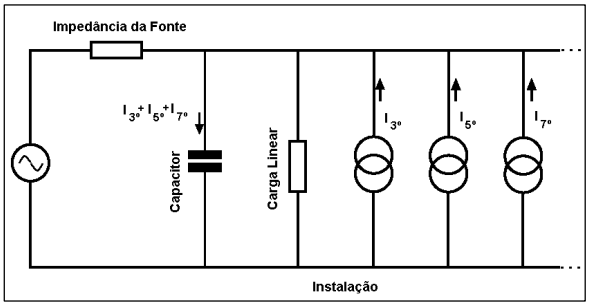

Fig. 7. Utilização de capacitor para correção do fator de potência.

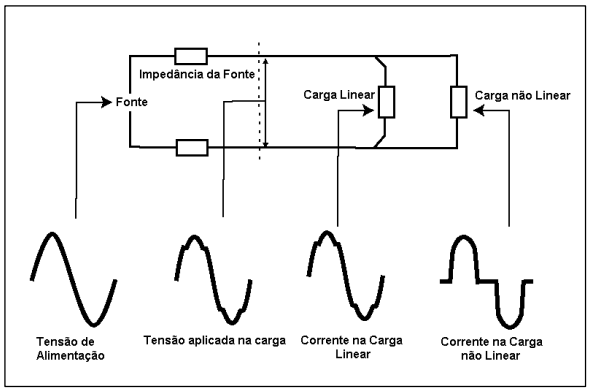

Fig. 8. Distorção da forma de onda da tensão causada por uma carga não linear.

impedância: aquele referente aos cabos de transmissão da energia desde o barramento até as cargas, e o referente à impedância interna dos geradores ou transformadores da rede de distribuição. A Figura 8 ilustra esta situação.

A solução é separar os circuitos que alimentam as cargas geradoras de harmônicos daquelas que são sensíveis a estes harmônicos, assim a tensão distorcida causada pelas cargas não lineares não afetam as cargas sensíveis.

Na instalação de transformadores, devem ser selecionados aqueles que apresentam impedância de saída extremamente baixa ou que tenham capacidade de suportar o aquecimento adicional; em outras palavras, transformadores sobredimensionados.

Soluções para o Problema de Correntes Harmônicas

As medidas disponíveis para o controle na magnitude das correntes harmônicas podem ser classificadas em dois grupos: filtros passivos e filtros ativos.

Filtros passivos

Filtros passivos são usados para prover um caminho de baixa impedância (*shunt*) para correntes harmônicas. Assim estas correntes fluem pelo filtro e não pela fonte de alimentação. O filtro pode ser projetado para uma única freqüência de harmônico ou para uma ampla faixa de freqüências, dependendo das necessidades.

Em alguns casos é necessário projetar um filtro mais complexo para aumentar a impedância em série nas freqüências harmônicas e então reduzir a proporção de corrente que flui de volta para a fonte (Figura 9). Filtros rejeita banda são comumente propostos, instalados no neutro ou na fase. Um filtro em série é adicionado mais para bloquear as correntes harmônicas do que para prover um caminho controlado para elas. Como pode haver queda de tensão significativa no filtro em série, este deve ser empregado com cautela.

Filtros ativos

Os filtros passivos são soluções eficazes somente para freqüências particulares de harmônicos. Em algumas instalações o conteúdo harmônico da corrente varia ao longo do tempo. Uma solução conveniente nestes casos é o emprego de filtros ativos.

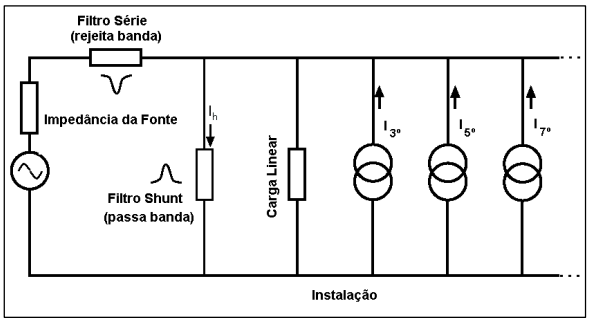

Fig. 9. Filtros série e shunt de harmônicos passivo.

Como mostrado na Figura 10, o filtro ativo é um elemento *shunt*. Um transformador de corrente amostra o conteúdo harmônico da corrente circulando pela carga e comanda um gerador de corrente que produz uma réplica exata que é introduzida na linha no próximo ciclo. Uma vez que as correntes harmônicas são fornecidas pelo filtro ativo, apenas a corrente fundamental é drenada da fonte de alimentação. Na prática, correntes harmônicas são reduzidas em até 90% de sua magnitude e, consequentemente, a distorção na forma de onda de tensão também. O alto custo desta solução pode tornar proibitivo o seu uso.

SISTEMAS DE SONORIZAÇÃO

As freqüências comumente encontradas na composição de correntes distorcidas por cargas não lineares estão dentro do espectro de freqüências audíveis. Trata-se de um grande problema para sistemas de gravação e sonorização, pois estas freqüências podem atingir os circuitos que processam e amplificam os sinais de áudio, principalmente porque estes são sinais de tensão de baixa magnitude e seus circuitos de amplificação envolvem elevados ganhos de tensão (20-80 dB).

Estúdios de gravação são projetados para evitar este tipo de problema. Alguns cuidados podem ser tomados como a instalação de transformadores isoladores na entrada de energia do estúdio. Também as fontes de alimentação internas dos próprios equipamentos contém filtros para atenuar a entrada de ruído harmônico proveniente da rede elétrica. Topologias de circuitos eletrônicos de processamento de sinais com alta rejeição a ruídos provenientes da fonte de alimentação são cada vez mais utilizados pelos projetistas dos equipamentos.

No caso de sistemas de sonorização o problema é mais

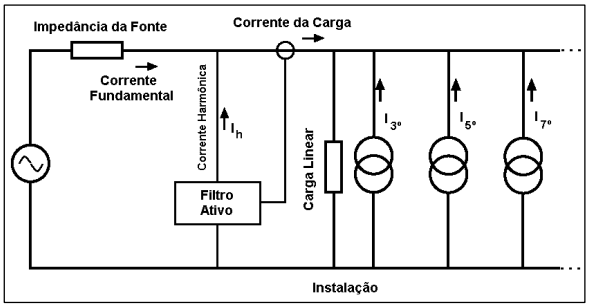

Fig. 10. Filtro ativo de harmônicos.

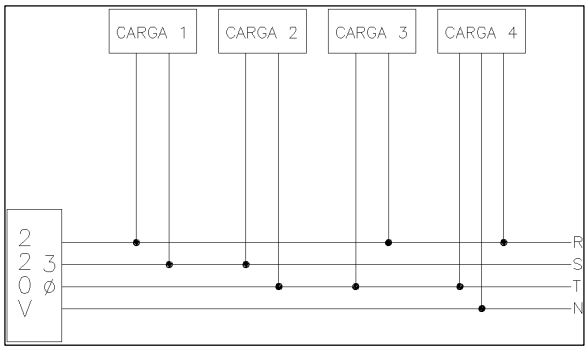

Fig. 11. Configuração típica de um sistema de sonorização de médio porte.

grave: os amplificadores de potência utilizados para o reforço sonoro demandam alta potência do barramento, que normalmente alimenta todo o sistema, incluindo equipamentos sensíveis a harmônicos como pré amplificadores de microfones e processadores de pequenos sinais de áudio.

Na grande maioria dos casos os amplificadores de potência são alimentados por conversores CA-CC que retiram a corrente alternada da rede e entregam corrente contínua para o circuito amplificador. Estes conversores são vistos pelo barramento como cargas extremamente não lineares, pois constituem-se de transformadores abaixadores de tensão e uma ponte retificadora de onda completa filtrada por capacitores. A demanda de potência dos amplificadores é relativamente alta e varia de acordo com o programa musical, de onde conclui-se que a magnitude instantânea das correntes harmônicas é dependente do nível do sinal de áudio a ser amplificado.

A contaminação dos equipamentos pelas freqüências harmônicas da corrente e da tensão é percebida auditivamente como um ruído nos alto-falantes que, misturado ao sinal de áudio, pode tornar a execução sonora desconfortável para músicos, locutores e ouvintes.

Um exemplo típico de *layout* de um sistema de sonorização de médio porte (público de até 2000 pessoas) é mostrado na Figura 11. Temos conectadas ao barramento trifásico (220 V fase-fase) três tipos de carga diferentes, descritas pela Tabela 1.

A energia elétrica é transmitida do ponto comum (fornecido pela concessionária local) até o local da sonorização através de cabos relativamente longos. A impedância destes cabos influencia decisivamente na introdução de distorção harmônica na forma de onda da tensão que será entregue aos equipamentos do sistema.

As cargas 1 e 2 são compostas por grupos idênticos de amplificadores de potência destinados ao acionamento das caixas acústicas de endereçamento de som ao público (*public address*). São as cargas de maior potência e as maiores geradoras de harmônicas para o sistema. A carga 3 é composta pelo conjunto de amplificadores de potência responsáveis pelas caixas acústicas que fornecem o monitoramento do som para os músicos no palco. Consomem aproximadamente 2/3 da potência das cargas 1 e 2 sendo consideradas também grandes geradoras de harmônicos. A carga 4 é constituída pelos equipamentos de baixa potência que, por serem equipamentos que tratam com sinais de áudio

de baixa amplitude, normalmente mesas misturadoras, equalizadores, compressores e processadores de efeito digitais, são os mais sensíveis aos ruídos introduzidos no barramento pelas outras três cargas.

É importante notar a distribuição desequilibrada das cargas no barramento, o que leva a problemas de sobrecorrente no condutor neutro, como discutido anteriormente.

Conclui-se que os sistemas de sonorização de médio e grande porte estão susceptíveis a problemas relacionados a correntes harmônicas gerados por cargas não lineares. Resta quantificar o conteúdo harmônico das correntes envolvidas nas cargas de maior potência, a influência destas correntes na distorção da forma de onda da tensão do sistema e determinar as conseqüências de uma tensão distorcida em equipamentos sensíveis a estes harmônicos.

MEDIÇÃO E ANÁLISE DE CORRENTES E TENSÕES DE ALIMENTAÇÃO DE UM AMPLIFICADOR

São apresentados e discutidos os resultados de medições de formas de onda de corrente e tensão em um típico amplificador de potência e a influência do programa musical na magnitude da potência consumida por este equipamento.

O amplificador ensaiado foi o Nashville NA1600, que consome 660 W quando alimentado por 127 V (dados especificados pelo fabricante).

O sistema de medição foi composto por um osciloscópio digital portátil do fabricante Tektronix modelo THS 720, acompanhado das pontas de prova de tensão e corrente. Os dados adquiridos pelo osciloscópio foram transferidos para o microcomputador e analisados pelo *software* Tektronix WaveStar, versão 1.3.

A Figura 13 mostra a forma de onda da tensão fase-neutro no barramento quando nenhuma carga está conectada à alimentação. A taxa de distorção harmônica (relação entre a soma da amplitude de todas componentes da forma de onda e a amplitude apenas da fundamental) é de 2,76%, determinada principalmente pelos três primeiros harmônicos ímpares. Esta medição serve como uma referência para a análise do sistema a carga plena.

Na Figura 14 é mostrada a forma de onda da corrente consumida pelo amplificador ensaiado em repouso, isto é, sem nenhum sinal de áudio aplicado à sua entrada. Trata-se de uma forma de onda complexa, com grande conteúdo harmônico. O quinto harmônico é o de maior magnitude (52%), seguido pelo terceiro (42%). Os demais componentes apresentam magnitude inferior a 25%. Esta forma de onda mostra que o equipamento medido possui duas etapas de retificação no secundário do transformador de entrada, recebendo tensões defasadas entre si.

Quando é aplicado um sinal (ruído rosa) na entrada do amplificador a máxima potência, tem-se o consumo máximo de corrente. Sua forma de onda pode ser vista na Figura 15, juntamente com o conteúdo harmônico. Trata-se do pior caso, onde o terceiro harmônico chega a representar 68% da magnitude da freqüência fundamental. Percebe-se aqui que o conteúdo harmônico da corrente varia com a potência consumida pelo equipamento, o que dificulta a eliminação destas correntes com a utilização de filtro passivos. A Figura 12 ilustra a modulação na corrente consumida pelo amplificador causada pela variação na amplitude do sinal de áudio (programa musical).

Na Figura 16 é analisada a forma de onda da tensão quando o amplificador trabalha com carga máxima. Em comparação com a forma de onda da Figura 13, vemos que a DHT subiu para 2,78%, um aumento de 0,12%. A queda de tensão foi de aproximadamente 1 V, dado que reflete a regulação da fonte de alimentação. Embora sejam diferenças pequenas, é importante frisar que num sistema de médio/grande porte o impacto é muito maior e, dependendo das especificações de cada elemento do sistema (linhas de transmissão de energia, blindagem de equipamentos contra interferências da rede), as correntes harmônicas geradas pelos amplificadores de potência podem efetivamente prejudicar o funcionamento dos equipamentos conectados ao barramento.

TABELA 1

CARGAS NUM SISTEMA DE SONORIZAÇÃO

CARGA 1 e 2			*(Public Adress – PA)*
Quant.	Tipo	Modelo	Potência [W]
2	Amplificador de Potência	TIP 3000	4500
2	Amplificador de Potência	TIP 2000	3000
2	Amplificador de Potência	TIP 800	1760
		Potência Total [W]	**18520**
CARGA 3			*(Monitores de Palco)*
Quant.	Tipo	Modelo	Potência [W]
2	Amplificador de Potência	NA 1600	660
1	Amplificador de Potência	TIP 1000	1760
2	Amplificador de Potência	TIP 1200	2200
1	Amplificador de Potência	TIP 1500	2640
1	Amplificador de Potência	TIP 2000	3000
		Potência Total [W]	**13120**
CARGA 4			*(Processadores e Periféricos)*
Quant.	Tipo	Modelo	Potência [W]
6	Equalizadores	vários	25
6	Compressores/Gates	vários	30
4	Processadores Digitais	vários	50
2	Misturadores de Sinais	vários	90
		Potência Total [W]	**710**

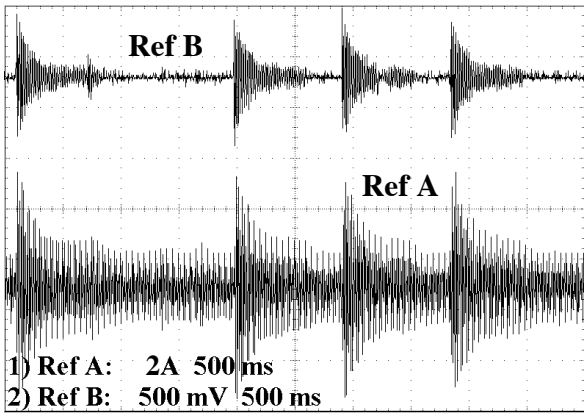

Fig. 12. Sinal de áudio (acima) modulando a corrente de alimentação do amplificador (baixo).

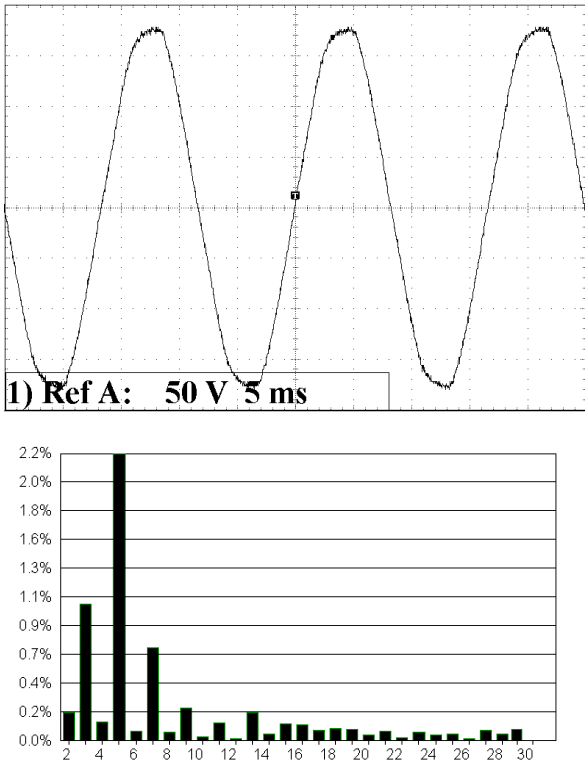

Fig. 13. Forma de onda e conteúdo harmônico da tensão do barramento sem carga.

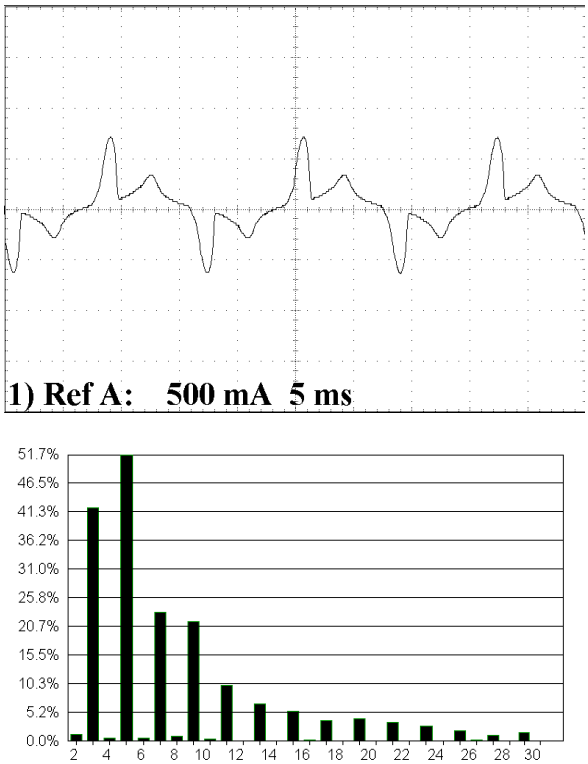

Fig. 14. Forma de onda e conteúdo harmônico da corrente de repouso.

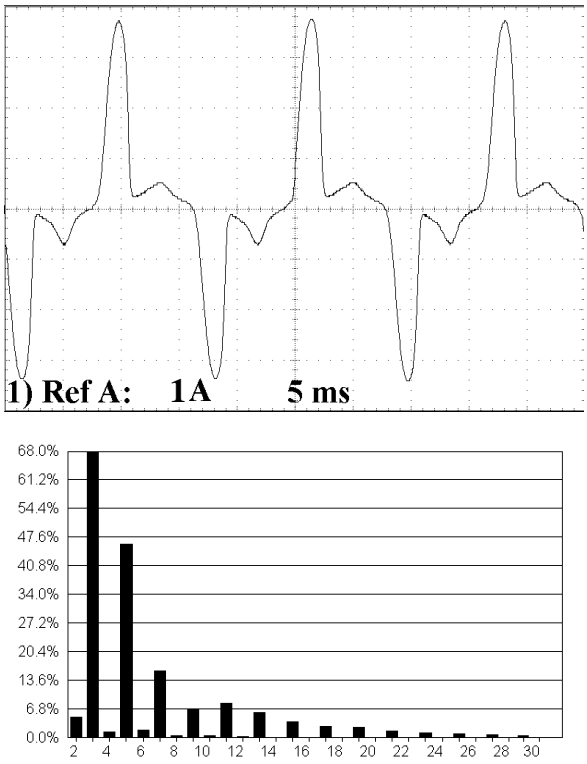

Fig. 15. Forma de onda e conteúdo harmônico da corrente a carga máxima.

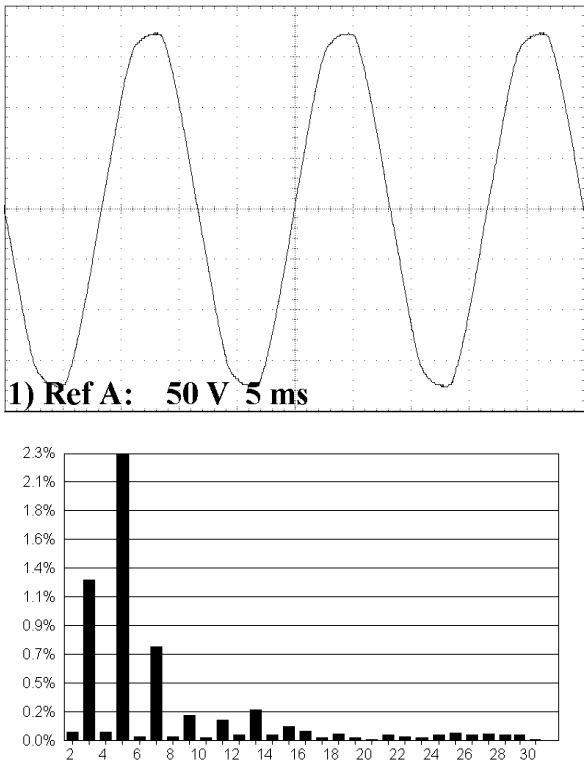

Fig. 16. Forma de onda e conteúdo harmônico da tensão a carga máxima.

CONCLUSÃO

Em sistemas de áudio, componentes harmônicos na rede elétrica dos equipamentos podem, além de comprometer o seu funcionamento, introduzir ruídos audíveis no programa sonoro, prejudicando a qualidade da reprodução sonora.

Analisando um sistema de sonorização de médio porte, conclui-se que os principais geradores de harmônicos são os amplificadores de potência, uma vez que incorporam conversores CA-CC, que são cargas não lineares.

Os ensaios realizados num amplificador de potência comprovam a geração de correntes harmônicas significativas por estes equipamentos e, conseqüentemente, a possibilidade de introdução de tensões com distorção harmônica no barramento de energia elétrica.

Como sugestão para o complemento deste estudo, uma análise da influência de tensões de alimentação distorcidas sobre o sinal de áudio nos equipamentos pode ser feita. Esta análise buscaria identificar se as harmônicas são transmitidas para o sinal de áudio diretamente através da filtragem inadequada da fonte de alimentação interna dos equipamentos, ou por irradiação eletromagnética dada a proximidade dos cabos de energia dos cabos que transmitem sinais de áudio e a alta impedância dos circuitos de entrada dos equipamentos.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] DUGAN, R. C., *Electrical Power Systems Quality*. McGraw Hill, USA, 1996.

[2] CHAPMAN, D., *Harmonics - Causes and Effects*. Copper Development Association. London, March 2001.

[3] GREBE, T. E. *Solving Harmonics Problems in Industrial Plants and Harmonic Mitigation Techniques for Adjustable-Speed Drives*. Proceedings of Electrotech, Montreal, 1992.

[4] IEEE Standard 519-1992, IEEE Recommended Practice and Requirements for Harmonic Control in Electrical Power Systems. Piscataway, N. J., 1992.

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na IX Convenção Nacional
11 – 13 de Abril de 2005, São Paulo, SP

Este artigo foi reproduzido do original entregue pelo autor, sem edições, correções e considerações feitas pelo comitê técnico deste evento. 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 reservados. Não é permitida a reprodução total ou parcial deste artigo sem autorização expressa da AES Brasil.

Obtenção de Marcas de Pitch em Sinais de Voz para Síntese por Concatenação Temporal

Vagner L. Latsch e Sergio L. Netto
Programa de Engenharia Elétrica, COPPE/UFRJ
Código Postal 68503, Rio de Janeiro, RJ, 21941-972, Brasil
latsch@lps.ufrj.br sergioln@lps.ufrj.br

RESUMO

Neste artigo é proposto um método para obtenção das marcas de pitch em segmentos de sinais de voz a serem concatenados pelo algoritmo TD-PSOLA. O sistema proposto utiliza a captação de um sinal auxiliar, através de um microfone de contato, para obter informações mais intensas a respeito da atividade das cordas vocais. Isto contribuiu para melhorar o desempenho de uma detecção automática de marcas. Os resultados se mostraram promissores, inclusive para casos críticos de detecção, onde se mostrou necessária pouca ou nenhuma correção manual.

INTRODUÇÃO

Os conversores texto-fala, também conhecidos como TTS (text-to-speech), são sistemas que produzem fala sintética correspondente à leitura de um texto. De modo geral, estes sistemas se baseiam na concatenação de unidades sonoras que são devidamente processadas para se incorporar a entoação desejada à voz sintetizada. Um dos algoritmos mais comumente utilizados para este processamento, devido ao seu baixo custo computacional, é o chamado TD-PSOLA (Time Domain - Pitch Synchronous Overlap and Add). Tal algoritmo requer a detecção das chamadas marcas de pitch que indicam o instante de tempo do fechamento da glote em um dado sinal de voz. Este processo como um todo é bastante intenso e consistirá no foco do presente artigo.

Neste contexto, após uma breve introdução ao problema de conversão TTS, faremos uma apresentação do algoritmo TD-PSOLA. Mostraremos, então, a necessidade da obtenção das marcas de pitch, ilustrando claramente este conceito e a dificuldade de detectá-las de forma precisa. Será vista, em seguida, uma metodologia de obtenção das marcas utilizando-se de um sinal auxiliar obtido por um microfone de contato colocado junto à garganta do locutor. Mostraremos por fim como este sinal auxiliar facilita a implementação de todo o processo e torna o resultado obtido muito preciso e confiável. Ao fim, concluímos o artigo apresentando os principais resultados obtidos.

CONVERSÃO TEXTO-FALA IRRESTRITA

Sistemas TTS têm aplicação em diferentes áreas, como por exemplo, na consulta de *emails* por telefone, na leitura de menus e orientações em centros de atendimento automático e até mesmo no auxílio a deficientes visuais em diferentes áreas.

Em alguns casos, como nos serviços de auxilio às listas telefônicas onde se requer a pronúncia automática dos números telefônicos, o "texto" a ser convertido está limitado aos algarismos de 0 a 9. Neste caso, para gerar a fala sintética correspondente à pronúncia do número de telefone desejado, é intuitivo sugerir a concatenação de segmentos de voz contendo a pronúncia dos algarismos.

Já os demais exemplos acima citados de aplicações de sistemas TTS tratam da conversão de texto *irrestrito*, onde o texto a ser convertido não está limitado a um conjunto de palavras ou frases. Em tais casos, a concatenação pura e simples de palavras se torna impraticável se considerarmos a quantidade de palavras existentes e as suas variantes. Baseado em conceitos fonéticos, que tratam de mapear os diversos sons existentes em uma língua (fones), e fonológicos, que observam a organização destes sons construindo significado, tem-se a proposta do uso de unidades menores de concatenação, limitadas pela sílaba, pelos próprios fones, ou unidades intermediárias.

Assim, depois de definido o tipo de unidade a ser usada e quais serão necessárias para gerar as palavras e frases de uma língua em um sistema TTS, é feita a coleta destas unidades a partir de sinais de fala previamente gravados, compondo um banco de unidades [1].

Deste modo, um sistema de conversão de texto irrestrito, inicialmente fará a conversão dos caracteres do texto (grafemas) em unidades fonológicas (fonemas), e em seguida irá obter do banco as unidades necessárias para gerar a seqüência dada e concatená-las gerando o sinal de fala sintético correspondente.

É preciso levar em conta que as unidades de concatenação, sejam elas quais forem, estão sujeitas à variação, de acordo com a posição ocupada dentro de uma frase ou com a entoação aplicada. Por exemplo, no caso dos números de telefone, considerando cada algarismo uma unidade, a pronúncia do algarismo 2 na seqüência numérica **2**555-5555 será diferente na seqüência 5555-555**2**. Assim, para obter uma entoação correta e natural, seria necessário armazenar todas as variantes de uma unidade ou então usar um método capaz de modificá-las principalmente em intensidade, duração e freqüência fundamental, que são os principais fatores para caracterizar a entoação ou prosódia.

Um dos métodos mais populares, que tem sido usado em diversos sistemas TTS atuais, devido à sua simplicidade de implementação e ao reduzido custo computacional, é algoritmo TD-PSOLA descrito brevemente a seguir.

ALGORITMO TD-PSOLA

O algoritmo TD-PSOLA [2] é baseado na técnica de *overlap and add* (OLA), na qual um sinal periódico com diferente escala temporal e/ou pitch é reconstruído através da aplicação janelas síncronas com o pitch. Estas janelas são aplicadas ao sinal, centradas em *marcas de pitch* com largura típica de dois períodos, e são alongadas ou encurtadas, removidas ou repetidas para obter o sinal modificado [3].

A qualidade oferecida pelo TD-PSOLA no contexto de síntese por cópia é perto da perfeição [3]. Porém, quando o algoritmo é utilizado na modificação de segmentos concatenados, provenientes de outros contextos, se as marcas não são posicionadas de forma consistente, o resultado são erros de fase na superposição das janelas, principalmente na vizinhança de concatenação [3].

Para a determinação destas marcas, podem ser usados algoritmos de detecção de pitch chamados de PDAs (*Pitch Detection Algoritms*) [4]. Estes algoritmos podem ser divididos em duas categorias: PDAs no domínio do tempo e PDAs por análise em termo curto. Os PDAs no domínio do tempo oferecem a estimativa do pitch período a período, mas são sensíveis as degradações do sinal na janela de análise. PDAs de termo curto, por outro lado, são mais robustos mas oferecem uma estimativa do pitch médio ao longo de um número de períodos, isto porque o método conta com a similaridade do sinal de voz entre períodos de pitch adjacentes. Assim, se vários períodos de pitch estão contidos em um segmento de análise, o valor do pitch estimado é um valor médio para todo o segmento [5].

Idealmente as marcas de pitch deveriam ser introduzidas nos trechos sonoros na posição de um evento específico no ciclo de pitch e nos trechos surdos regularmente espaçadas [3]. Um evento no ciclo de pitch muito utilizado é o instante de fechamento glotal (GCI, g*lotal closure instant*), ponto onde ocorre a maior excitação do trato vocal. Porém, a detecção precisa destes instantes diretamente do sinal de voz, seja de forma automática ou mesmo de forma visual, apresenta uma enorme dificuldade significando um grande consumo de tempo [3].

Uma solução possível para a detecção precisa do instante de fechamento glotal é o uso de um equipamento chamado *eletroglotógrafo* (EGG) que mede a atividade das cordas vocais [6]. Este equipamento, porém, apresenta um custo elevado, da ordem de alguns milhares de dólares. Uma alternativa de baixo custo é vista a seguir.

MICROFONE DE CONTATO

Alguns sistemas de aquisição de voz em ambientes extremamente ruidosos têm usado microfones em contato com o pescoço, chamado de *throat microphone*, por apresentar reduzida captação de ruído ambiente. Alguns autores [7] têm proposto a utilização deste tipo de microfone para melhorar o desempenho de sistemas de reconhecimento em ambientes ruidosos. Em [8] os autores utilizaram um acelerômetro em contato com a pele, na altura da glote, e observaram que o sinal captado, quando comparado com o sinal de um EGG, representa o som gerado pela vibração das cordas vocais.

Neste trabalho propõe-se o uso de um microfone de contato para a captação da vibração da glote com o objetivo de obter os GCIs de forma barata e precisa. O "microfone" utilizado trata-se de um disco piezoelétrico cerâmico, ilustrado na Fig. 1, geralmente utilizado como captador em instrumentos musicais acústicos, como violão, violino etc.

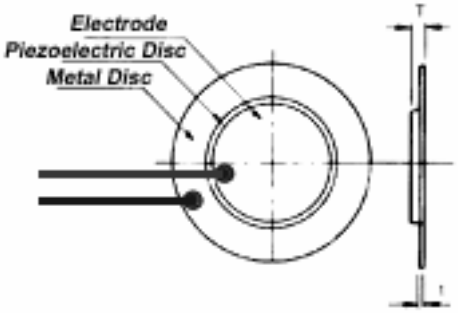

Fig. 1: Formato de um disco piezoelétrico.

O disco de metal, mostrado na Fig. 1, é colado a uma base plástica somente pelas bordas, de maneira que o centro fique livre. Esta base plástica é então fixada a uma fita de velcro, formando um colar, mostrado na Fig. 2. Este colar é colocado ao redor do pescoço de maneira que o disco piezoelétrico

fique localizado na região frontal do pescoço, o mais baixo possível, conforme mostrado na Fig. 2. O ajuste de pressão do colar (apertado ou frouxo), esta diretamente associado à qualidade do sinal, portanto a pressão ideal é aquela em que o colar fique o mais justo possível, sem causar grande desconforto.

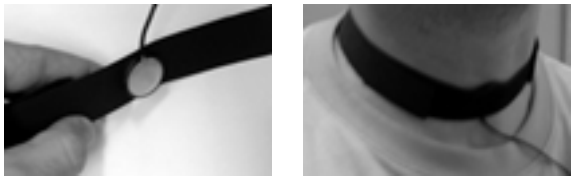

Fig. 2: Disco piezoelétrico fixado ao colar de velcro e colocação do colar na base do pescoço.

Além do sinal do microfone de contato, o sinal de voz precisa ser captado em simultâneo, no entanto, a maioria das placas de som não possui entrada para dois microfones em simultâneo (estéreo). Uma solução é utilizar a entrada *line-in* da placa de som, que pode ser utilizada em modo estéreo, porém é necessário um pré-amplificador para os microfones. Deste modo, foram montados dois pré-amplificadores, conforme a nota referenciada em [9]. Os sinais obtidos em simultâneo pelos dois microfones se mostram defasados de acordo com a distância entre os microfones, deste modo para manter esta distância fixa ao longo da gravação, foi usado um microfone acoplado aos fones de ouvido e próximo a boca. Deste modo, o atraso dependerá principalmente das características físicas do usuário. Em média este atraso equivale ao tempo de propagação para o som percorrer uma distância típica de 20 cm variando em +/- 5 cm. Deste modo, temos um atraso no sinal do microfone convencional da ordem de (0,6 +/- 0,15) ms. Os sinais obtidos foram amostrados na freqüência de 22050 Hz o que resulta em um atraso em torno de 12 amostras. Na Fig. 3 é mostrado no gráfico superior o sinal obtido pelo microfone convencional, onde foi compensado o atraso de 12 amostras, e no gráfico abaixo o sinal do microfone de contato.

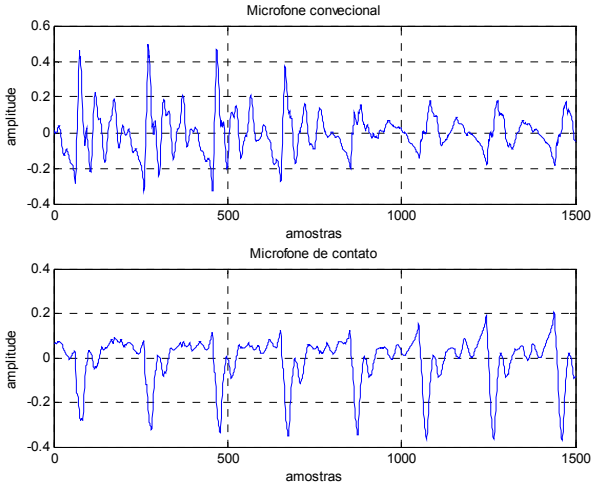

Fig. 3: sinal de voz e do microfone de contato

Observando os sinais obtidos pelo microfone de contato e pelo microfone convencional, nota-se que durante os trechos sonoros, ou seja onde ocorre atividade glotal, o sinal demonstra características mais regulares do que o sinal de voz. Esta característica se justifica pelo fato de que as vibrações captadas pelo contato provêm principalmente da laringe (considerada um tubo com dimensões constantes) que produz harmônicos quase invariantes ao longo do tempo. Por outro lado, para o sinal de voz, as diferentes configurações do trato vocal para a produção de diferentes sons, produzem diferentes harmônicos (formantes), que se mantém regulares por curtos períodos de tempo.

Na Fig 3, observa-se que a ocorrência de picos no sinal do microfone de contato podem ser bons indicativos para os GCIs. Porém, em alguns casos, estes picos têm sua amplitude reduzida não sendo possível detectá-los como máximos locais. Isto, de modo geral, inviabiliza a obtenção dos GCIs diretamente do sinal do microfone de contato, sendo necessário então um procedimento de detecção um pouco mais elaborado, como visto a seguir.

OBTENÇÂO DOS GCIs

De modo geral, destacamos neste trabalho duas técnicas para determinação dos GCIs a partir do sinal de voz: as técnicas baseadas no resíduo da predição linear e as técnicas baseadas no conceito de máxima verossimilhança. Estas famílias de algoritmos são discutidas em seqüência.

Métodos Baseados no Resíduo da Predição Linear

Muitos sistemas de análise da voz são baseados no modelo linear fonte-filtro, constituído por um filtro digital linear autoregressivo, que modela o trato vocal e uma fonte de excitação periódica considerada como um sinal representativo da atividade glotal.

Vários algoritmos precursores da detecção automática de eventos no sinal de voz, como por exemplo, os descritos em [10], [11], [12], [13] e [14], baseiam-se na idéia de que em segmentos curtos (menores do que um período de pitch) que não contêm uma excitação, o modelo de predição linear é mais adequado e conseqüentemente o erro de predição é menor. Por outro lado, quando um instante de excitação, ou o instante de fechamento glotal, está incluído no segmento de análise o erro de predição linear é maior. Deste modo, o ponto onde ocorre um grande erro de predição pode ser usado para indicar o instante do fechamento glotal [13].

Em [14], a partir da suposição de que o resíduo de predição linear exibe picos correspondentes aos GCIs, os autores observam que devidos a alguns fatores, como por exemplo, a estimação não acurada das formantes e da largura de bandas na etapa de análise, múltiplos picos podem ocorrer no sinal de resíduo, tornando difícil a estimação precisa [14] dos GCIs. Deste modo, o método proposto em [14] procura reduzir estas ambigüidades. Para isto, inicialmente o sinal de resíduo é processado no domínio da freqüência, aplicando-se uma janela de Hanning à sua FFT, para reduzir as componentes de baixas e altas freqüências. Em seguida, é obtido o contorno da Transformada de Hilbert de modo a atenuar os efeitos de fase introduzidos na obtenção do resíduo.

Na aplicação do método ao sinal do microfone de contato, inicialmente foi aplicado um filtro de pré-ênfase enfatizando as altas freqüências, para tornar seu decaimento espectral similar ao sinal de voz. Em seguida a aplicação do método de detecção pode ser observada na Fig. 4.

Observa-se que os máximos, correspondentes aos pontos de excitação do trato vocal, ocorrem com maior amplitude do que no sinal de voz, sendo mais fácil detectá-los em meio ao ruído. Porém, os picos intermediários aos supostos GCIs também ocorrem com maior amplitude.

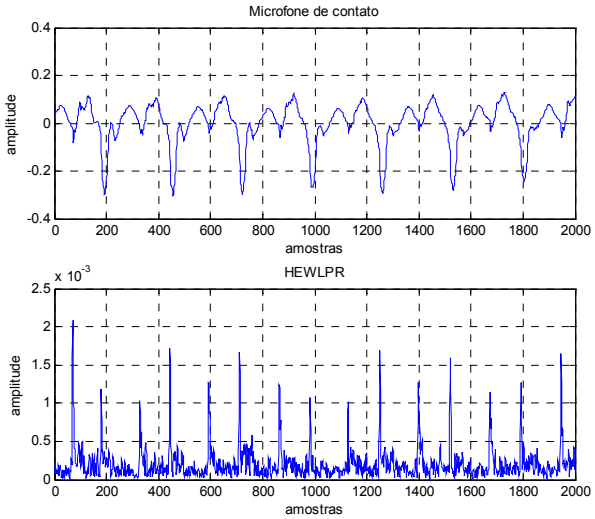

Fig. 4: Obtenção dos GCIs por resíduo da predição linear.

Supõe-se que o sinal do microfone de contato forneça informações mais intensas não somente sobre o instante de fechamento da glote, mas também em outros instantes de batimento das cordas vogais, como por exemplo, o instante de abertura da glote. Por um lado, o método confirma a suposição de que um microfone de contato forneceria informações mais nítidas sobre o movimento da glote, por outro lado, a dificuldade em separar o GCIs torna o método ineficiente para esta proposta.

Métodos Baseados na Máxima Verossimilhança

Esta metodologia foi proposta em [5] para estimar os GCIs, adaptada da teoria de detecção de épocas (ou eventos) por máxima verossimilhança em aplicações para radar. Assim como na seção anterior, este método assume que o sinal de voz dentro de um período de pitch é induzido por um pulso em uma época, geralmente definida como a representação de um GCI.

Assumindo que a produção da voz pode ser modelada por um sistema linear autoregressivo, o sinal modelo devido a uma época pode ser expresso como:

$$\hat{s}(n) = \begin{cases} \sum_{i=1}^{p} a_i \hat{s}(n-i) & 0 < n \le \infty \\ G & n = 0 \\ 0 & n < 0 \end{cases} \qquad (1)$$

onde G é uma constante positiva arbitrária e p a ordem do polinômio.

Em seguida, é suposto que a diferença entre o sinal observado, $s(n+n_0)$ $n \in [0, N-1]$ (onde n_0 é uma seqüência de atrasos de alinhamento) e o sinal modelo é um processo gaussiano e que as N observações constroem um processo gaussiano com N dimensões independentes e variância uniforme σ.

Assim, dado $x(n) = s(n+n_0) - \hat{s}(n)$, a densidade de probabilidade condicional, ou função de verossimilhança, será descrita por:

$$p(X \mid \theta) = \frac{1}{(2\pi\sigma^2)^{N/2}} . \exp\left\{ \frac{-\sum_{n=0}^{N-1} \left[s(n+n_0) - s(n) \right]^2}{2\sigma^2} \right\} \qquad (2)$$

onde θ é o espaço de parâmetros $\theta = \{\sigma, a_1, a_2, a_3, ..., n_0\}$.

Deste modo, quando o valor dos parâmetros maximizarem a função de verossimilhança, significa que uma época ocorreu. Maximizar a função de verossimilhança pode ser substituído por maximizar o logaritmo da verossimilhança e portanto a função a ser maximizada torna-se:

$$\ln\left[p(X \mid \theta) \right] = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{\sum_{n=0}^{N-1} \left[s(n+n_0) - s(n) \right]^2}{2\sigma^2} \qquad (3)$$

Assim conclui-se que não é possível encontrar uma expressão explícita para um valor ótimo de n_0. Porém, resolvendo algebricamente a potência interna ao somatório, e observando as possibilidades de máximos na função de $\ln[p(X \mid \theta)]$ em função de n_0, tem-se que o termo $\sum_{n=0}^{N-1} \left[s(n+n_0) \hat{s}(n) \right]$ é dominante. Este termo é chamado de sinal MLED (m*aximum-likelihood epoch determination*) e trata-se da correlação cruzada entre o sinal de voz e o sinal modelo. Portanto, em função de n_0, os máximos no sinal MLED equivalem aos máximos na função de verossimilhança. Em seguida, os coeficientes do sinal modelo, que produzem um máximo na função de verossimilhança, são deduzidos como os coeficientes de predição linear obtidos pelo método da autocorrelação. Neste sentido, o sinal modelo é considerado como os coeficientes de um filtro casado [5].

Em um período do sinal MLED, aparecem não só os máximos locais, onde a correlação cruzada é máxima, que correspondem aos GCIs, mas também a outros falsos candidatos. A razão em amplitude entre o pulso principal e os outros pulsos varia substancialmente e depende das propriedades do sinal, criando ambigüidade na decisão de escolha [5]. Para contornar este problema, os autores propõem o uso de um "sinal de seleção", similar à aplicação de uma janela, para enfatizar o contraste entre o pulso principal e os pulsos secundários. Os autores demonstram que o contorno da transformada de Hilbert (ou módulo do sinal analítico) do sinal MLED pode ser utilizado como sinal de seleção. A média pode ainda ser subtraída para tornar o sinal de seleção mais parecido com um pulso, sendo possível anular o sinal entre pulsos adjacentes.

Aplicando o método ao sinal de voz, os autores observaram experimentalmente que o indicativo para o GCI é melhor definido a 50% da amplitude do máximo (de zero até o ponto máximo do pulso, à esquerda) e este critério é empírico [5]. Esta imprecisão no posicionamento do GCI relatada pelos autores foi verificada em vários sinais, porém o critério de correção sugerido pelos autores nem sempre é eficiente.

Aplicando o método ao mesmo sinal do experimento anterior, passado igualmente por um filtro de pré-ênfase, o resultado da detecção é mostrado na Fig 5.

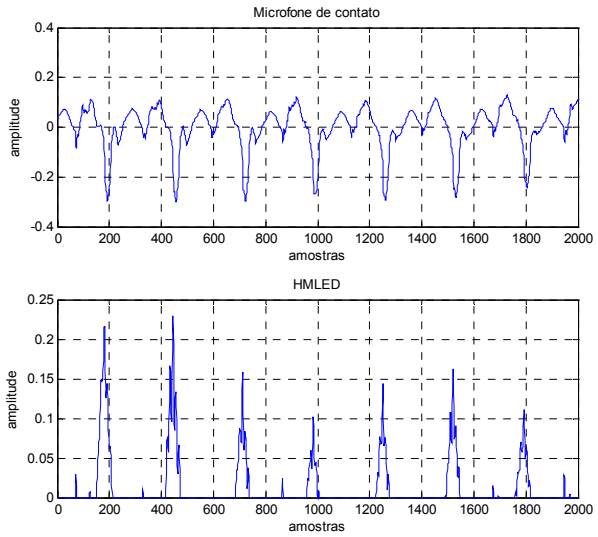

Fig. 5: Obtenção dos GCIs por máxima verossimilhança.

Conforme dito anteriormente, o método quando aplicado ao sinal de voz, apresenta uma imprecisão na detecção dos GCIs, no qual era necessário um método empírico de correção. Surpreendentemente, o método quando aplicado ao sinal do microfone de contato, apresentou esta imprecisão somente em alguns poucos casos e da ordem de 0,3 ms. Além disso, foi constatado que o módulo da transformada de Hilbert do sinal MLED é mais preciso do que próprio sinal MLED ou da multiplicação dos dois sinais conforme proposto em [5]. De fato, somente a subtração da média global do módulo da transformada de Hilbert permitiu a diferenciação entre os trechos sonoros e surdos.

SISTEMA PARA OBTENÇÃO DOS CGIs

Para o auxílio na construção de um banco de unidades para síntese por concatenação, foi implementado um aplicativo na linguagem C++, para o sistema operacional Windows, que permite ao usuário gravar os sinais em modo estéreo exibindo-os em janelas paralelas; detectar automaticamente o atraso entre os sinais; obter as marcas de GCIs automaticamente e editar estas marcas para efeito de sintonia fina.

A detecção do atraso é feita baseada no resíduo de predição linear. Apesar do sinal de resíduo do microfone de contato exibir outros picos de excitação além daqueles contidos no sinal de voz, observa-se que a correlação entre os resíduos é capaz de determinar o atraso presente no sinal de voz. Porém, diferenças de polaridade nos picos em amplitude nos resíduos, influenciam o resultado da correlação. Deste modo, foi utilizado o método [14] para reduzir as ambigüidades nos sinais de resíduo e em seguida calcular a correlação entre os dois num intervalo característico para o atraso.

A obtenção automática dos GCIs foi feita utilizando o módulo da transformada de Hilbert do sinal MLED, onde os máximos locais são detectados em segmentos de curta duração.

Para verificação do método, foram observados casos onde a detecção dos GCIs a partir do sinal de voz é extremamente difícil, principalmente para consoantes vozeadas.

Nas Fig. 6, 7, 8 e 9 são mostrados dois gráficos onde são mostrados o sinal de voz e os GCIs obtidos diretamente do sinal de voz e do sinal do microfone de contato, respectivamente. Para notação dos segmentos e na transcrição fonética foram utilizados os símbolos da Associação Internacional de Fonética (IPA) e a nomenclatura utilizada para as consoantes segue a definida em [15].

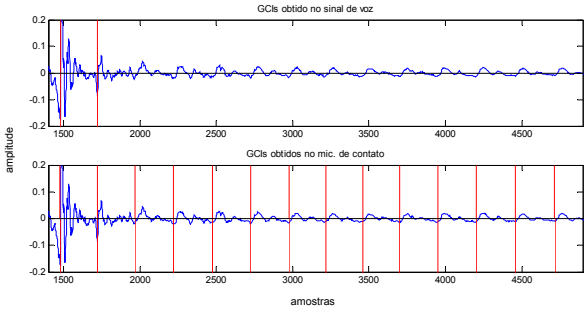

Fig. 6: Segmento do sinal de voz contendo a consoante Oclusiva Bilabial Vozeada /b/, recortada da palavra "abril" - /aˈbriɫ/, e as marcas dos GCIs obtidos do sinal de voz e do microfone de contato.

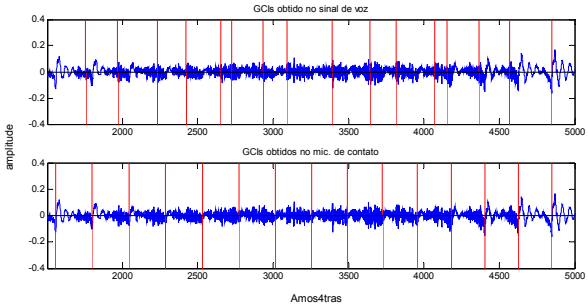

Fig. 7: Segmento do sinal de voz contendo a consoante Fricativa Alveolar Vozeada /z/, recortada da palavra "casa" - /ˈkaza/, e as marcas dos GCIs obtidos do sinal de voz e do microfone de contato.

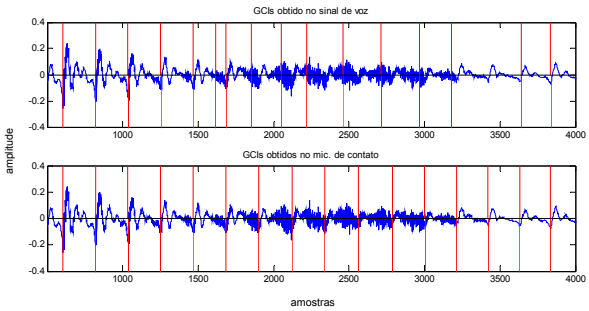

Fig. 8: Segmento do sinal de voz contendo a consoante Fricativa Alveopalatal Vozeada /ʒ/, recortada da palavra "mesmo" - /ˈmeʒmu/ e as marcas dos GCIs obtidos do sinal de voz e do microfone de contato.

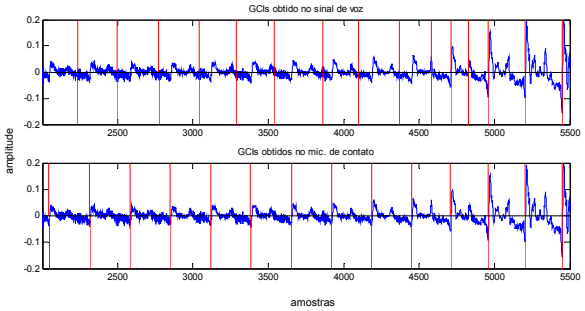

Fig. 9: Segmento do sinal de voz contendo a consoante Fricativa Alveolar Vozeada /v/, recortada da palavra "avante" - /aˈvãtʃi/, e as marcas dos GCIs obtidos do sinal de voz e do microfone de contato.

CONCLUSÕES

Foi estabelecido um método para obtenção precisa e automática das marcas de pitch em segmentos de sinais de voz, a serem utilizadas pelo algoritmo TD-PSOLA na concatenação destes segmentos. As marcas coincidem com os instantes de fechamento glotal, reduzindo a possibilidade de erros de fase na concatenação.

O método utiliza a gravação simultânea do sinal de voz e o sinal obtido por um disco piezoelétrico em contato com a pele, localizado na base do pescoço, no qual se mostrou conter mais informações sobre a atividade glotal do que o próprio sinal de voz. O disco piezoelétrico usado, assim como o circuito pré-amplificador teve caráter experimental, deste modo sugere-se que as características deste dispositivo sejam melhor exploradas assim como circuitos adequados à instrumentação. Observa-se ainda que o sinal obtido pelo microfone de contato é menos sujeito a ruído ambiente; que em trechos sonoros o sinal apresenta características mais estacionárias do que o sinal de voz; e que as consoantes sonoras são enfatizadas.

Foi descrito assim todo um método (semi-)automático para detecção automática do GCIs. Os resultados se mostraram bastante promissores, especialmente quando o método é aplicado a casos críticos onde a obtenção precisa dos GCIs a partir apenas do sinal de voz apresenta grande dificuldade quando realizada pelos métodos tradicionais.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Black, A.W., Lenzo, K.L., "Building Synthetic Voices", *FestVox 2.0*, 2003.

[2] Charpentier, F., Moulines, E. "Pitch-Synchronous wave form processing techniques for text-to-speech synthesis using diphones", *Proceedings of Eurospeech 89,* v. 2, pp. 13-19, 1989.

[3] Dutoit, T., *An introduction to text-to-speech synthesis*, Kluwer Academic Publishers, London, 1997.

[4] Hess, W., *Pitch Determination of speech Signals*, Springer-Verlag, Berlin, 1983.

[5] Cheng, Y. M., O'Shaughnessy, D., "Automatic and Reliable Estimation of Glottal Closuse Instants and Period", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 37, n. 12, pp. 1805-1815, 1989.

[6] Krishnamurthy, A. K., Childers, D. G., "Two-Channel Speech Analysis", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 34, No. 4, pp.730-743, 1986.

[7] Graciarena, M., Franco, H., Sonmez, K., et al., "Combining Standard and Throat Microphones for Robust Speech Recognition", *IEEE Signal Processing Letters*, v. 10, n. 3, 2003.

[8] Askenfelt, A., Gauffin, J., Sundberg, J., et al., "A comparison of contact microphone and electroglottograph for the measure of fundamental frequency", *Journal of Speech and Hearing Research*, v. 23, n. 2, pp. 258-273, 1980.

[9] Cittadinni, R., Poulan, F., "TS971 based Electret Condenser Microphone amplifier", *AN1534 Application Note*, STMicroeletronics, 2002.

[10] Wong, D.Y., Markel, J.D., Gray, A.H., "Least squares glottal inverse filtering from the acoustic speech waveform", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 27, pp. 350-355, 1979.

[11] Smits R. , Yegnanarayana, B., "Determination of Instants of Significant Excitation in Speech Using Group Delay Function", *IEEE Transactions on Speech Audio Processing*, v. 3, pp. 325-333, 1995.

[12] Strube, H. W., "Determination of the instants of glottal closure from the speech wave", *J. Acoust. Soc. Amer.,* v. 56, n. 5, pp. 1625-1629, 1974.

[13] ChangXue MA, Kamp Y. K., Willems L. F., "Frobenius Norm Approach to Glottal Closure Detection from the Speech Signal", *IEEE Transactions on Speech and Audio Processing*, v.2, pp. 258-265, 1994.

[14] Ananthapadmanabha T.V., Yegnanarayana, B., "Epoch Extration from linear Prediction Residual for Identification of Closed Glottis Interval", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 27, n. 4, pp. 309-319, 1979.

[15] Thaís C. S., *Fonética e Fonologia do Português*, Editora Contexto, São Paulo, 2003.

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na IX Convenção Nacional
11 – 13 de Abril de 2005, São Paulo, SP

Este artigo foi reproduzido do original entregue pelo autor, sem edições, correções e considerações feitas pelo comitê técnico deste evento. 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 reservados. Não é permitida a reprodução total ou parcial deste artigo sem autorização expressa da AES Brasil.

Codificação Perceptiva de Sinais de Voz de Banda Larga

P. A. Berger, F. A. O. Nascimento, L. M. da Silva.
Departamento de Engenharia Elétrica, Universidade de Brasília, Brasil.
pedrobg@uol.com.br

RESUMO

Esse artigo apresenta o estudo e simulação de um esquema de codificação perceptiva para sinais de voz de banda larga. O esquema é uma versão modificada do codificador de áudio AC-3, para atender especificamente às características de sinais de voz de banda larga. Sinais de voz codificados com taxas de bits de 32, 24 e 16 kbps foram obtidos e avaliados subjetivamente. Avaliou-se também a robustez do esquema em relação à perda de blocos de informação. Verificou-se o esquema tolera razoavelmente uma perda de até 10% dos blocos de coeficientes MDCTs.

INTRODUÇÃO

A maioria dos atuais codificadores de voz realiza a compressão do sinal pelo uso de modelos de produção da voz e explorando o fenômeno do mascaramento auditivo. Contudo, esses modelos não são, em geral, apropriados para outros sinais tais como voz ruidosa, sinal contendo vozes de vários locutores, música e ruído de fundo. Por outro lado, codificadores baseados em transformadas e modelos perceptivos podem comprimir eficientemente diferentes tipos de sinais, uma vez que eles não são tão especializados quanto os codificadores baseados em modelos de produção da voz [8].

Com o uso crescente da transmissão de sinais de voz por redes de pacote, a robustez dos codificadores em relação à perda de pacotes passou a ser uma característica de grande importância. A informação contida nos pacotes perdidos deve ser substituída de uma forma que minimize a deterioração do sinal de voz reproduzido. Nos codificadores baseados em predição linear, os efeitos dessa perda podem se propagar por vários quadros através da realimentação feita pelo preditor. Os codificadores baseados em transformadas podem, em princípio, apresentar maior robustez a esse tipo de degradação. Primeiro, porque a propagação dos efeitos da perda de um bloco de informações será mais restrita. Segundo, porque talvez os esquemas baseados em transformadas permitam que a informação perdida seja substituída de forma mais satisfatória por versões obtidas por interpolação.

Um outro aspecto a ser considerado é a largura de banda do sinal de voz a ser codificado. Larguras de banda maiores que a largura de banda telefônica (300 a 3400 Hz) propiciam significativas melhorias subjetivas na qualidade da voz

codificada. Uma largura de banda de 50 a 7000 Hz não só melhora a inteligibilidade e naturalidade da voz, mas acrescenta também um sentimento de transmissão transparente e facilita o reconhecimento do locutor. Sinais de voz com essa largura de banda são denominados sinais de voz de banda larga. Várias aplicações relevantes demandam esse tipo de voz, exemplos: sistemas de teleconferência RDSI e comunicações audiovisuais interativas multipontos.

Esse artigo apresenta o estudo e simulação de um esquema de codificação perceptiva para sinais de voz de banda larga. O esquema é uma versão adaptada do codificador de áudio AC-3 [1]-[3], que é um esquema de codificação baseado em modelo perceptivo e transformada MDCT (modified discrete cosine transform). Foram feitas modificações consideráveis no esquema de codificação original do AC-3 para que o codificador proposto pudesse torna-se eficiente na codificação de sinais de voz de banda larga. Dentre as modificações mais importantes podemos destacar:

• A codificação diferencial dos expoentes foi modificada para adequar-se à largura de banda de 50-700Hz (largura de banda para sinais de voz de banda larga);

• A reutilização de expoentes codificados foi reduzida em relação ao esquema original do AC-3, uma vez que este procedimento torna o codificador muito sensível à perda de pacotes;

• A integração dos coeficientes transformados em bandas críticas foi alterada para atender às novas especificações de largura de banda;

• Os cálculos dos limiares de mascaramento foi também alterado para adequar-se ao novo esquema de integração em bandas críticas;

A estrutura básica do esquema de codificação é apresentada na próxima seção. Duas etapas importantes do esquema e os testes práticos e avaliações são, então, detalhadas nas seções seguintes. As conclusões e possíveis melhorias a serem intentadas no futuro são reportadas na última seção.

METODOLOGIA

Estrutura Básica de Codificação e Decodificação

A Fig. 1 mostra o diagrama de blocos do esquema de codificação investigado: uma versão simplificada e adaptada do codificador de áudio AC-3 [1]-[3]. A codificação é feita no domínio da freqüência, usando uma MDCT (modified discrete cosine transform) de 2M pontos com 50% de superposição [4]-[7]. Isto é, a transformada é aplicada a blocos de 2M amostras PCMs, com 50% de superposição entre blocos — veja a Fig. 2. Contudo, somente coeficientes MDCTs são gerados para cada bloco de entrada de amostras; ou seja, essa transformação é equivalente a um banco de M filtros cujas saídas são criticamente amostradas. As funções de base MDCTs estendem-se, portanto, por dois blocos no tempo, levando à eliminação virtual dos efeitos de bloco que degradam o sinal reconstruído quando a transformada usada é sem superposição. Dado um bloco de entrada x(n) ,0≤n≤2M-1, a MDCT direta pode ser definida como [5, 7]:

$$X(k) = \sum_{n=0}^{2M-1} x(n)\, h_k(n), \quad 0 \le k \le M-1 \qquad (1)$$

Onde,

$$h_k(n) = w(n)\sqrt{\frac{2}{M}}\cos\left[\frac{(2n+M+1)(2k+1)\pi}{4M}\right], \quad 0 \le k \le M-1 \qquad (2)$$

sendo que w(n) é uma janela temporal de comprimento 2M — na abordagem de banco de filtros, ela corresponde ao filtro passa-baixos protótipo a partir do qual são gerados os filtros passa-faixas $h_k(n)$, 0≤k≤M-1, que comporão o banco de filtros. A MDCT inversa (IMDCT), nesse caso, é definida como:

$$x(n) = \sum_{n=0}^{M-1}\left[X(k)h(k,n) + X^P(k)h(k,n+M)\right], \quad 0 \le n \le M-1 \qquad (3)$$

Onde $X^P(k)$ denota o bloco precedente de coeficientes MDCTs. Para que a reconstrução de x(n) seja perfeita é preciso que a janela satisfaça as seguintes condições:

$$\left.\begin{aligned} w(2M-1-n) &= w(n) \\ w^2(n) + w^2(n+M) &= 1 \end{aligned}\right\}, \quad \text{para } 0 \le n \le M-1 \qquad (4)$$

Nas simulações realizadas, o sinal de voz (ou de áudio) tem largura de banda de 50 a 7000 Hz e foi amostrado a taxa de 16 kHz. O comprimento dos quadros foi fixado em 512 amostras ou, equivalentemente, 32 ms. Portanto, são gerados 256 coeficientes espectrais (ou MDCTs) a cada 16 ms. O espaçamento freqüêncial entre coeficientes é de $8.000 \div 256 = 31{,}25$ Hz.

Cada um dos coeficientes MDCTs é representado na notação binária de ponto flutuante como um expoente e uma mantissa. O conjunto de expoentes é codificado em uma representação grosseira do espectro do sinal que é referida como envoltória espectral. Com base nessa envoltória espectral, é obtida, pela aplicação de um modelo perceptivo, uma curva de limiar de mascaramento, que indica qual o nível máximo que o ruído de codificação pode ter para que ele não seja percebido pelo ouvinte. Essa curva é usada pelo algoritmo de alocação de bits para determinar quantos bits serão usados para codificar cada uma das mantissas.

O processo de decodificação é basicamente o inverso do processo de codificação. Uma vez que os expoentes tenham sido decodificados, eles são usados para reproduzir o modelo perceptivo e a alocação de bits usados pelo codificador. Os bits que representam as mantissas são, então, decodificados. Os coeficientes MDCTs reconstruídos são transformados de volta para o domínio do tempo para produzir as amostras PCMs decodificadas — veja ilustração na Fig. 2(b) e equação (3).

O algoritmo de alocação de bits usado nesse trabalho é do tipo backward, ou seja, ele não requer a transmissão de qualquer informação lateral entre o codificador e o decodificador. Ele é uma versão simplificada do algoritmo backward/forward híbrido usado pelo codificador AC-3 [1].

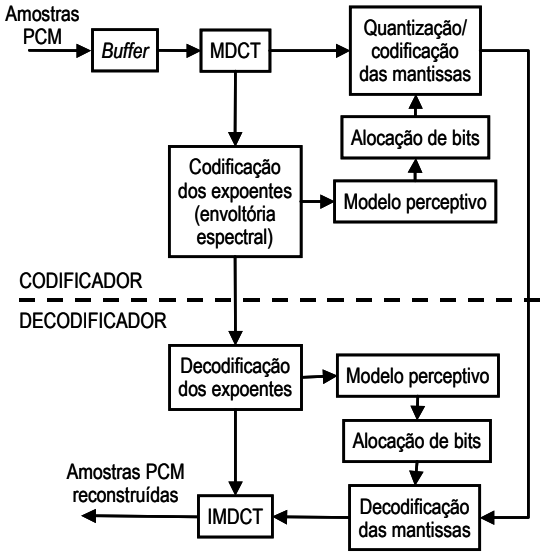

Figura 1 – Diagrama de blocos do codificador e decodificador.

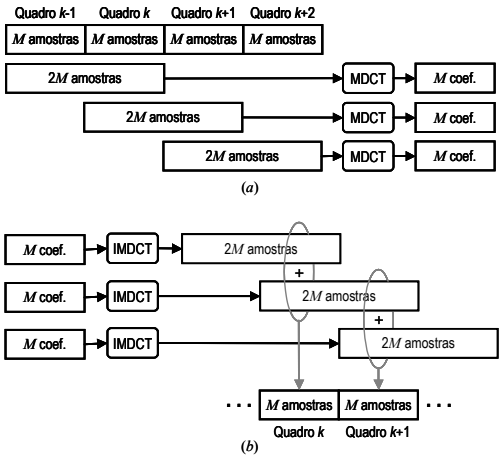

Figura 2 – Ilustração da aplicação da MDCT: (a) MDCT direta e (b) MDCT inversa (IMDCT).

Codificação dos Expoentes

Os coeficientes MDCTs de índice 224 até 255 correspondem a freqüências fora da banda do sinal (50-7000 Hz) e, por isso, seus expoentes e mantissas não são transmitidos: o decodificador os faz iguais a zero. Os expoentes dos 224 coeficientes restantes, de índice 0 até 223, são divididos em grupos. De cada grupo, apenas o expoente de maior valor é codificado e enviado, os demais expoentes do grupo são feitos iguais a esse de maior valor. A mantissa que tem seu expoente alterado é escalonada para compensar a mudança feita no expoente. Apenas um expoente é enviado integralmente, codificado em uma palavra de 5 bits. Os demais expoentes são codificados diferencialmente, com deltas de 0, ±1 ou ±2. Os deltas são divididos em conjuntos de 3 deltas e cada conjunto é codificado em uma palavra de 7 bits. Foram experimentados dois tipos de agrupamento dos expoentes:

- Agrupamento Uniforme (AU): Os 224 expoentes são divididos em 112 grupos, cada um contendo um par de expoentes. Portanto, são gastos $5+(111\div 3)\times 7=264$ bits para representar os expoentes de um bloco de coeficientes MDCTs.
- Agrupamento Não-Uniforme (ANU): Os 224 expoentes são divididos em 67 grupos, que podem ter 2, 3, 4 ou 5 elementos. Grupos correspondentes a faixas de freqüências baixas contém 2 elementos e grupos que correspondem a faixas de freqüências altas podem conter até 5 elementos. Essa divisão não uniforme é inspirada nas bandas críticas da audição humana, que também têm largura crescente com a freqüência. Nesse caso, são gastos $5+(66\div 3)\times 7=159$ bits para representar os expoentes de um bloco de coeficientes MDCTs.

Uma forma de reduzir o dispêndio de bits com a codificação dos expoentes é a reutilização dos expoentes codificados de um bloco de coeficientes MDCTs em blocos subseqüentes. No codificador AC-3, essa reutilização pode-se estender a até cinco blocos subseqüentes. Nas simulações realizadas nesse trabalho, apenas a reutilização em um bloco subseqüente foi considerada. Isso porque essa reutilização torna o codificador muito sensível a perda de informações: por exemplo, devida à perda de pacotes em uma transmissão baseada em pacotes.

Com uma reutilização, o dispêndio médio de bits para representar os expoentes dos coeficientes MDCTs é de 132 bits por bloco (ou por 15 ms), caso seja implementado o agrupamento uniforme. E é de 79,5 bits, se for usado o agrupamento não-uniforme.

Alocação de Bits

Uma vez fixada a taxa de bits, (bps), a ser despendida na codificação do sinal, o número de bits que pode ser gasto a cada quadro é dado por $B_{\mathrm{Q}}=R_b T_{\mathrm{Q}}$, onde $T_{\mathrm{Q}}=15$ ms é a duração de um quadro. No codificador simulado, $B_{\mathrm{Q}}=B_{\mathrm{E}}+B_{\mathrm{M}}$, onde B_{E} e B_{M} são, respectivamente, as quantidades de bits gastas na codificação dos expoentes e das mantissas dos coeficientes MDCTs. A codificação dos expoentes e o valor de B_{E} foram assuntos da Seção 3. Nesta seção discute-se o processo pelo qual os B_{M} bits restantes são distribuídos entre as mantissas dos coeficientes MDCTs de modo a minimizar a distorção audível presente no sinal reconstruído. Essa distribuição é dinâmica, enquanto que a distribuição dos B_{E} bits entre os expoentes é estática.

O objetivo é determinar $b(k)$, $k=0,1,\ldots,M-1$, sujeito à seguinte restrição:

$$\sum_{k=0}^{M-1} b(k)=B_{\mathrm{E}} \tag{5}$$

onde $b(k)$ é o número de bits alocado para representar a k-ésima mantissa e que define o número de níveis do quantizador a ser usado. O critério usado é o da distorção perceptiva. Bits são distribuídos buscando fazer com que a envoltória do espectro de potência do ruído de quantização fique sempre abaixo do limiar de mascaramento. Todo componente de ruído com nível inferior a esse limiar é encoberto pelo próprio sinal de voz (ou de áudio) e não é

percebido pela audição humana [7]. O limiar de mascaramento é calculado de acordo com um modelo perceptivo (ou psicacústico). A seguir é descrito de forma sucinta o modelo psicacústico usado nesse trabalho: ele é uma versão simplificada do modelo psicacústico usado pelo AC-3 [1]-[3].

O cálculo do limiar de mascaramento é realizado em três etapas [2]-[3]. Primeiro, os expoentes codificados $\exp(k)$ são transformados em uma densidade espectral de potência (DEP) logarítmica da seguinte forma:

$$\sum_{k=0}^{M-1} b(k) = B_{\mathrm{E}} \tag{6}$$

Na escala logarítmica resultante, 128 unidades corresponde a 6 dB. Então, as amostras $\mathrm{dep}(k)$ são agrupadas em bandas cuja largura corresponde, aproximadamente, à metade da largura das bandas críticas da audição humana. No codificador simulado nesse trabalho, o espectro na faixa entre 0 e 7000 Hz ($0 \leq k \leq 223$) é dividido em 37 bandas, contendo o número de amostras indicado na Tabela 1 — o espaçamento entre duas amostras adjacentes é de 31,25 Hz. As amostras contidas em uma única banda são somadas (em escala linear), resultando, então, uma DEP integrada com apenas 37 amostras. O propósito dessa integração é reduzir o esforço computacional requerido para realizar a terceira etapa.

Finalmente, o limiar de mascaramento versus freqüência é obtido convoluindo a DEP integrada e um protótipo de função de espalhamento — isto é, um modelo da curva de mascaramento da audição humana. O modelo usado nas simulações é o modelo proposto para o AC-3, com parâmetros estáticos [3]. O resultado desse cálculo é uma curva indicando, para cada banda, o nível de mascaramento estimado. Essa curva é comparada com limiar absoluto da audição humana e o maior dentre os dois é mantido.

Tabela 1 – Número de amostras nas 37 bandas em que a faixa entre 0 e 7 kHz é dividida.

Banda	1-9	10-16	17-21	22-24	25-26	27-28	29-30	31-32	33	34	35	36	37
Tamanho	2	3	4	5	6	7	9	11	14	15	17	18	20

A curva estimada de mascaramento é subtraída da $\mathrm{dep}(k)$ para determinar a razão sinal-ruído (RSR) desejada para cada coeficiente MDCT. A cada bit adicional alocado a um quantizador uniforme, a RSR aumenta por 6 dB aproximadamente. Assim, para cada coeficiente MDCT, pode-se mapear a RSR desejada no número de bits $b(k)$ que deve ser alocado para o quantizador. O critério dos 6 dB por bit é impreciso para um número pequeno de bits alocados. O mapa utilizado nas simulações é o usado pelo AC-3 [3], que considera um ganho de RSR menor que 6 dB por bit para $b(k) \leq 4$. O valor máximo para $b(k)$ é 16.

Uma vez que $b(k)$, $k = 0, 1, \ldots, M-1$, esteja determinado, as mantissas são quantizadas usando quantizadores uniformes. Os quantizadores com até 15 níveis são simétricos e os que possuem mais de 15 níveis são assimétricos.

Várias mantissas não recebem bits para a sua codificação. Isso porque o componente espectral correspondente tem nível abaixo do limiar de mascaramento e, portanto, não é audível — logo, ele não precisa ser transmitido. Mas mantissas não recebem bits também porque os bits disponíveis acabaram, tendo sido gastos com mantissas mais relevantes. Seja qual for o caso, a providência natural a ser tomada pelo decodificador seria substituir o valor real dessas mantissas por zero. Contudo, em [3] é recomendado que os valores dessas mantissas sejam substituídos por valores aleatórios [3]. Testes subjetivos mostraram que essa estratégia realmente melhora qualidade subjetiva do sinal reconstruído: ele soa menos abafado, apresentado maior riqueza espectral de alta freqüência. Contudo, percebe-se também um ruído causado por essa estratégia. Assim, a amplitude máxima desses números aleatórios deve ser escolhida com cuidado. Nas simulações realizadas, foram utilizados números aleatórios com distribuição uniforme entre – 0,35 e + 0,35.

TESTES PRÁTICOS

Foram realizadas simulações do codificador descrito para três taxas de bits: 32, 24 e 16 kbps. A Tabela 2 apresenta a estratégia de codificação dos expoentes e a divisão, entre expoentes e mantissas, dos bits disponíveis por quadro.

Os sinais codificados com taxa de 32 e 24 kbps têm qualidade subjetiva equivalente. A qualidade desses sinais é muito boa, mas é possível notar um leve abafamento, provavelmente devido à falhas na reprodução dos componentes de alta freqüência. Nota-se também um ruído de baixa intensidade sibilado, que é devido aos componentes cujas mantissas não foram transmitidas e que foram substituídas por valores aleatórios. Contudo, o sinal com esse ruído é preferível ao sinal em cuja reconstrução faz-se iguais a zero as mantissas não transmitidas. Nesse último sinal é bem perceptível a ausência dos componentes de alta freqüência. A figura 3 mostra uma comparação entre um segmento de sinal de voz original e um segmento de sinal de voz decodificado para uma taxa de 32 kbps..

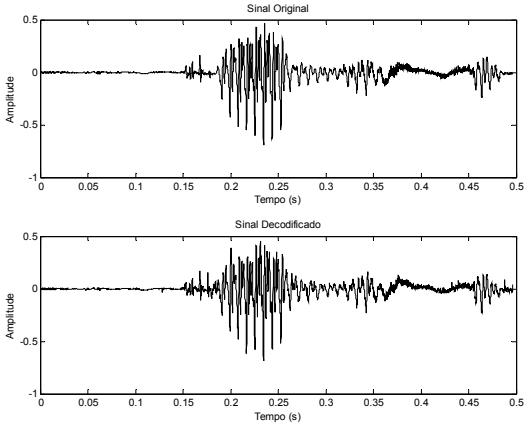

Figura 3 – Comparação entre um segmento de sinal de voz original e um segmento de sinal de voz decodificado para uma taxa de 32 kbps.

Os sinais codificados com 16 kbps apresentam uma degradação bem mais perceptível. Nota-se nesses sinais um ruído mais intenso e de espectro amplo, enquanto que no ruído presente nos sinais codificados com 32 ou 24 kbps predomina os componentes de alta freqüência. Medidas objetivas também detectam essa maior degradação. A Tabela 3 mostra exemplos de valores medidos de razão sinal-ruído segmentar (RSRseg) para sinais codificados com as três taxas de bits. Enquanto que a redução a taxa de bits de 32 para 24 kbps levou a uma queda de 1,33 dB na RSRseg, a redução de 24 para 16 kbps levou a uma queda de 2,48 dB na RSRseg.

Tabela 2 – Estratégia de codificação e a alocação de bits para as taxas de 32, 24 e 16 kbps.

Taxa de bits global		32 kbps	24 kbps	16 kbps
Agrupamento dos expoentes		AU	AU	ANU
Reutilização dos expoentes		Não	Sim	Sim
Número médio de bits gasto por quadro (16 ms)	Expoentes	264	264 ÷ 2 = 132	159 ÷ 2 = 79,5

Tabela 3 – Desempenho dos codificadores simulados, em termos da razão sinal-ruído segmentar (RSRseg).

Taxa de bits	32 kbps	24 kbps	16 kbps
RSRseg	13,12 dB	11,79 dB	9,31 dB

Um outro tipo de teste realizado foi a simulação de perda de blocos de coeficientes MDCTs e a substituição desses blocos por versões obtidas pela interpolação dos coeficientes MDCTs dos blocos anterior e o posterior ao bloco (ou blocos) perdido(s). Com o esquema que resultou em taxa de 32 kbps, a perda de até 10% dos blocos causa degradação quase imperceptível e com perda de 20% dos pacotes a qualidade é ainda razoável, sendo que a inteligibilidade ainda é muito boa. Com os outros dois esquemas, há uma deterioração maior, mas com perda de até 10% dos pacotes se tem ainda uma qualidade razoável e boa inteligibilidade. A razão para a maior sensibilidade desses dois esquemas é o fato deles reutilizarem os expoentes codificados. Nesse caso, quando um bloco que contém expoentes é perdido, o próximo bloco também fica perdido.

CONCLUSÃO

Neste artigo, apresentou-se o estudo de um esquema de codificação perceptiva para sinais de voz de banda larga. O esquema é uma versão simplificada do codificador de áudio AC-3. Foram realizadas simulações que geraram sinais codificados com taxas de bits de 32, 24 e 16 kbps. Avaliações subjetivas da qualidade dos sinais codificados indicam que o esquema é promissor, mas para operar com taxas de 16 kbps ou menor precisa ser melhorado. Estratégias que podem vir a propiciar a melhoria necessária são, por exemplo, o uso da quantização vetorial e a codificação paramétrica da envoltória espectral. Essas possibilidades serão investigadas em trabalhos futuros. Foi testada também a robustez do esquema simulado em relação à perda de blocos de informação. Verificou-se que perdas de até 10% dos blocos de coeficientes MDCTs é razoavelmente tolerada pelo esquema.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] C. Todd, G. Davidson, M. Davis, L. Fielder, B. Link e S. Vernon, "AC-3: Flexible Perceptual Coding for Audio Transmission and Storage", 96th Convention of the Audio Engineering Society, preprint 3796, fevereiro de 1994.

[2] G. Davidson, L. Fielder e B. Link, "Parametric Bit Allocation in a Perceptual Audio Coder", 97th Convention of the Audio Engineering Society, preprint 3921, novembro de 1994.

[3] Digital Television Standard, United States Advanced Television Systems Committee (ATSC), "ATSC Standard: Digital Audio Compression (AC-3), Revision A", Doc. A/52A, agosto de 2001.

[4] J.P. Princen e A.B. Bradley, "Analysis/Synthesis Filter Bank Design Based on Time Domain Aliasing Cancellation", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-34, pp. 1153-1161, outubro de 1986.

[5] H. Malvar, "Lapped Transforms for Efficient Transform/Subband Coding", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-38, pp. 969-978, junho de 1990.

[6] R.D. Koilpillai e P.P. Vaidyanathan, "Cosine-Modulated FIR Banks Satisfying Perfect Reconstruction", IEEE Trans. Signal Processing, vol. SP-40, pp. 770-783, abril de 1992.

[7] T. Painter e A. Spanias, "Perceptual Coding of Digital Audio", Proceedings of the IEEE, vol. 88, no. 4, pp. 451-513, abril de 2000.

[8] H. N-Azghandi e P. Kabal, "Perceptual Coding of Narrowband Audio Signals at 8 kbit/s", Proc. IEEE Workshop on Speech Coding for Telecom., pp. 109-110, setembro de 1997.

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na IX Convenção Nacional
11 – 13 de Abril de 2005, São Paulo, SP

Este artigo foi reproduzido do original entregue pelo autor, sem edições, correções e considerações feitas pelo comitê técnico deste evento. 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 reservados. Não é permitida a reprodução total ou parcial deste artigo sem autorização expressa da AES Brasil.

Amplificador Classe D de 1000W RMS com Realimentação

Fábio Vincenzi Romualdo da Silva, Luiz Carlos Gomes de Freitas,
Normandes José Moreira Júnior, João Batista Vierira Júnior,
Luiz Carlos de Freitas

Universidade Federal de Uberlândia
Campus Santa Mônica-Bloco 3N, Av. João Naves de Ávila, 2160
Código Postal, Uberlândia, Minas Gerais, Brasil
freitas@ufu.br

RESUMO

Este artigo apresenta uma nova topologia de amplificador Classe D de potência, com rendimento superior a 90%, adequado para operar com sub-woofer e woofer. A proposta não necessita de filtro passa baixa no estágio de saída, uma vez que possui como princípio de funcionamento, modular o sinal a ser amplificado por meio de duas fontes de corrente sobre um capacitor de filtro. O referido modo de operação possibilita que o amplificador opere com fontes de tensão não estabilizadas com THD inferior a 1%.

1 INTRODUÇÃO

A principal vantagem dos amplificadores Classe D sobre os amplificadores lineares tradicionais (Classe A, B e AB) é que, se bem projetados, o rendimento pode ser superior a 90% [1]. O alto rendimento do Classe D sobre os amplificadores lineares torna-o mais atrativo em equipamentos alimentados a bateria como é o caso dos amplificadores utilizados em carros, walkman, rádios portáteis, megafones, walkie-talkie, etc. Outra desvantagem dos amplificadores lineares, para aplicações de potência elevada, é o peso e o volume devido o uso de grandes dissipadores de calor que são necessários para evitar danos nos dispositivos semicondutores que compõe o circuito de potência dos amplificadores lineares.

Os amplificadores Classe D convencionais, como mostrado na Fig. 1, possuem um modulador de largura de pulso (PWM), um estágio de potência e um filtro passa baixa na saída. No estágio PWM, o sinal de referência é comparado com um sinal dente de serra de alta freqüência. O PWM que é resultado dessa comparação é composto por pulsos digitais cuja largura de pulso é proporcional ao valor instantâneo do sinal de entrada.

O estágio de potência apresenta, com maior freqüência, configuração em ponte completa para que, desta forma, possa propiciar maior potência de saída, principalmente em aplicações de baixa tensão. O filtro passa baixa é usado para remover os componentes harmônicos do sinal PWM compondo desta forma o sinal de referência amplificado na carga.

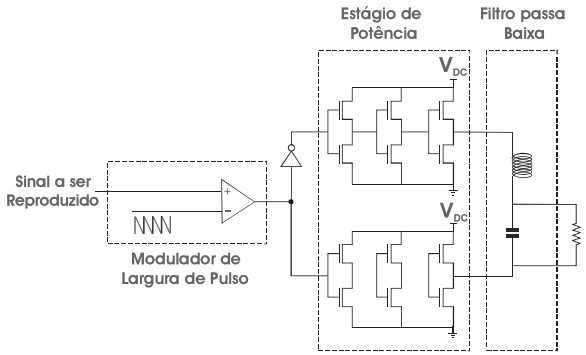

Fig. 1. Conversor Classe D convencional

Nos amplificadores não realimentados, a amplitude do sinal PWM gerado depende do nível de tensão da fonte que alimenta o circuito de potência. Qualquer oscilação na tensão de alimentação é repassada para o sinal PWM que provocará distorção na forma de onda do sinal de áudio[2].

No novo arranjo de amplificador Classe D proposto, Fig. 3, as fontes de alimentação não necessitam ser estabilizadas pelo fato da topologia ser realimentada além de não possuir filtro passa baixa adicional na saída.

A topologia proposta possui as seguintes vantagens com relação a topologia apresentada (Fig. 2) em trabalhos anteriores [3]-[6]:

- possui duas chaves ao invés de quatro e conseqüentemente dois drivers isolados ao invés de quatro.
- não é necessário utilizar snubber nas chaves
- rendimento mais elevado

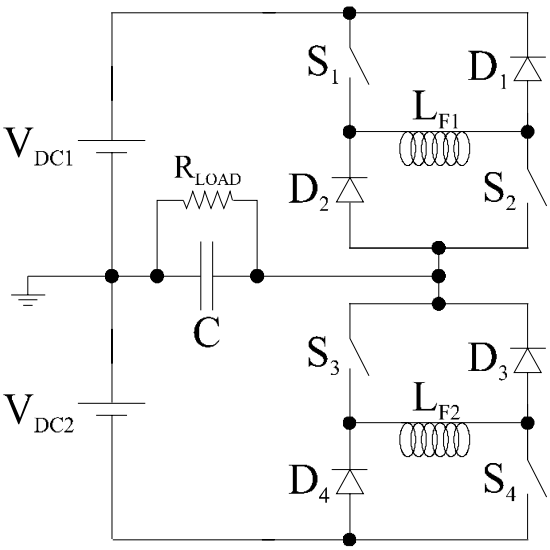

Fig. 2. Conversor Classe D apresentado em trabalhos anteriores

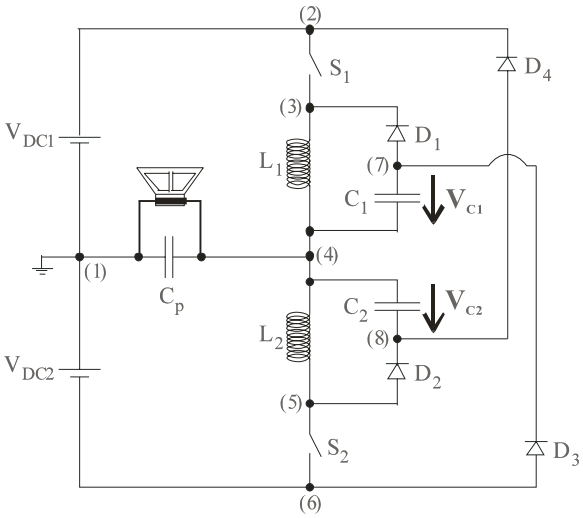

Fig. 3. Conversor Classe D proposto

2 ROTEIRO DE PROJETO

A máxima variação da tensão por tempo (slew rate) tem de ser considerada para o correto projeto do amplificador. O slew rate máximo é determinado considerando-se a máxima freqüência de onda senoidal ou triangular a ser amplificada. Quando maior a freqüência maior é a taxa de variação de tensão necessária para reproduzir uma forma de onda senoidal ou triangular. A forma de onda quadrada não pode ser considerada no cálculo da máxima taxa de variação de tensão já que teoricamente o slew rate necessário para reproduzi-las é infinito.

A análise matemática baseada num sinal de entrada senoidal, se comparado ao sinal triangular, simplifica a matemática estudada. E neste caso o sinal de entrada instantâneo é representado pela Equação (1).

$$V(t) = V_{pk}.\sin(\omega.t) \qquad (1)$$

onde:

- V_{pk} - tensão máxima de pico da senóide de saída
- ω - freqüência angular $2.\pi.f$;
- t – tensão instantânea de corrente

Derivando a Equação (1) resulta no slew-rate desejado, Equações (2).

$$\frac{dV(t)}{dt} = \omega.V_{pk}.\cos(\omega.t) \qquad (2)$$

A variação máxima ocorre na passagem por zero para a máxima freqüência a ser amplificada $\cos(0) = 1$. Isso

significa que $\cos(\omega.t)$ na Equações (2) pode ser eliminado o que resulta na Equações (3).

$$\frac{dV(t)}{dt} = \omega.V_{pk} \qquad (3)$$

A Equações (4) relaciona corrente e capacitância. Esta equação é muito importante para calcular a corrente necessária para produzir o slew-rate necessário no capacitor de filtro de saída.

$$I = C.\frac{dV}{dt} \qquad (4)$$

Combinando (3) e (4), resulta que a corrente pode ser calculada pela Equação (5).

$$I_{pk} = 2.\pi.f_{\max}.C.V_{pk} \qquad (5)$$

onde:

$f_{\max}$ - é a máxima freqüência a ser amplificada

Considerando taxa de crescimento positiva no capacitor C_p as Fig. 4 e 5 mostram que o indutor L_1 carrega o capacitor C_p e C_1 e descarrega o capacitor C_2. O processo de carga e descarga ocorre basicamente ao mesmo tempo devido a freqüência de chaveamento ser elevada. Desta forma o circuito mostrado na Fig. 6 pode ser usado para encontrar a equação de corrente que produz o máximo slew-rate.

onde:

- capacitor C_T é a soma das capacitâncias C_p, C_1 e C_2;
- em t = 0s a corrente no indutor é zero;
- em t = 0s a tensão no capacitor é zero.

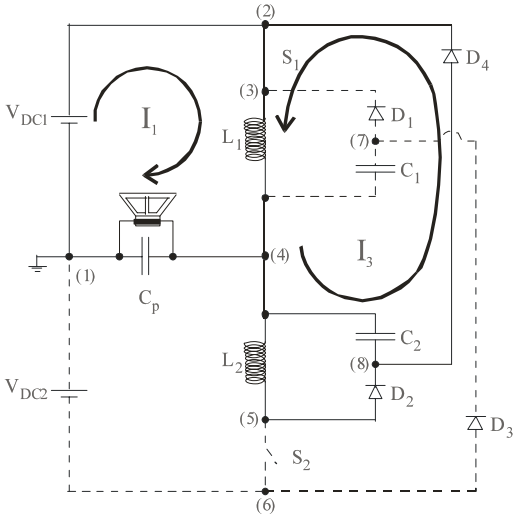

Fig.4: Carregando Cp e descarregando C2 (Taxa de Crescimento Positiva).

A Equação (6) mostra a corrente instantânea no circuito LC mostrado na Fig. 6.

$$I(t) = \frac{V_{DC}}{Zo}.\cos(\omega.t) \qquad (6)$$

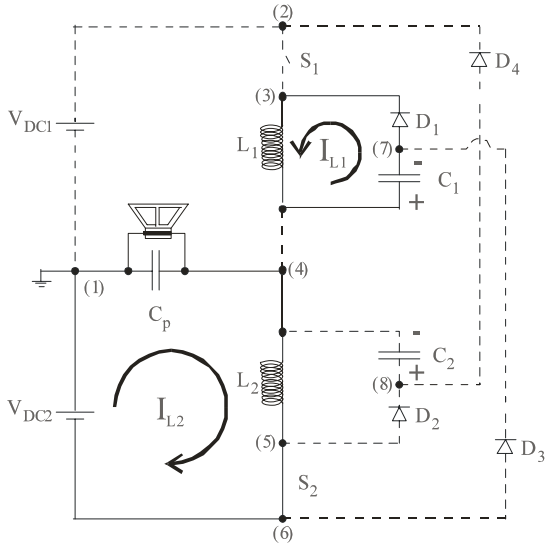

Fig.5: Carregando C1 e descarregando Cp (Taxa de Crescimento Positiva).

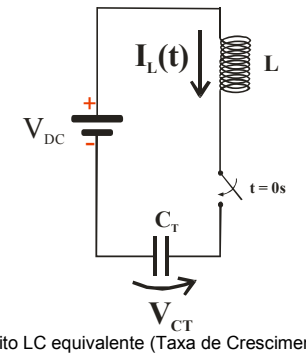

Fig.6: Circuito LC equivalente (Taxa de Crescimento Positiva).

A corrente máxima ocorre quando t=0, pois $\cos(0) = 1$. Isso significa que este termo, na Equações (6), pode ser eliminado o que resulta na Equações (7).

$$I_{pk} = \frac{V_{DC}}{Zo} \qquad (7)$$

onde:

$$Zo = \sqrt{\frac{L}{C_T}}$$

Combinando as Equações (5) e (7) resulta na Equação (8).

$$L = \frac{V_{DC}^{2}}{4.\pi^{2}.f_{\max}^{2}.V_{pk}^{2}.C_{T}} \qquad (8)$$

É interessante notar que a impedância carga nunca é considerada. Isso ocorre porque a carga não está relacionada com o slew-rate. A impedância é relevante para análise de resposta em freqüência e/ou análise de filtro. É interessante notar que o slew-rate máximo ocorre (na passagem por zero) quando a corrente é máxima e a tensão é zero. Se a tensão é zero a carga é incapaz de drenar corrente.

A capacitância total C_T é a soma das C_p , C_1 e C_2. Valores adequados para C_1 e C_2 são $C_p \leq \left(C_1 = C_2\right) \leq 2.C_p$.

3 EXEMPLO DE PROJETO

O objetivo deste projeto é alimentar um alto falante de 18 polegadas de 800W RMS e 8Ω de impedância. Como pode ser observado na Fig. 7 [2], o sistema de três vias divide as freqüências em três faixas que são:

- 10 – 700Hz para woofer
- 700 – 4kHz para midrange
- acima de 4kHz para tweeter

Desta forma pretende-se projetar um amplificador que possua uma freqüência de corte de 2kHz destinado a suprir os 45% de potência destinada ao woofer. O projeto de um amplificador para midrange deve levar em consideração a freqüência de operação máxima de 4kHz.

O exemplo a seguir é destinado, somente, a alimentação do woofer pelo fato da freqüência de corte de 2kHz não atender toda a faixa de operação do midrange.

A Fig. 7 mostra que 45% da potência de um sistema de som de 3 vias é destinado a alimentação do woofer, 45% é direcionado ao midrange e 10% ao tweeter. Isto posto, é apresentado a seguir as especificações de projeto.

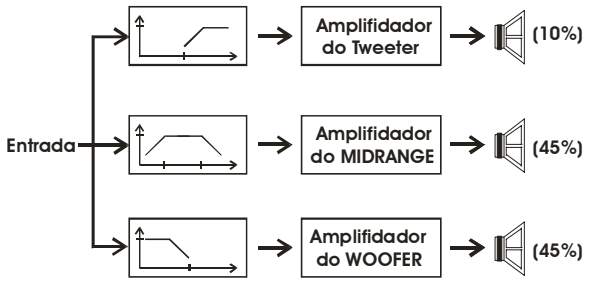

Fig.7: Sistema de auto-falantes ativo (three-way) composto de 3 amplificadores para cada banda.

A Tabela 1 mostra a especificação de projeto de um amplificador chaveado de potência de 800W

Table 1: Especificação de projeto

Po	800W
$Vo_{(pk)}$	114V
$f_{\max}$	2kHz
$I_{L1(pk)} = I_{L2(pk)}$	$\simeq I_{CARGA}$
V_{DC}	144V

A tensão nominal RMS na carga (Vo) é:

$$Vo = \frac{Vo_{(pk)}}{\sqrt{2}} \qquad (9)$$

$$Vo = \frac{114V}{\sqrt{2}} = 80V$$

$$Po = Vo.Io \qquad (10)$$

onde:

Po = Potencia nominal RMS de saída
Vo = Tensão nominal RMS de saída
Io = Corrente nominal RMS de saída

A corrente nominal RMS na carga é dada pela equação (10).

$$Io = \frac{800W}{80V} = 10A$$

A corrente de pico na carga:

$$Io_{(pk)} = 10.\sqrt{2} = 14,14A$$

Usando a Equação (5) C_T :

$$C_T = \frac{14,14A}{2.\pi.2000Hz.114V} = 10\mu F$$

Adotando $C_1 = C_2 = 2.C_S$ a capacitância C_p, C_1 e C_2 é:

$$C_1 = C_2 = \frac{2}{5}.C_T = \frac{2}{5}.10\mu F = 4\mu F$$

$$C_p = \frac{C_T}{5} = \frac{10\mu F}{5} = 2\mu F$$

Finalmente o valor da indutância dos indutores L_1 e L_2 é obtido pela Equação (8).

$$L_1 = L_2 = \frac{\left(144V\right)^2}{4.\pi^2.\left(114V\right)^2.\left(2kHz\right)^2.\left(10,1.10^{-6}F\right)}$$

$$L_1 = L_2 = 1mH$$

4 Estratégia de Controle

O controle é realizado por meio de histerese. A freqüência de chaveamento, portando, opera em uma faixa de freqüência que depende da dinâmica da realimentação do controle. No conversor implementado, a referida freqüência, se encontra acima de 50kHz.

Nos amplificadores convencionais a operação em freqüência fixa propicia vantagens com relação ao projeto do filtro de saída. Por outro lado, a topologia proposta opera com duas fontes de corrente na modulação do sinal amplificado sobre um capacitor de filtro Cp no estágio de saída. Motivo pela qual pode-se obter baixo THD na operação do controle via histerese.

A histerese é implementada por meio de um comparador que recebe em sua entrada inversora uma amostra do sinal modulado sobre o capacitor de filtro Cp,. O sinal a ser reproduzido é aplicado na entrada não inversora do mesmo comparador e o resultado desta comparação é utilizado no acionamento da chave S1 e S2 de forma complementar.

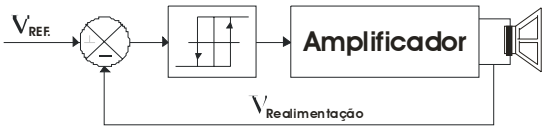

Fig. 8. Diagrama do circuito de controle

4 RESULTADOS EXPERIMENTAIS

A Fig. 9 mostra que o rendimento atinge 92% na condição de 800W de potência de saída com carga de 8Ω e THD de 0,84% (Fig. 10). A potência de 1137W exibida no gráfico da Fig. 9 foi extraída com uma carga resistiva de 5Ω. O teste de 1137W foi possível porque tanto os indutores quanto os MOSFETs utilizados estavam dimensionados acima das especificações de projeto. Por fim, o diagrama de Bode da Fig. 11 mostra o ganho do amplificador em função da freqüência onde fica claro que as especificações de projeto, com relação a freqüência de corte, estão de acordo com os resultados experimentais obtidos.

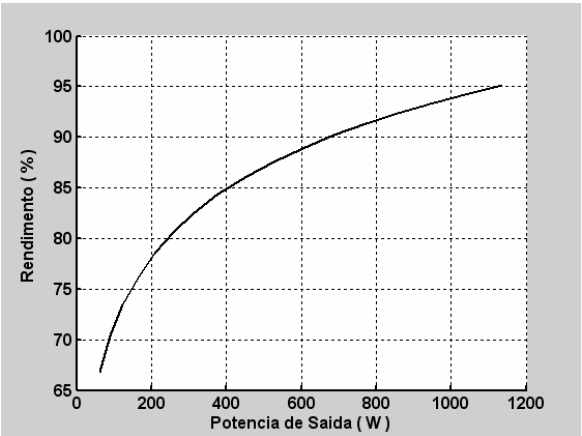

Fig. 9. Curva do rendimento em função da potência de saída

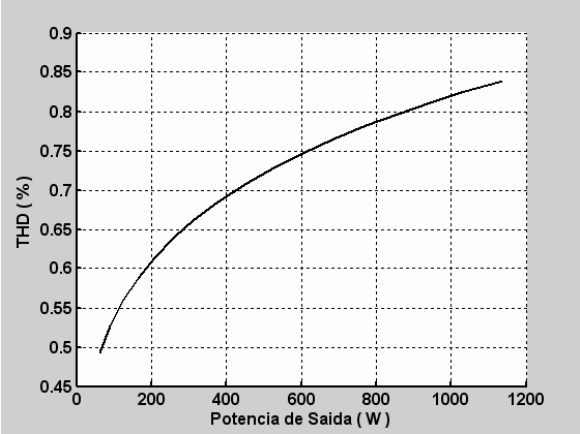

Fig. 10. THD em função da potência de saída

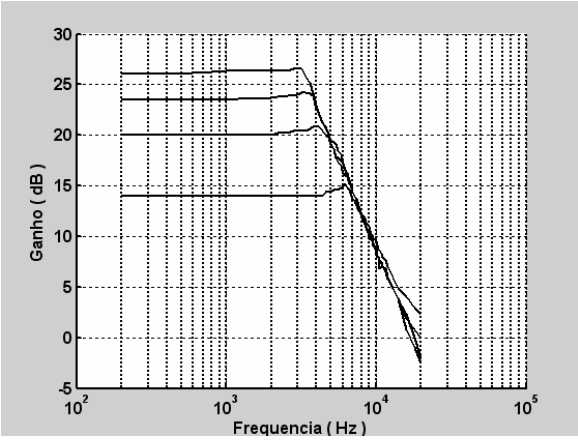

Fig. 11. Curva do ganho em dB em função da freqüência.

5 Conclusão

Este artigo apresentou um novo amplificador de áudio Classe D com rendimento superior a 90%, THD menor que 1% que não necessita de fonte de tensão de alimentação regulada para operar. Outra vantagem que cabe salientar é a não existência de filtro externo que nos amplificadores de áudio convencionais filtra as componentes harmônicas dos pulsos PWM gerados pelo estágio de saída do amplificador.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Chang, M. T. Tan, Z. Cheng and Y. C. Tong. "Analisys and Design of Power Efficient Class D Amplifier Output Stages." IEEE Trans. Circuits and Sustems I: Fundamental Theory and Applications, vol. 47, no. 6, pp. 897-902, 2000.

[2] K. Nilsen "High-Fidelity PWM-Based Amplifier Concept for Active Loudspeaker Systems with Very Low Energy Consumption" J. Audio Eng. Soc., vol. 45, no. 7/8, pp. 554-570, 1997.

[3] F. R. S. Vincenzi, et al., "Amplificador de Áudio Chaveado" SEMEA- Seminário de Engenharia de Áudio, Belo Horizonte - MG. Editora da UFMG, CD-ROM, vol. único, 2002.
[4] F. R. S. Vincenzi, et al., "A New Audio Switched Power Amplifier" INTELEC - Internation Telecommucations Energy Conference, , Copenhege. Proceedings of IEEE INTELEC'1999. Psicataway - NJ - USA: IEEE PRESS. vol. único, pp. 01-06, 1999.
[5] F. R. S. Vincenzi, et al., "A Switched power Amplifier for Audio Application" COBEP - V Congresso Brasileiro de Eletrônica de Potência, Foz do Iguaçu PR. Anais do COBEP'1999. Foz do Iguaçu PR: Imprensa Universitária da UFPR, vol. 02, pp. 521-526, 1999.
[6] F. R. S. Vincenzi, et al., "A new switched Power Amplifier for High Power Applications". INTELEC – Internation Telecommucations Energy Conference Copenhage. Proceedings of IEEE'1999. Psicataway - NJ- USA: IEEE PRESS, vol. único, 1999.

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na IX Convenção Nacional
11 - 13 de Abril de 2005, São Paulo, SP

Este artigo foi reproduzido do original entregue pelo autor, sem edições, correções e considerações feitas pelo comitê técnico deste evento. 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 reservados. Não é permitida a reprodução total ou parcial deste artigo sem autorização expressa da AES Brasil.

Simulação de Som 3D em um Ambiente de Realidade Virtual Imersiva Utilizando HRTF

Leandro Ferrari Thomaz, Marcelo K. Zuffo, João Antonio Zuffo, Regis Rossi A. Faria
LSI – Escola Politécnica da USP
05508-900, São Paulo, SP, Brasil
{lfthomaz, mkzuffo, jazuffo, regis}@lsi.usp.br

RESUMO

Este artigo descreve a implementação de um sistema de auralização 3D em um ambiente de realidade virtual imersiva utilizando-se técnicas bi-aurais. Para avaliar a eficácia desta solução em ambientes do tipo CAVE um programa está sendo desenvolvido para calcular a posição do usuário e da fonte sonora e convoluir o som proveniente desta fonte com respostas impulsivas do tipo HRTF previamente calculadas, sintetizando-se o som em fones de ouvido. O *software* será integrado ao conjunto de soluções para sonorização tridimensional previstas no projeto AUDIENCE.

INTRODUÇÃO

A motivação deste projeto é poder implementar um sistema de auralização 3D de baixo custo, utilizando partes de *software* e medições de HRTF previamente feitas, por ser uma técnica amplamente conhecida e empregada. A integração com o *software* de navegação e síntese desenvolvido na CAVERNA Digital [1], o Jinx [2], é uma premissa fundamental para investigarmos o potencial de aplicação desta técnica em um ambiente de RV imersiva.

CAVERNA DIGITAL

A CAVERNA Digital (CAVE[1] - *Audio Visual Experience Automatic Virtual Environment*) é um sistema de realidade virtual, que consiste de uma sala, na qual cada parede é uma tela de projeção. O observador que se encontra dentro do ambiente experimenta um alto grau de imersão. Este sistema não é intrusivo, e não existem equipamentos pesados para serem carregados pelo usuário.

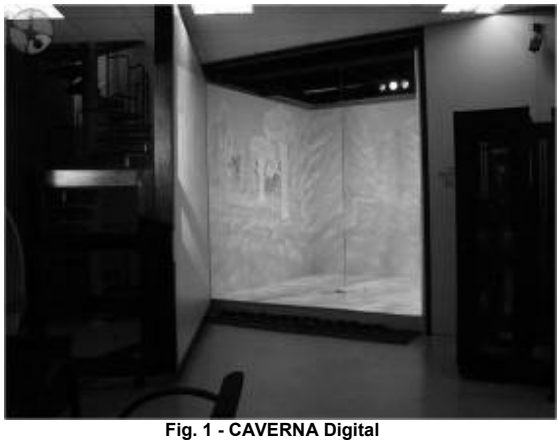

Fig. 1 - CAVERNA Digital

A CAVERNA é composta por um cubo de dimensão 3m x 3m x 3m com projeção nas quatro paredes laterais mais o

[1] CAVE™ é marca registrada da University of Illinois.

piso (fig. 1). O sistema pode comportar até 6 usuários no seu interior, compartilhando a mesma experiência virtual.

Os sistemas cúbicos apresentam atualmente uma qualidade de imersão melhor que qualquer outra forma de sistema de realidade virtual como, por exemplo, os capacetes de realidade virtual HMD (*Head Mounted Display*), que são óculos pesados e apresentam sérios problemas de rastreamento, navegabilidade, campo de visão e podem induzir ao mal estar no usuário. Na CAVERNA Digital os recursos incorporados ao usuário para a imersão são geralmente mínimos, baseando-se apenas em óculos para suportar a estereoscopia e dispositivos de rastreamento.

Uma grande vantagem da CAVERNA Digital é sua flexibilidade, em ter a possibilidade de compartilhamento do mundo virtual por vários usuários simultaneamente, a quantidade de pessoas está limitada diretamente às dimensões físicas da CAVERNA utilizada, como também é possível o recálculo do ponto de vista do usuário, baseado em um sistema de rastreamento para um usuário e assim se produzir o efeito perfeito de imersão e navegabilidade.

Dentre as possíveis aplicações da CAVERNA Digital, destacamos: a visualização de lençóis petrolíferos para otimizar o projeto de perfurações, as maquetes digitais na indústria automobilística para solucionar problemas de *design* e aerodinâmica, a análise meteorológica como a visualização de massas de ar, além de simulações cirúrgicas na medicina, planetário virtual, projetos arquitetônicos, a navegação em ambientes virtuais, como museus e salas de espetáculos, e aplicações avançadas de entretenimento.

No leiaute arquitetônico da CAVERNA Digital (fig. 2), um dos aspectos levados em conta foi embutir totalmente os projetores e outros dispositivos, escondendo do usuário final as particularidades técnicas do sistema.

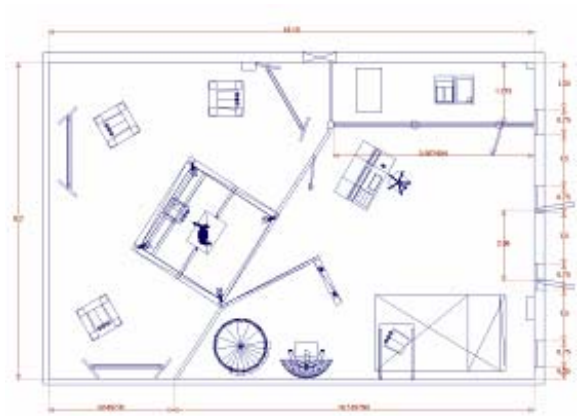

Fig. 2 - Leiaute da CAVERNA Digital

Do ponto de vista estrutural deve-se levar em conta também aspectos como compatibilidade eletromagnética, iluminação, isolamento e tratamento acústico, controle de temperatura e umidade, e finalmente a possibilidade de montagem de periféricos adicionais no interior da CAVERNA Digital, como fones de ouvido. A estrutura do cubo é toda feita em madeira e plástico, envolvendo o uso de metal apenas em alguns parafusos estruturais e projetores, buscando assim minimizar a distorção dos campos magnéticos gerados pelos rastreadores de posição eletromagnéticos. Os pilares e vigas do prédio próximos à CAVERNA foram construídos de madeira, também para evitar interferência no sistema eletromagnético.

Nas CAVEs, em geral, verifica-se que o sistema de áudio é secundário, existindo apenas sistemas estéreo ou com poucos recursos multicanais. Desta forma, não existe uma adequada reprodução de ambiência, localização dos objetos e correta indicação da diretividade dos objetos sonoros.

Com relação à parte de áudio da CAVERNA, está sendo desenvolvido um projeto para a implementação de um sistema de áudio 3D flexível e escalável, o AUDIENCE [3] . Este projeto conta com a incorporação de soluções comerciais e *surround* até desenvolvimentos específicos para sonorização 3D e auralização, prevendo-se a correta geração de campos sonoros espaciais. Embora focado em reprodução multicanal, dentro do escopo do AUDIENCE estão previstas investigações de diversos sistemas, incluindo aqueles bi-aurais com reprodução via fones de ouvido, para se conhecer melhor a adequação a aplicações especificas, e estudar sua usabilidade.

HRTF

HRTF (*Head-Related Transfer Function*), ou função de transferência relativa à cabeça, consiste numa representação matemática da transformação que um som sofre desde a fonte sonora até o ouvido humano, usualmente é expressa segundo uma medição da resposta impulsiva (ou *Impulse Response*, IR) na altura dos ouvidos humanos. Conforme podemos observar na figura 3, a geração de um som x(t) na posição da fonte sonora é ouvido pela cabeça como $x_R(t)$ no ouvido direito e $x_L(t)$ no ouvido esquerdo, ou seja, existe uma diferença entre esses sons. As respostas impulsivas medidas são relativas a cada ouvido, $h_R(t)$ e $h_L(t)$ (fig. 3).

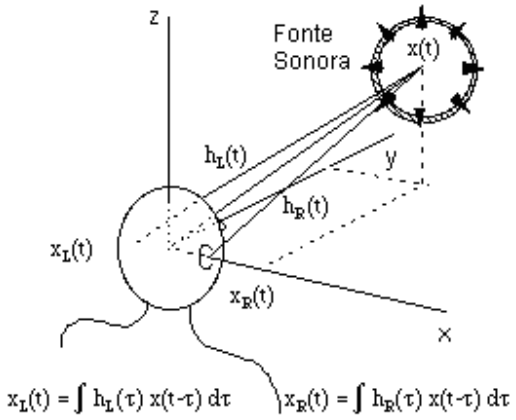

Fig. 3 Medição da HRTF

O experimento para a geração das HRTFs consiste em, para uma dada posição relativa a cabeça, gerar um impulso sonoro (com um estouro, por exemplo) e, através de microfones posicionados nas entradas dos dois ouvidos, gravar este impulso.

Com estas respostas impulsivas, h_R e h_L, gravadas podemos simular como um som seria ouvido pela cabeça, caso fosse gerado naquela dada posição da fonte e recebida naquela posição da cabeça. Isto é possível através da aplicação do teorema da convolução [4], onde o som simulado seria obtido

através da resposta impulsiva convoluida com um som qualquer.

AURALIZAÇÃO NA CAVERNA ATRAVÉS DE HRTF

As respostas impulsivas que são utilizadas neste primeiro trabalho foram obtidas pelo Media Labs do MIT [5]. O trabalho realizado por eles gerou um grupo de arquivos que contém respostas impulsivas para vários ângulos de elevação (de 10 em 10 graus) e rotação (de 5 em 5 graus). Todas as respostas são para o ouvido direito, sendo que para obter do ouvido esquerdo basta utilizar a resposta impulsiva suplementar (por exemplo, se para o direito é utilizada a rotação de 15 graus, para o esquerdo se utiliza a de 165 graus).

Para o cálculo da posição do usuário em relação à fonte sonora utilizamos os dados obtidos do rastreador, que fornece a posição *xyz*, o grau de rotação da cabeça do usuário em relação a um eixo fixo de coordenadas e a posição da fonte sonora (fig. 4).

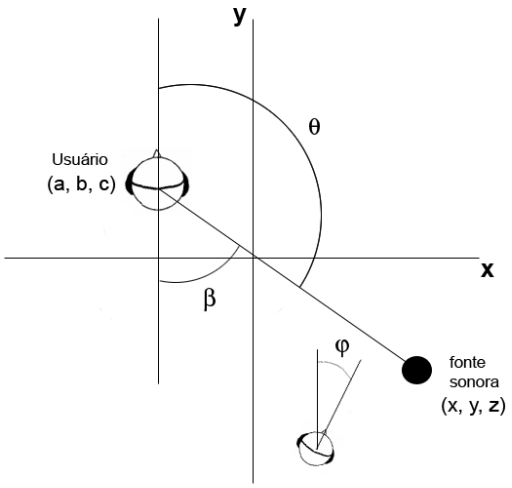

Fig. 4 Cálculo do ângulo θ entre a cabeça e a fonte

O eixo de coordenadas é deslocado para a cabeça do usuário, a fonte sonora é transferida para o 1º quadrante do plano e o ângulo entre a cabeça e a fonte é calculado através da seguinte equação (sendo *ab* as coordenadas da cabeça e *xy* as coordenadas da fonte sonora deslocada):

$$\beta = arctg\left(\frac{|x-a|}{|y-b|}\right) \tag{1}$$

A partir de β, do ângulo de rotação da cabeça φ (obtido através do rastreador) e do quadrante em que se encontrava a fonte originalmente, obtemos o ângulo final θ:

1º quadrante: $\theta = \beta - \varphi$ (2)

2º quadrante: $\theta = 180° - \beta - \varphi$ (3)

3º quadrante: $\theta = 180° + \beta - \varphi$ (4)

4º quadrante: $\theta = 360° - \beta - \varphi$ (5)

Se o ângulo θ for negativo é porque a fonte sonora está à esquerda do usuário. Assim, selecionamos a resposta impulsiva com ângulo de rotação mais próxima do obtido pelo cálculo.

O cálculo do ângulo de elevação γ da elevação é feito através da seguinte equação:

$$\gamma = arctg\left(\frac{|z-c|}{\sqrt{(x-a)^2+(y-b)^2}}\right) \tag{6}$$

A intensidade da fonte sonora (I) é calculada a partir da distância (d) desta em relação à cabeça, utilizando a equação de onda que diz que a intensidade é proporcional ao inverso da distância ao quadrado [6].

$$I \propto \frac{1}{d^2} \tag{7}$$

A auralização é feita a partir da aplicação da operação matemática da convolução. A partir da resposta impulsiva mais próxima da posição da fonte sonora, o som emitido por essa fonte é convoluido com esta resposta impulsiva tanto para o ouvido direito quanto para o esquerdo e o resultado é reproduzido nos fones de ouvido.

$$x_R = x \otimes h_R(\theta) \tag{8}$$

$$x_L = x \otimes h_L(180° - \theta) \tag{9}$$

Neste primeiro experimento, o modelo utilizado leva em consideração algumas simplificações. Não são levados em conta objetos que estão entre a fonte sonora e o usuário, o que causaria uma atenuação, reflexão e/ou difração da onda sonora. Também não foi considerada a reverberação do mundo virtual.

IMPLEMENTAÇÃO

A implementação do sistema envolve a integração do *hardware* disponível na CAVERNA, o *software* utilizado para navegação (Jinx) e um módulo específico de áudio que se encaixa no Jinx. O diagrama de blocos a seguir mostra as partes do sistema (fig. 5).

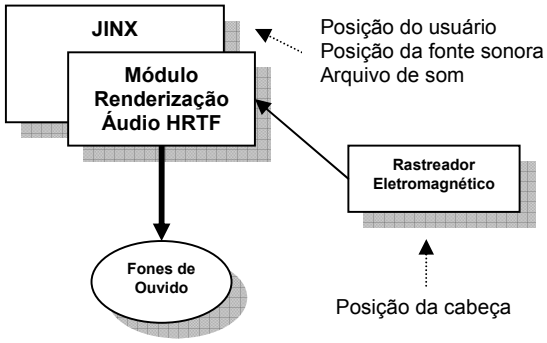

Fig. 5 Diagrama de Blocos do Sistema

Hardware

O *hardware* utilizado para a implementação do projeto está descrito a seguir:

- CAVERNA Digital

A infra-estrutura já disponível da CAVERNA Digital foi utilizada, incluindo um aglomerado computacional (*cluster*), responsável pela geração de imagens e processamento dos dados de áudio. Futuramente, um novo nó será adicionado ao cluster e será responsável exclusivamente pelo processamento do áudio, tornando possíveis simulações mais complexas.

- Rastreador

A CAVERNA já possui um sistema de rastreamento eletro-magnético de marca Ascension Technology, modelo Flock of Birds [7]. Este dispositivo permite rastrear o usuário em todo o espaço da CAVERNA, fornecendo a posição *xyz*, bem como o grau de rotação. A resolução do aparelho é de 0,5mm para a posição e 0,1" para a rotação.

- Fone de ouvido

Para a auralização do som, foi utilizado um fone de ouvido com fio disponível na CAVERNA. O ideal seria a utilização de um fone de ouvido sem fio, que não interfira com o campo magnético gerado pelo rastreador. Com este tipo de fone poderia ser proporcionado ao usuário um maior grau de liberdade, aumentando a imersão. O rastreador foi acoplado ao topo do fone de ouvido e calibrado.

- Placa de som

A placa de som utilizada é responsável por transformar o som digital, gerado pelo *software*, em som analógico e enviar ao fone de ouvido. É importante ressaltar que a placa de som tem que ter baixa latência de forma a não interferir com os cálculos de síntese sonora (que já são extremamente complexos).

Software

O *software* utilizado para a renderização e navegação da CAVERNA Digital é o Jinx. Ele é um navegador de ambientes virtuais distribuído, ou seja, ele foi feito para ser usado em aglomerados de computadores, ou *clusters*. O Jinx é baseado no padrão X3D, dando flexibilidade ao sistema e permitindo que o usuário concentre-se apenas no desenvolvimento do conteúdo.

Está sendo desenvolvido um módulo de *software* que funciona acoplado ao Jinx (conforme figura 5). O modulo é responsável por obter a posição do usuário e de sua cabeça, selecionar a resposta impulsiva que mais se aproxima desta posição com relação à fonte sonora e fazer a convolução do som com a resposta impulsiva.

O módulo tem basicamente duas seções, que são executadas em paralelo, indefinidamente, até que a simulação seja finalizada. Estas seções são descritas a seguir.

A primeira seção é responsável por obter os dados do usuário, calcular a posição relativa entre o usuário e a fonte sonora, selecionar uma resposta impulsiva e, caso seja diferente da utilizada anteriormente, fornecer à segunda seção do *software* a nova resposta impulsiva.

A segunda seção é responsável por fazer a convolução entre o som emitido pela fonte sonora e a resposta impulsiva fornecida pela primeira seção, e por sintetizar o sinal através da placa de som. A convolução é feita sob demanda, utilizando pequenos blocos de amostra do sinal original e convoluindo com a resposta impulsiva. Desta forma, é possível fazer a mudança na resposta impulsiva. A convolução é feita por um software de código aberto chamado BruteFIR [8]. Utilizando extensões dos processadores AMD e Intel, este software consegue fazer os cálculos da convolução em um tempo extremamente baixo.

É muito importante que haja um mecanismo que trave a execução da segunda seção enquanto a resposta impulsiva estiver sendo atualizada.

Abaixo, mostramos o algoritmo básico, em pseudocódigo, de cada seção do módulo de auralização via HRTF proposto.

```
// 1ª Seção
...
enquanto(simulação_rodando) {
        obter_dados_rastreador(pos_xyz, rotacao);
        obter_dados_usuario(pos_xyz);
        obter_dados_fonte_sonora(pos_xyz, wave);
        calcular_pos_relativa(pos_rel);
        selecionar_IR(pos_rel);
        se(IR diferente da anterior) {
                lock();
                substituir_IR();
                unlock();
        }
}
...

// 2ª Seção
...
enquanto(simulação_rodando) {
        selecionar_bloco_wave(bloco, wave);
        convoluir(IR, bloco);
        tocar(bloco);
}
...
```

RESULTADOS

As respostas impulsivas obtidas pelo MIT foram testadas em um programa de processamento de áudio (AudioMulch [10]) com um módulo que faz as convoluções necessárias (Voxengo Pristine Space [11]). A partir destes testes, puderam ser observadas como se comportam estas respostas impulsivas para diversos pares de posições.

Em seguida, iniciou-se o desenvolvimento do *software* responsável pela convolução das HRTFs e integração com o Jinx. Este módulo já foi integrado com o sistema de rastreamento da CAVERNA, bem como ao sistema de navegação. No momento, o programa encontra-se em testes para que sejam feitos os ajustes finos como, por exemplo, a escolha do tamanho do bloco de som a ser convoluido para que o som exibido não tenha interrupções.

CONCLUSÕES

Embora o sistema de auralização não esteja concluído, mas algumas conclusões já podem ser obtidas a partir de testes feitos com partes do *software*.

A utilização das HRTFs, embora seja simples e sem requisitar excessivo custo computacional, não apresenta uma boa correção da localização ou diretividade verificada. Verifica-se que apenas os sons vindos da esquerda, centro e direita podem ser discernidos, enquanto que os sons provenientes de trás, cima e baixo não têm o realismo necessário.

Pensamos que este problema pode ser devido a uma falta de precisão no banco de dados de respostas impulsivas no caso de sons projetados atrás ou acima/abaixo da cabeça. Ou ainda, pelo fato do fone de ouvido ser colocado diretamente

na cabeça, fazendo com que sejam perdidas informações de diretividade dadas, tanto pela orelha humana, como pelo formato da cabeça. Vale lembrar que as medições de HRTF são diferentes para cada pessoa, e que a medição feita pelo MIT em uma cabeça artificial é apenas uma aproximação. Ou seja, o ideal seria, antes da simulação, obter as HRTFs do usuário e executar o sistema com estas.

Para desenvolvimentos futuros, pensamos em obter as nossas próprias respostas impulsivas, mais precisas que as já datadas medidas do MIT. Outra melhoria é a interpolação de HRTFs para posições intermediárias, minimizando o erro na transição entre duas respostas impulsivas. Também devemos incluir ao *software* de auralização a possibilidade de adicionar ao som produzido campos sonoros pré-definidos, como cavernas, catedrais etc.

O problema da baixa sensação de realismo pode ser resolvido utilizando algum sistema de síntese de som 3D mais robusto como, por exemplo, o Ambisonic [12]. Este sistema conta com uma codificação do som numa grade 3D mais elegante e, usualmente, utiliza diversas caixas de som posicionados em volta da CAVERNA, que podem assim gerar um campo de som perifônico. Esta solução está sendo explorada dentro do escopo do projeto AUDIENCE, em curso na CAVERNA Digital.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] *CAVERNA DIGITAL*. Disponível em: http://www.lsi.usp.br/~rv/p/cave_p.html. Consultado em 15/03/2005.

[2] Soares, L. P. and Zuffo, M. K. *JINX: an X3D browser for VR immersive simulation based on clusters of commodity computers*. In Proceedings of the ninth international conference on 3D Web technology, Monterey, California, USA, 79 – 86, 2004.

[3] *AUDIENCE – Audio Immersion Experience by Computer Emulation*. Disponível em http://www.lsi.usp.br/interativos/nem/audience/. Consultado em 15/03/2005.

[4] Oppenheim A., Schafer R., *Discrete-Time Signal Processing*, 2nd ed., Prentice Hall, 1998.

[5] Gardner B., Martin K., *HRTF Measurements of a KEMAR Dummy-Head Microphone*, 1994. Disponível em: http://sound.media.mit.edu/KEMAR.html. Consultado em 15/03/2005.

[6] Everest F., *Master Handbook of Acoustics*, 4th ed., McGraw-Hill, 2001.

[7] *Flock of Birds*. Disponível em: http://www.ascension-tech.com. Consultado em 15/03/2005.

[8] Torger A., *BruteFIR*, 2002. Disponível em: http://www.ludd.luth.se/~torger/brutefir.html. Consultado em 15/03/2005.

[9] Roederer J., *Introdução à física e psicofísica da música*, 1ª ed. 1ª reimpr., São Paulo: EDUSP, 2002.

[10] *AudioMulch Interactive Music Studio*. Disponível em: http://www.audiomulch.com/. Consultado em 15/03/2005.

[11] *Voxengo Pristine Space*. Disponível em: http://www.voxengo.com/pspace/. Consultado em 15/03/2005.

[12] *Ambisonic*. Disponível em: http://www.ambisonic.net/. Consultado em 15/03/2005.

[13] M.A. Gerzon, *Periphony: With-Height Sound Reproduction*, J. Audio Eng. Soc., vol. 21, pp. 2-10 (1973 Jan./Feb.)

[14] Michael Gerzon, *Surround Sound Psychoacoustics*, Wireless World, vol. 80, pp. 483-486 (1974 Dec.)

[15] CRUZ-NEIRA, C.; D.J. SANDIN; T.A. DEFANTI. *Surround-screen projection-based virtual reality: The design and Implementation of the CAVE*. In: SIGGRAPH 1993. ACM SIGGRAPH, Anaheim, Jul. 1993

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na IX Convenção Nacional
11 - 13 de Abril de 2005, São Paulo, SP

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

Granular Synthesis of Sounds through Fuzzyfied Markov Chains

EDUARDO RECK MIRANDA[1], JONATAS MANZOLLI[2], e ADOLFO MAIA JR.[3],
[1]Computer Music Research, Faculty of Technology, University of Plymouth, Plymouth, Devon PL4 8AA, United Kingdom.
[2]Instituto de Artes, Universidade Estadual de Campinas and Núcleo Interdisciplinar de Comunicação Sonora (NICS), (UNICAMP) - 13.081-970 - Campinas (SP), Brazil.
[3]IMECC, Universidade Estadual de Campinas and Núcleo Interdisciplinar de Comunicação Sonora (NICS), (UNICAMP) - 13.081-970 - Campinas (SP), Brazil.
eduardo.miranda@plymouth.ac.uk; jonatas@nics.unicamp.br; maia@ime.unicamp.br

ABSTRACT

In this paper we introduce a new model for granular synthesis using Markov Chains and Fuzzy Sets. Whereas Markov Chains are used to control the evolution of the sound in time, Fuzzy Sets are employed to define the internal structure of the sound grains. It is shown also how the fuzzy structure of grains changes the markov Process. We provide the mathematical foundations of the model and briefly discuss how we have implemented it in a MATLAB program named Fuzzkov 1.0.

INTRODUCTION

Granular synthesis [6] is commonly known as a technique that works by generating a rapid succession of tiny sounds, metaphorically referred to as sound grains or yet as microsounds [7, 8]. Granular synthesis is widely used by musicians to compose electronic or computer music because it can produce a wide range of different sounds, but it also has been used in speech synthesis [4, 5]. Clearly a discussion about musical aesthetics arise from these developments and although it is a very interesting topic by itself we will not deal with these matters in this paper. A good account of the aesthetics of microsound can be found in reference [9].

Granular synthesis is largely based upon Dennis Gabor idea of representing a sound using hundreds or thousands of elementary sound particles [3].

In this work we take Curtis Roads' definition of sound grain as a point of departure to develop a formal but flexible granular synthesis model. The model uses stochastic processes, namely Markov Chains with Transition Probability Matrix modulated by Membership Functions of the grains with values in the interval $[0, 1]$, which gives the grains their fuzzy characteristics. Thus, we propose a new method for controlling the grains by intertwining Stochastic Processes and Fuzzy Set Theory, where the content of the grains (or internal variables) can change their transition probabilities between states. For the sake of clarity, we have chosen a very simple State Space to introduce the model, where each grain is a state of a Grain Vector $\mathbf{G}$. Therefore, the membership functions in

this case modulate the transition probabilities between states (i.e., grains), changing their ordering position in the time domain. This paper introduced just one of several possible modes of interaction between internal and external control variables.

FUZZY GRAIN AND ITS MATRIX REPRESENTATION

Let us denote Ω the space of all possible oscillators, that is the *frequency* $\times$ *amplitude* space of the ordered pair (ω, a), where the variables ω and a varies in some suitable real intervals. Ω is referred to as a Parameters Space. We define, formally, a grain as a finite collection of points $\{(\omega_i(t), a_i(t)), i = 1, 2, \ldots, N\}$ in Ω, which is taken here as a state of a Markov Chain. A grain can be described by its Fourier Partials inside a real interval I. Its spectral content can be written, without loss of generality, as

$$G(t) = \sum_{n=1}^{N} a_n sin[2\pi\omega_n t + \delta_n], \tag{1}$$

where a_n, ω_n, δ_n reads for amplitude, frequency and a possible phase, respectively.

In granular synthesis a sound can be viewed as a quick stream of grains which, from a geometrical point of view, describes a trajectory in the Ω space.

A grain g_i with r Fourier Partials can be read as a $2 \times r$ matrix:

$$g^i = \begin{bmatrix} \omega_1^i & a_1^i \\ \omega_2^i & a_2^i \\ \vdots & \vdots \\ \omega_r^i & a_r^i \end{bmatrix} \tag{2}$$

Now, a fuzzy grain can be represented as a 3 column matrix

$$G^i = \left[\begin{array}{cc|c} \omega_1^i & a_1^i & \alpha_1^i \\ \omega_2^i & a_2^i & \alpha_2^i \\ \vdots & \vdots & \vdots \\ \omega_r^i & a_r^i & \alpha_r^i \end{array}\right] \tag{3}$$

where we have introduced a third column with the membership frequency and amplitude values of each partial of the grain G^i. Note that g^i is a particular case of G^i for $\alpha_1^i = \alpha_2^i = \ldots = \alpha_r^i = 1$.

MARKOV PROCESSES FOR FUZZY GRAINS

Fuzzy sets, first proposed by Lofti Zadeh [11] are able for handling uncertainty, imprecisions or vagueness. Below we show how the membership functions of fuzzy grains can modify the Markov Transition Matrix and so we get a fuzzy control for the Markov Chain. For a good account of Fuzzy Sets the reader is refereed to [1, 2]. Let us consider a grain described by its Fourier-like equation (1). Each subset of points in Ω represents a grain with particular Fourier partials, that is, it is a sum of basic sinusoidal frequencies.

With the above defined matrices G^i, it is possible to define an unambiguously time evolution of grains through out Markov Chains. This is usually accomplished through a Fuzzy Transition Table, constructed as follows: firstly, suppose that we have a transition matrix for ordinary grains, that is, with no membership vector yet defined. This can be written as follows:

$$\left[\begin{array}{c|cccc} & g^1 & g^2 & \ldots & g^{\mathrm{N}} \\ \hline g^1 & p^{11} & p^{12} & \ldots & p^{1\mathrm{N}} \\ g^2 & p^{21} & p^{22} & \ldots & p^{2\mathrm{N}} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ g^{\mathrm{N}} & p^{1\mathrm{N}} & p^{2\mathrm{N}} & \ldots & p^{\mathrm{NN}} \end{array}\right] \tag{4}$$

which can be viewed as a function

$$\begin{array}{rccl} p: & \mathbf{g} \times \mathbf{g} & \longrightarrow & [0,1] \\ & \left(g^i, g^j\right) & \longmapsto & p\left(g^i, g^j\right) = p^{ij} \end{array}$$

Now, we define a Fuzzy Extended Probability Transition Matrix (or simply Fuzzy Transition Matrix) Q : $\mathbf{G} \times \mathbf{G} \longrightarrow [0,1]$ as

$$Q^{ij} = Q\left(G^i, G^j\right) = \Phi^{ij} * p^{ij} \tag{5}$$

where the symbol $*$ means a matrix operation (e.g., a scalar product, a matrix product or any other well defined operation). The function Φ^{ij} is generated as a finite number of applications of the following basic operations of fuzzy sets: for $i, j = 1, 2, \ldots, N$, we define

1. $$\phi^{ij} = \max_{1 \le k \le r} \left\{\alpha_k^i, \alpha_k^j\right\}, \tag{6}$$

   where α^i and α^j are the membership vectors of the grains G^i and G^j respectively.

2. $$\phi^{ij} = \min_{1 \le k \le r} \left\{\alpha_k^i, \alpha_k^j\right\}, \tag{7}$$

   where α^i and α^j are the membership vector of the grains G^i and G^j respectively.

3. $$\alpha_c^i = 1 - \alpha^i. \tag{8}$$

These result in a product like $\Phi^{ij} = \phi_1^{ij}\phi_2^{ij}\ldots\phi_l^{ij}$, where the third operation above can be performed on any product of α_i vectors. These are basic operations on Fuzzy Sets. See Diamond and Kloeden [1] for a introduction to Fuzzy Sets and their metrics. Note that since the membership function modulates the probability values p^{ij}, the condition for the probability sum $\sum_{j=1}^{N} Q^{ij} = 1$ can be violated. In order to solve this problem we renormalize the matrix Q^{ij} as follows. Denoting $q_i = \sum_{k=1}^{N} Q_{ik}$ we define the elements of matrix $\mathbf{P}$ as

$$P_{ij} = Q^{ij}/q^i \qquad i, j = 1, 2, \ldots, N \tag{9}$$

Now the probability property $\sum_{j=1}^{N} P_{ij} = 1$ is clearly satisfied. The above definition shows that the internal fuzzy content of the grains have a weight (through the function Φ^{ij}) for their transition to a next state of the Markov Chain.

The Fuzzy Transition Matrix (or Table) now reads

$$\begin{array}{c|cccc} & G^1 & G^2 & \dots & G^{\mathrm{N}} \\ \hline G^1 & P^{11} & P^{12} & \dots & P^{1\mathrm{N}} \\ G^2 & P^{21} & P^{22} & \dots & P^{2\mathrm{N}} \\ \dots & \dots & \dots & \dots & \dots \\ G^{\mathrm{N}} & P^{\mathrm{N}1} & P^{\mathrm{N}2} & \dots & P^{\mathrm{NN}} \end{array} \tag{10}$$

In this simple model a transition from one state to another corresponds to a jump from a particular grain to another in the grain vector **G**. In adition the fuzzy content of a grain, that is, its membership vector, can have a significant weight on the probability transition. Since the process is finite, a criterium to halt the process is needed here. This will be discussed in the next section.

The above model is suitable for several kinds of matrix operations on internal as well external variables controlling the grains behaviour in time. There is plenty of room for the definition of a great number of different methods to generate and control the grains. We present one of such methods below.

CONTROL OF GRAIN STREAMS

There exist many different ways (algorithms) to control the evolution of the grains in time. We show here one by using the so called Hausdorf Metric which is suitable to measure distance between sets (grains are finite and discrete subsets of Ω). Its definition is as follows [1].

Suppose that the space $\Omega = \mathbb{R}^{2N}$ has a metric $d(x, y)$. Let x be a point in Ω and A a nonempty subset of Ω. We define the distance of the point x to the set A as:

$$\delta(x, A) = inf\left\{d(x, y), y \in A\right\}. \tag{11}$$

The Hausdorff separation of a set B from a set A is defined by

$$\Delta(B, A) = sup\left\{d(y, A), y \in B\right\} \tag{12}$$

In general, Δ is not symmetric, that is $\Delta(A, B) \neq \Delta(B, A)$. In order to get a symmetric one we define the so called Hausdorff distance by

$$d_H\left(A, B\right) = max\left\{\Delta(A, B), \Delta(B, A)\right\} \tag{13}$$

With this distance function (Ω, d_H) is a Metric Space.

Time evolution can be better controlled using a fuzzy metric that takes into account the degree of membership of the Fourier partials inside each grain. In other words, partials with low membership coefficients contribute little for the Hausdorff distance measure between the grains. Membership vectors define the fuzzy character of the grains, or in a musical jargon, their weighted harmonic content. A metric control is closely related to the notions of approximation and/or the maximal time (or number of steps) available to run a process. Below we indicate three stop criteria we devised to halt a Markov Chain in our model of granular synthesis.

Halting Criteria

1. *Convergent Type*: If the distance between the last generated grain and a fixed grain (target) is smaller than a prefixed arbitrary number ϵ, the process halts.
2. *Cauchy Type*: If the distance between two states is smaller than ϵ the process halts.
3. *Maximal Number of Steps Type (MNS)*: Fix the maximum number of steps for the process to halt.

Any of the above criteria can be used to halt the process. Of course *Maximal Number of Steps Type* is the simplest one, since no metric is required. In our program *Fuzzkov 1.0* we have implemented fully the MNS and partially, the Cauchy type, at the Hausdorff Metric level, but not at the Fuzzy Metric level. We have implemented the Hausdorff Metric as an inequality, so that FuzzKov 1.0 runs in loops until it is satisfied. We obtained good results for both controls of the grains streams working together.

In adition we can also specify a number of different formal settings to update the internal content of the grains at each step of the Markov Chain. This provides the means to control the evolution of the macrosound in the Ω Space. Below we indicate two updating methods.

Grains Updating

1. *No Updating: no change in the internal content*

   In this case, each grain G_i corresponds to a state of the grain vector **G** and no operation is applied to the internal structure of the grains. Nevertheless this procedure takes into account the fuzzy nature of grains as the role of the membership vectors is to produce new arrangements in time (that is, permutations) for the prefixed grains.

2. *Core Merged Grains*

   In this case we update l-th step grain as a subset merging l previous grains of the Markov Chain; e.g., $G^0, G^1, \dots, G^{l-1}$. For the sake of clarity, we ignored all the other subindexes. Define the *l-Mean Frequency* as

$$\overline{\omega}^{(l)} = \sum_{k=1}^{r} \frac{\omega_k^0 + \omega_k^1 + \dots + \omega_k^{l-1}}{l} \tag{14}$$

   and take the r closest frequencies from the set $U_l = \bigcup_{k=0}^{l-1} G^k$ to the mean frequency $\overline{\omega}^{(l)}$ and update G^l (with the same letter) as

$$G^l = \left[\left(\omega_{k_1}, a_{k_1}\right), \left(\omega_{k_2}, a_{k_2}\right), \dots, \left(\omega_{k_r}, a_{k_r}\right)\right].$$

This procedure leads to a concentration of frequencies within a narrow bandwidth, but with a large bandwidth for the amplitudes. The *halting criterion* here can be taken as the *Cauchy type*. Given an arbitrary (but small) number ϵ, the process stops if $d_H\left(G_i, G_{i+1}\right) \leq \epsilon$, where the distance between two points used for defining the above Hausdorf Distance is given by, for example:

$$d\left(\left(\omega_i, a_i\right), \left(\omega_j, a_j\right)\right) = \max_{1 \leq k \leq r} \left|\omega_i - \omega_j\right|. \qquad (15)$$

If we fix a particular grain in the Ω space, such as $\overline{G}$, we can consider the *Convergent halt criterion* , that is the proceess stops if $d_H\left(G_i, \overline{G}\right) \leq \epsilon$.

We can also take the *mean frequency* only of the last m grains and so it reads as

$$\overline{\omega}^{(l)} = \sum_{k=1}^{r} \frac{\omega_k^{l-m} + \omega_k^{l-m+1} + \ldots + \omega_k^{l-1}}{m} \qquad (16)$$

and take the r closest frequencies to $\overline{\omega}^{(l)}$ from the set $U_l = \bigcup_{k=l-m}^{l-1} G^k$. Clearly, for $m = l$ we get the previous model.

A short description of the Diagram of Fuzzkov 1.0 (Fig 1) is as follows. The grains are generated by randomly (uniform and gaussian) 3-dimensional matrices A with dimensions $2 \times r \times N$ which include r normalized frequencies and amplitudes for N grains (Fourier Partials). We have taken the uniform as well the Gaussian distribution of probability to gerated them. From this we get a Matrix $B(2, 1, N)$ with the sum of Fourier Partials for the N grains. A Markov transition Matrix $p(N, N)$ is generated and modified by a Membership Matrix $Memb(N, N)$. A number of diferent operations are available to do this modification. So we get a fuzzyfied Markov Matrix $Q(N, N)$ which operates on an aaray of probabilities vectors $u(n+1, N)$. Next, a particular filter choose the index of the maximal value of each probability vector $I(1, n+1)$. Finally, the program reorder the Grain matrix $B(2, 1, N)$ along the index vector $I(1, n+1)$ and produce the sound as well other outputs for analysis.

EXAMPLES AND ANALYSIS OF THE RESULTS

We have implemented a prototype of our model using Matlab in which Membership Matrices modulate a Transition Probability Matrix of a Markov Chain, but the internal content of the grains are not changed during the Process. Thus, our model can be thought of as a *Coarse Grain Fuzzy Synthesis*. We have used the MNS and Cauchy criteria, by using the Hausdorff Metric on the Grain Space, in order to halt the process. Weakly convergent process have lead to rich varities of timbre along sound streams. This is because the system has time enough to explore the possibilities during the Markov Process. We have in adition included some special effects commonly used in granular syntehsis such as modulation through time windows for the grains, which avoids glitches and, in the macro scale, we included crescendo and decrescendo effects. In fact this can be done more generally by using an modulation function with an arbitary number of peaks and regions with increasing as well decreasing rates.

We have normalized all the sound signals, so they are more suitable to analysis and comparison. After record the digital signal as a *wav.file* the program has three outpus: *sound stream, spectrogram* and *plot of the probability vector evolution* from MATLAB. In adition,in order to analyse in a easy way our results, we used the WAVELAB program to get a *3D-analysis* in a time $\times$ frequency space.

CONCLUSION

We have presented a model for granular synthesis as a Markov Chain in which each grain is a possible state of a Grain Vector **G**. A major feature of our model is that the spectral components of the grains are coupled with the state transition probability through grain's membership vectors. This allows the user more flexibility in a higher level as well more variability to control the sequence of grains of the Markov Chain. We have implemented a computer algorithm named *Fuzzkov 1.0*, written in MATLAB, in which the membership functions modulate the Transition Probability Matrix. Nevertheless the internal contents of the grains are not changed by them in this first version of Fuzzkov. In this way, the present model can be understood as a *Coarse Grain Fuzzy Synthesis*. A more complex program should allow the use of Fuzzy Functions to emphasize some particular components of the spectral content of the grains and then to drive the sound flow by updating them at each step through mergin and selecting the most representative frequencies and amplitudes to construct the next grain of the stream. As Halt Criteria we used the Maximal Number of Steps, as well the Cauchy type with the Hausdorff Metric between grains. A model in which the states of the Markov Chain are related to grains's subsets (*Fine Grain*), as well an effective use of Fuzzy Metrics, will be presented elsewhere.

A distintive point of our approach is its flexibility of macro manipulation by fuzzy matrices parameters. At present time we can consider this control as a toy model. Nevertheless it has a great potential to include new aspects of the fuzzy approach. Also all effects are included in the algorithm and so no external device is needed. Our experiments with *Fuzzkov 1.0* has shown that its audio output is comparable with the most recent granular synthesizers and in adition, depending on the inputs parameters it can provide surprisingly new sounds

ACKNOWLEDGMENTS

This work was supported by CAPES (Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior), Brazil, under Project 0205-04-4 and FAEP (Fundação de Apoio à Pesquisa e ao Ensino), UNICAMP. AMJ is

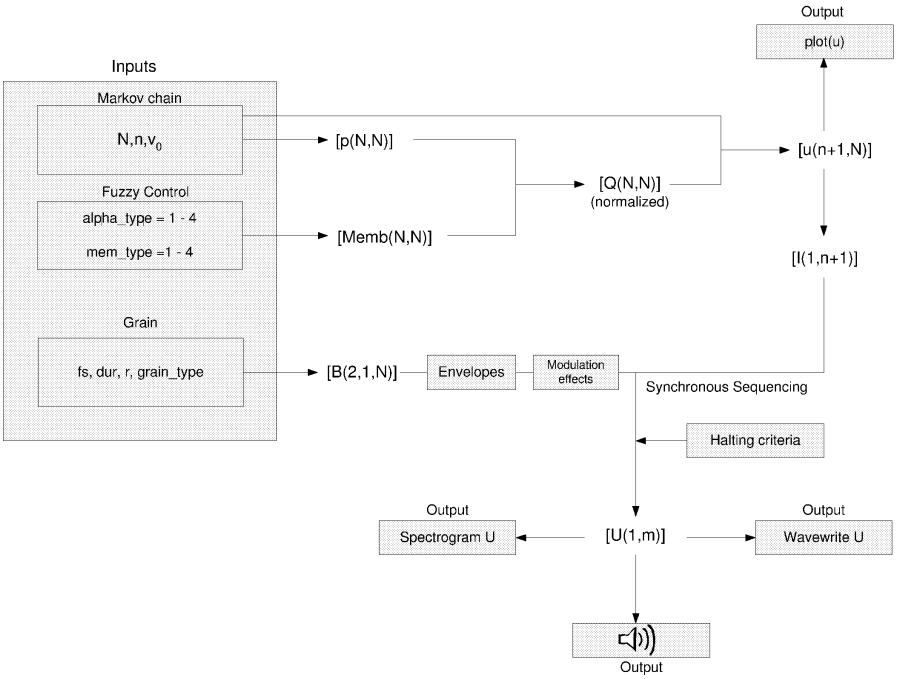

Fig. 1: Diagram of *FuzzKov 1.0*

grateful to the School of Communications, Electronics and Engineering of University of Plymouth for the hospitality during his sabbatical.

REFERENCES

[1] P. Diamond and P. Kloeden, *Metrics of Fuzzy Sets: Theory and Applications*, World Scientific, (1994).

[2] A. Kandel, *Fuzzy Techniques in Pattern Recognition*, John Wiley Sons, (1982).

[3] D. Gabor, *Acoustical Quanta and the Theory of Hearing*, Nature**159** (4044), 591-594,(1947).

[4] E. R. Miranda, *Generating Source Streams for Extralinguistic Utterances*, Journal of the Audio Engineering Society, 50(3):165-172, (2002).

[5] E. R. Miranda, *Computer Sound Design: Synthesis techniques and Programming*, Oxford: Focal Press (2002).

[6] C. Roads, *Introduction to Granular Synthesis*, Comp. Mus. Jour. **12**(2), 11-13 (1988.)

[7] C. Roads, *Computer Music Tutorial*, MIT Press, Cambridge, MA (1996).

[8] C. Roads, *Microsound*, MIT Press, Cambridge,MA, (2001).

[9] P. Thomson, *Atoms and errors: towards a history and aesthetics of microsound*, Organized Sound **9**(2), 207-218, (2004).

[10] B. Truax, *Real Time Granular Synthesis with a DSP Computer*, Computer Music Journal, Vol. 2, No. 2, pp. 14-26, (1988).

[11] L.A. Zadeh, *Fuzzy Sets*, Informat. Control, **8**, 338-353, (1965).

Revisão

André Luís Dalcastagnê
Antonio Carlos Moreirão de Queiroz
Fábio Pacheco Freeland
Fernando Antônio Pinto Barúqui
Fernando Santana Pacheco
Filipe Castello da Costa Beltrão Diniz
Joarez Bastos Monteiro
Leandro Ferrari Thomaz
Leonardo Gomes Baltar
Miguel Arjona Ramírez
Monique Virório Nicodem
PauloAntonio Andrade Esquef
Regis Rossi Alves Faria
Ronaldo de Freitas Zampolo
Rosalfonso Bortoni
Sergio Lima Netto
Sidnei Noceti Filho
Tadeu Nagashima Ferreira

Créditos

Coordenação Geral:

João Américo (Presidente-AES Brasil)

Coordenação Técnica:

Luiz Wagner Pereira Biscainho (UFRJ)

Comissão Técnica:

Marcelo Knörich Zuffo - USP
Regis Rossi Alves Faria - USP
Rosalfonso Bortoni - Clever Tecnologia
Rui Seara - UFSC
Sergio Lima Netto - UFRJ
Sidnei Noceti Filho - UFSC