Anais do II Congresso Brasileiro de Engenharia de Áudio da AES-Brasil, apresentado na VIII Convenção Nacional da AES-Brasil

Centro de Convenções Blue Tree Towers
São Paulo, SP
7 a 9 de junho de 2004

COPPE/Poli/UFRJ

X

sumário

Apresentação

Créditos

Trabalhos

Autores

apresentação

Sejam todos bem vindos à VIII 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 II 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 um marco memorável na história da Sociedade de Engenharia de Áudio – AES Brasil. Todos nós, Comissão Organizadora, Comissão Técnica, revisores e ainda os autores dos artigos técnicos submetidos ao Congresso, fizemos o máximo para alcançar essa meta.

Eng. Joel Brito - Coordenador Geral

Prof. Luiz Wagner Pereira Biscainho - Coordenador Técnico

créditos

II Congresso Brasileiro de Engenharia de Áudio da AES Brasil

São Paulo, 5 a 7 de junho de 2004

Coordenação Geral:

Joel Brito (AES-Brasil)

Coordenação Técnica:

Luiz Wagner Pereira Biscainho (UFRJ)

Comissão Técnica:

Marcelo Knörich Zuffo (USP)

Regis Rossi Alves Faria (USP)

Rosalfonso Bortoni (Studio R)

Rui Seara (UFSC)

Sergio Lima Netto (UFRJ)

Sidnei Noceti Filho (UFSC)

créditos

Relação de Revisores:

André Luís Dalcastagnê

Antonio Carlos Moreirão de Queiroz

Fernando Antônio Pinto Barúqui

Fernando Santana Pacheco

Filipe Castello da Costa Beltrão Diniz

Joarez Bastos Monteiro

Leandro Ferrari Thomaz

Marcelo Knörich Zuffo

Phabio Junckes Setubal

Regis Rossi Alves Faria

Sergio Gilberto Rodríguez Soria

Sergio Lima Netto

Sidnei Noceti Filho

Solimar de Souza Silva

Uma Revisão sobre Metodologia para Avaliação Subjetiva de Componentes de Áudio.

Chistian G. Herrera

Sistema Adaptativo para Caracterização de Alto-Falantes.

Rafael C. Haical; Fábio P. Itturrie; Márcio H. Costa; Delmar B. Carvalho

Análise de Viabilidade da Aplicação de Filme Fino Metálico em Cones e Domos de Alto-Falantes.

Hugo D'A. Jucosky; Roberto K. Yamamoto; Walter Ullmann

Impermeabilização de Cones de Papel de Alto-Falantes com Plasma de CF4/H2.

Hugo D'A. Jucosky; Roberto K. Yamamoto

Estado da Arte da Classificação de Sinais de Áudio.

Jayme G. A. Barbedo; Amauri Lopes

Interpolação de Sinais de Áudio Usando Polinômios do Par de Linhas Espectrais

Paulo A. A. Esquef

Interpolação Bilinear Generalizada de HRTFs para Geração de Som Tridimensional.

Fábio P. Freeland; Luiz W. P. Biscainho; Paulo S. R. Diniz

Controle em Tempo Real da Temperatura da Bobina e do Deslocamento do Cone de Alto-falantes para Operação em Alta Potência.

Constâncio Bortoni; Sidnei Noceti Filho; Rosalfonso Bortoni; Rui Seara

autores

Barbedo, Jayme G. A.

Estado da Arte da Classificação de Sinais de Áudio.

Biscainho, Luiz W. P.

Interpolação Bilinear Generalizada de HRTFs para Geração de Som Tridimensional.

Bortoni, Constâncio

Controle em Tempo Real da Temperatura da Bobina e do Deslocamento do Cone de Alto-falantes para Operação em Alta Potência.

Bortoni, Rosalfonso

Controle em Tempo Real da Temperatura da Bobina e do Deslocamento do Cone de Alto-falantes para Operação em Alta Potência.

Carvalho, Delmar B.

Sistema Adaptativo para Caracterização de Alto-Falantes.

Costa, Márcio H.

Sistema Adaptativo para Caracterização de Alto-Falantes.

Diniz, Paulo S. R.

Interpolação Bilinear Generalizada de HRTFs para Geração de Som Tridimensional.

autores

Esquef, Paulo A. A.

Interpolação de Sinais de Áudio Usando Polinômios do Par de Linhas Espectrais.

Freeland, Fábio P.

Interpolação Bilinear Generalizada de HRTFs para Geração de Som Tridimensional.

Haical, Rafael C.

Sistema Adaptativo para Caracterização de Alto-Falantes.

Herrera, Chistian G.

Uma Revisão sobre Metodologia para Avaliação Subjetiva de Componentes de Áudio.

Itturri, Fábio P.

Sistema Adaptativo para Caracterização de Alto-Falantes.

Jucosky, Hugo D'A.

Análise de Viabilidade da Aplicação de Filme Fino Metálico em Cones e Domos de Alto-Falantes.

Impermeabilização de Cones de Papel de Alto-Falantes com Plasma de CF4/H2.

Lopes, Amauri

Estado da Arte da Classificação de Sinais de Áudio.

Noceti Filho, Sidnei

Controle em Tempo Real da Temperatura da Bobina e do Deslocamento do Cone de Alto-falantes para Operação em Alta Potência.

Seara, Rui

Controle em Tempo Real da Temperatura da Bobina e do Deslocamento do Cone de Alto-falantes para Operação em Alta Potência.

Ullmann, Walter

Análise de Viabilidade da Aplicação de Filme Fino Metálico em Cones e Domos de Alto-Falantes.

Yamamoto, Roberto K.

Análise de Viabilidade da Aplicação de Filme Fino Metálico em Cones e Domos de Alto-Falantes.

Impermeabilização de Cones de Papel de Alto-Falantes com Plasma de CF4/H2.

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na VIII Convenção Nacional
7-9 de Junho de 2004, São Paulo, SP, Brasil

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.

Uma Revisão Sobre Metodologias Para Avaliação Subjetiva de Componentes de Áudio

Christian Gonçalves Herrera
UFMG – Programa de Pós Graduação em Engenharia Elétrica
Belo Horizonte, Minas Gerais, Brasil
herrera@cpdee.ufmg.br

RESUMO

Neste trabalho é discutida a necessidade da execução de testes auditivos subjetivos para a avaliação de componentes de áudio. A abordagem do tema sob um ponto de vista científico é defendida e considerada de vital importância na obtenção de resultados confiáveis e significantes. Ao final, um exemplo ilustra o caso típico onde existem discrepâncias entre a avaliação objetiva e a avaliação subjetiva.

INTRODUÇÃO

A aplicação final da maioria dos sistemas de áudio é a audição de um programa (e.g.: voz, música), o que consiste numa experiência altamente subjetiva. Muito embora o projeto destes sistemas lance mão de técnicas de engenharia essencialmente objetivas, para a avaliação do produto final sempre se faz necessário um procedimento que leve em consideração os aspectos subjetivos envolvidos na audição.

A validade dos resultados obtidos numa avaliação subjetiva de componentes de áudio depende de quatro fatores: a) as condições do ambiente onde os testes são realizados; b) o tipo do programa a ser executado pelos equipamentos em análise; c) as condições individuais dos ouvintes participantes do teste; d) a qualidade dos eventuais componentes utilizados como referência. A legitimidade do teste só é garantida quando se exerce o máximo controle sobre estes fatores.

O nível de interesse da comunidade do áudio neste tipo de avaliação, e consequentemente a abordagem seguida, apresenta três perfis típicos: a) os revisores de produtos e algumas revistas especializadas, que freqüentemente se munem de metodologias questionáveis do ponto de vista científico e são assim levados a tecer críticas pouca fundamentadas sobre um determinado produto; b) o mercado consumidor, que busca por métodos padronizados de avaliação para orientar sua decisão na compra de equipamentos que conformam suas necessidades; c) a comunidade científica, que acredita na correlação entre as

características audíveis e os parâmetros objetivos, e trabalha na busca de metodologias que considerem os limites da audição humana e o estado da arte nas técnicas de medição.

O objetivo deste artigo é discutir as metodologias científicas desenvolvidas nos últimos vinte anos para avaliação subjetiva de componentes de áudio. São abordados procedimentos a serem tomados para maximizar o controle sobre os diversos fatores que vêm a influenciar a percepção dos ouvintes. Ao final, um exemplo de teste é apresentado onde é avaliado o codificador de áudio digital para compressão com perdas denominado MPEG Layer 3 (MP3).

AVALIAÇÕES SUBJETIVAS

Existem muitas razões que justificam a necessidade e a utilidade dos testes subjetivos para avaliação de componentes de áudio. A principal delas refere-se ao fato de que nem todas as características audíveis podem ser objetivamente medidas e correlacionadas com o que é ouvido, com as atuais ferramentas disponíveis. Este fato não implica de forma alguma que os resultados de medições objetivas sejam irrelevantes ao caracterizar um componente. Ao contrário, há muita informação a ser extraída destas medições, principalmente se forem realizadas de maneira mais completa e adequada do que vem sendo feito.

De fato, podem ser questionáveis os motivos que levam um pesquisador a investigar diferenças audíveis tão aparentemente insignificantes entre componentes que fatalmente não seriam percebidas em situações reais. Contudo, existe o argumento de que a cadeia de equipamentos pelos quais trafega o sinal de áudio pode ser tão complexa que o acúmulo de imperfeições ao longo dos diversos equipamentos degrada o sinal a ponto de tornar inaceitável o resultado final. Também existe a teoria que o homem é capaz de detectar *qualquer* diferença audível entre dois sinais, se lhe for dado o tempo suficiente [1].

Embora cada indivíduo possua sua própria opinião e preferência, esta opinião só terá sentido para alguém se contiver elementos claros de objetividade, consistência e repetitibilidade. Tais elementos só podem ser conseguidos através da prevenção de variações indesejadas de parâmetros que venham a influenciar o resultado, e da utilização de referências absolutas sempre que possível de modo que a exatidão possa ser verificada objetivamente, e não apenas adivinhada.

Alguns requisitos para garantir a validade dos resultados de avaliações subjetivas devem ser seguidos [2, 3, 4]:

a) atenção na interface entre dois equipamentos. O adequado casamento de impedância evita sobrecargas e, consequentemente, alterações na resposta em freqüência de um sistema. Sempre que possível a entrada do equipamento em teste (EET) deve ser excitada por uma fonte de baixa impedância de saída (*buffer*);
b) diferenças lineares mensuráveis devem ser eliminadas antes que qualquer conclusão seja tomada a respeito de distorções originadas por não linearidades. Estas diferenças referem-se principalmente à anomalias na resposta em freqüência do EET, e que podem ser prontamente corrigidas através de filtragens adequadas (equalização). Também são fontes de diferenças lineares a resposta de fase, a polaridade e, finalmente, o ganho do EET. Experimentos descritos em [2] revelam que diferenças de apenas 0,2 dB entre oitavas adjacentes podem ser percebidas pelo aparelho auditivo humano. Por outro lado, diferenças de mais de 1 dB podem ser facilmente produzidas na saída de amplificadores de potência, devido à grande variação da impedância dos alto falantes com a freqüência. Embora estas diferenças sejam efetivamente audíveis, elas não se referem a artefatos gerados a partir de comportamento não linear do EET;
c) focalizar o teste na detecção de similaridade ou diferença entre dois equipamentos. Este procedimento deve ser seguido nos casos em que uma referência absoluta não esteja disponível ou viável de ser utilizada no teste. Quanto melhor o teste, mais simples é a natureza do julgamento que o ouvinte terá que fazer. Um exemplo típico de um teste com uma referência absoluta presente é o caso do teste de caixas acústicas: a comparação é feita entre a voz do locutor ao vivo na sala de teste e a reprodução da mesma voz gravada numa câmara anecóica;
d) execução de teste cegos, ou preferencialmente, duplo cegos. Um teste cego é aquele onde o ouvinte não sabe qual componente está sendo avaliado no momento da audição, enquanto que um teste duplo cego é aquele que nem o ouvinte nem a pessoa que está conduzindo o teste tem acesso à identidade do componente. Neste caso, um gabarito com as respostas deve ser gerado por um agente inexpressivo, como por exemplo um microcomputador.

CONDIÇÕES ACÚSTICAS DO AMBIENTE

Se o objetivo é eliminar ao máximo quaisquer fontes de interferência num teste auditivo, é bastante intuitivo se pensar em realizar os testes usando fones de ouvido. Desta maneira, as condições acústicas (e.g.: tempo de reverberação, ondas estacionárias) de uma sala de audição são evitadas. Esta facilidade não pode ser

implementada quando o componente a ser avaliado é uma caixa acústica, ou quando se deseja avaliar o componente simulando condições normais de operação.

Sendo assim, alguns requisitos a respeito das condições acústicas do ambiente onde acontece o teste devem ser levadas em consideração [3]:

a) tamanho e geometria: devem ser usadas salas com área superior a 20 m^2, com pé direito superior a 2,1 m. A geometria deve apresentar certa simetria longitudinal, porém devem ser evitadas paredes paralelas, pois propiciam o aparecimento de ondas estacionárias;
b) reflexões: o tempo de reverberação típico (RT60) compatível com salas de audição domésticas ou salas de controle em estúdios de gravação é de 0,45 ± 0,15 s. Reflexões que caracterizem ecos devem ser eliminadas;
c) ruído de fundo: o ruído de fundo na sala não deve exceder 35 dBA ou 50 dBC, nem tampouco deve conter tons periódicos dominantes (e.g.: ruídos de 60 Hz de reatores de lâmpadas fluorescentes);
d) posicionamento das caixas acústicas: a proximidade a regiões limites da sala (paredes, teto, piso) altera a carga acústica vista pela caixa, levando a variações na radiação de baixas freqüências, principalmente. Também as reflexões primárias de amplitude significativa no período de até 3 ms após o som direto atingir o ouvinte podem enfatizar as freqüências médias, trazendo uma coloração indevida ao programa executado. Em avaliações de imagem sonora de sistemas multicanais as caixas devem fazer um ângulo de, no mínimo, 40º com o ouvinte no vértice do ângulo;
e) posicionamento dos ouvintes: uma distância de, no mínimo, 1 metro de qualquer parede deve ser mantida para evitar colorações e/ou distorções de imagem sonora. A distância do ouvinte à caixa acústica deve ser coerente ao tipo da caixa (campo próximo, campo médio, etc.)

CARACTERÍSTICAS DOS PROGRAMAS EXECUTADOS

Deve ser dada preferência a programas armazenados em mídias digitais, dada a sua maior capacidade de manter a integridade dos dados durante a sua vida útil, além da possibilidade de serem feitas cópias precisas do material gravado.

Obviamente, as gravações utilizadas devem ter sua procedência garantida no que diz respeito às técnicas de microfonação e condicionamento dos sinais captados. Qualquer tipo de processamento (e.g.: equalização, compressão dinâmica) desabilita uma gravação por remeter o ouvinte a uma referência inexistente no mundo real.

Gravações de sons familiares, como a voz humana ou instrumentos musicais populares, devem ser preferidos. O conteúdo do espectro de freqüências do conjunto de gravações deve cobrir toda a faixa audível humana.

Alguns tipos de sinais como senóides e ruídos branco ou rosa podem ser utilizados em avaliações auditivas. Porém certo cuidado deve ser tomado na interpretação dos resultados, pois estes sinais podem revelar características do sistema em questão que não poderiam ser percebidas no caso da audição de música. Um exemplo típico é o fenômeno do sistema auditivo conhecido como mascaramento, onde um tom com amplitude relativamente alta encobre tons de freqüências próximas adjacentes [1]. Neste caso, distorções harmônicas de baixa ordem podem passar desapercebidas quando programas de conteúdo espectral complexo como música são ouvidos.

Determinados estilos musicais são estratégicos para identificação de parâmetros subjetivos. Sons percussivos são ideais para avaliação da capacidade de resposta a transientes; rock'n'roll serve bem na avaliação da resposta em freqüência do sistema; música sinfônica pode ser interessante devido às grandes variações de dinâmica normalmente exploradas.

O nível de pressão sonora (SPL) do programa executado deve ser compatível ao SPL das fontes sonoras originais. Uma orquestra sinfônica, por exemplo, pode chegar aos 95 dBA. Todos os componentes avaliados devem ser previamente calibrados de maneira a não apresentarem diferenças de intensidade maiores que ± 0,5 dB entre si.

CONDIÇÕES DOS OUVINTES

A experiência mostra que as habilidades dos ouvintes em perceber diferenças sônicas, em se concentrar devidamente no teste, ou de reagir de maneira consistente a diversas repetições de um mesmo estímulo sonoro, varia consideravelmente [4]. Os principais fatores que levam a esta variação de comportamento são as condições psico-fisiológicas do ouvinte, bem como a sua experiência prévia.

A correta seleção e treinamento de indivíduos para a realização de avaliações subjetivas em componentes de áudio proporciona maior confiabilidade ao teste, uma vez que fatores como a repetitibilidade e a significância dos resultados são mantidos sob maior controle. O grau de consistência das repostas é mensurado a partir do desvio padrão da taxa de acertos para um conjunto de avaliações.

Limiar da Audição

De acordo com a ISO 7092 [5], uma pessoa otologicamente normal é "aquela em condições normais de saúde, que no momento do teste não apresenta excesso de cera no canal auditivo, sem nenhuma patologia auditiva diagnosticada e que não apresenta histórico de exposição indevida a ruídos". Esta norma prega ainda que, caso seja considerado aceitável um desvio no limiar de audição de até 15 dB entre 250 e 8000 Hz, menos de 10% da população masculina otologicamente normal com idade de 20 anos será excluída.

Toole [6] realizou experimentos para analisar a influência causada pela magnitude do desvio do limiar de audição na variação da taxa de acertos de diversos ouvintes para experimentos repetidos. Foi detectado que para freqüências acima de 1 kHz a correlação entre o desvio padrão da taxa de acertos e o desvio do limiar de audição é pequena, enquanto que para freqüências abaixo de 1 kHz a correlação é alta. Em outras palavras, isto mostra que um ouvinte em más condições otológicas tem menor capacidade de manter coerência entre repetidas avaliações subjetivas. Entretanto, o grupo de ouvintes que Toole utilizou como amostra apresentava características como idade e histórico de exposição a ruídos que os impediam de serem classificados como otologicamente normais, segundo a ISO 7092.

Já no experimento conduzido por Bech [4], um grupo de 15 ouvintes otologicamente normais não apresentou correlação significante entre os dois parâmetros citados, o que leva a concluir que a escolha de ouvintes com condições adequadas de saúde otológica garante a habilidade destes em repetir de maneira consistente as taxas de acerto nas avaliações subjetivas.

Experiência Prévia

É bastante intuitivo e lógico acreditar que um grupo de ouvintes com comprovada experiência em tarefas de avaliação auditiva apresente taxas de acerto mais homogêneas que um grupo de ouvintes inexperientes.

Esta experiência inclui familiaridade com instrumentos musicais e presença em concertos, experiência com análise crítica de sons reproduzidos artificialmente ou executados ao vivo, e a aptidão geral em detectar diferenças audíveis em sons gravados.

Hansen [7] publicou resultados onde esta tendência se fez mostrar, e discute o esforço necessário em selecionar participantes com graus de experiência auditiva conhecidos.

A diferença no desvio padrão da taxa de acertos entre dois ou mais grupos com diferentes perfis de experiência revela que, dada uma magnitude de intervalo de confiança esperada, o número de ouvintes necessários para um determinado grupo deve ser ajustado.

O treinamento de ouvintes é um procedimento cujos resultados psicoacústicos são bem conhecidos. Após várias repetições de um mesmo estímulo, certas capacidades auditivas de um ouvinte podem ser alteradas, como por exemplo o limiar absoluto da audição. O objetivo de um programa de treinamento num grupo de ouvintes é garantir que os diferentes aspectos do desempenho auditivo destes alcance um nível assintótico, resultando na máxima eficiência de tempo, precisão e custos para uma avaliação.

MÉTODO A/B/X

O tipo de teste mais recomendado na literatura é o teste duplo cego A/B/X [8]. Este teste (Figura 1) tem como objetivo promover a comparação direta entre dois componentes de áudio (A e B). Em certos casos, o EET é ligado ao canal B e o canal A é curtocircuitado, servindo como referência. Assim, qualquer diferença é atribuída às distorções geradas pelo EET. O ouvinte tem à sua disposição um controle de comandos munido de uma chave seletora de três posições. As posições A e B conectam, logicamente, o componente A ou B à cadeia do áudio. A posição X dispara um circuito gerador de sinais aleatórios responsável por escolher entre conectar o canal A ou B. Ninguém tem acesso à escolha efetuada pelo dispositivo; um relatório, ou gabarito, do teste é gerado para averiguação posterior e determinação da taxa de acertos.

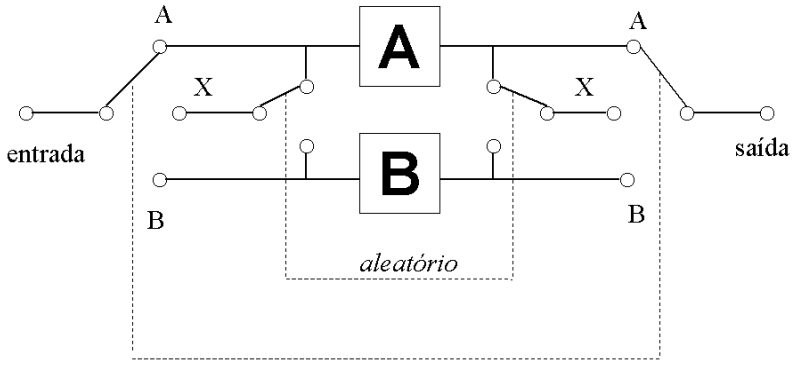


Figura 1: Diagrama esquemático do dispositivo para realização do teste A/B/X.

Alguns procedimentos operacionais são recomendados [8] para a maximização da resolução alcançada com o teste:

a) justaposição instantânea dos sinais comparados: vários aspectos qualitativos do som permanecem na memória do ouvinte por curtos intervalos de tempo;
b) questões de natureza comparativa: descrições qualitativas devem ser consideradas de acordo com a experiência do ouvinte;
c) tempo ilimitado: situações de pressão por prazos podem inibir a sensibilidade do ouvinte a certos parâmetros audíveis;
d) decisão forçada: o ouvinte sabe que sempre existe uma resposta certa, A ou B, o que o encoraja a explorar toda a sua capacidade auditiva em responder corretamente;
e) teste controlado pelo ouvinte: se desejado, o teste pode ser realizado individualmente, o que possibilita a adequação dos procedimentos (tempo de exposição ou chaveamento) às preferências do ouvinte;

EXEMPLO PRÁTICO

Para ilustrar a aplicação de algumas das metodologias discutidas, um teste subjetivo foi realizado com o objetivo de detectar diferenças entre as codificações digitais de sinais de áudio PCM e MP3. A PCM (Pulse Code Modulation) é uma codificação sem perdas utilizada nas tradicionais mídias digitais de áudio, principalmente nos CDs (Compact Disc). Nesta aplicação específica, a taxa de amostragem é de 44100 amostras por segundo e a resolução é de 16 bits, quantizados uniformemente. O MP3 é uma codificação com perdas que foi proposta para possibilitar uma compressão no tamanho do arquivo de áudio digital para a transmissão em meios com largura de banda limitada (INTERNET). Entretanto, os algoritmos que realizam a compressão do sinal de áudio no formato MP3 descartam certas informações que são pouco significativas para a percepção auditiva humana [9].

O objetivo da avaliação é, então, concluir se a codificação MP3 com suas perdas inerentes pode ser diferenciada auditivamente da codificação PCM 16 bits/44.1 kHz, empregada nos CDs. Assim, ambas codificações foram comparadas utilizando-se dois programas: o primeiro contendo uma gravação de voz masculina e o segundo contendo um trecho de uma peça musical executada por um conjunto de instrumentos de corda. Ambos os programas têm duração de quinze segundos. Duas taxas de compressão diferentes (128 e 92 kbps) para a codificação MP3 foram testadas, o que resultou num total de quatro testes comparativos. Para cada teste, os ouvintes eram expostos aos sinais originais, seguido de três audições do tipo X, onde foi questionado a procedência do sinal ouvido (A ou B).

O grupo de ouvintes foi composto por treze alunos e professores de pós graduação com idades entre 26 e 47 anos. Nenhuma avaliação física foi realizada nestes ouvintes para determinar suas condições otológicas.

Um formulário para a anotação das respostas foi distribuído contendo todas as instruções para a realização do teste (Anexo A). Por se tratar de um teste comparativo entre dois programas gravados, a influência dos equipamentos de reprodução utilizados pôde ser desconsiderada. A relação sinal ruído medida com ponderação A ficou acima de 50 dB SPL.

O resultado com as taxas de acerto para os quatro testes é apresentada na Tabela 1. Os resultados sugerem que a codificação MP3 com taxa de 128 kbps é estatisticamente improvável de ser diferenciada da codificação PCM, para ouvintes com condições otológicas adversas.

Nas Figuras 2 a 4 são apresentadas análises no domínio da freqüência (FFT) para um dos programas musicais executados e para um tom senoidal, o que permite constatar que a codificação MP3 além de introduzir distorções no sinal, ainda limita a sua banda de freqüências. No entanto, como foi apontado pelo teste subjetivo, em determinadas condições estes fenômenos são pouco percebidos pela audição humana.

É importante deixar claro que a execução deste teste não teve definitivamente a intenção de gerar resultados cientificamente comprovados sobre a qualidade sonora de qualquer uma das codificações. O objetivo aqui foi ilustrar alguns dos procedimentos defendidos ao longo deste texto, com foco principal na técnica ABX, além de confrontar resultados objetivos com subjetivos.

CONCLUSÃO

É possível realizar testes subjetivos para avaliação de componentes de áudio de maneira científica. As metodologias aqui apresentadas têm como objetivo conformar procedimentos com alto grau de controle para que resultados confiáveis e significantes sejam gerados.

A observância a fatores como condições do ambiente, condições dos ouvintes participantes e o método de avaliação são decisivos. Situações de teste com boas condições baseadas nestes quesitos são viáveis de ser implementadas à baixo custo.

REFERÊNCIAS

[1] D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966).

[2] S. P. Lipshitz and J. Vanderkooy, "The Great Debate: Subjective Evaluation," *J. Audio Eng. Soc*, vol. 29, pp. 482-491 (1981 July/Aug).

[3] AES 20-1996, "AES Recommended Practice for Professional Audio – Subjective Evaluation of Loudspeakers," Audio Engineering Society, 1996.

[4] Søren Bech, "Selection and Training of Subjects for Listening Tests of Sound-Reproducing Equipment," *J. Audio Eng. Soc*, vol. 40, pp. 590-610 (1992 July/Aug).

[5] ISO 7029-1984, "Threshold of Hearing Air Conduction as a Function of Sex and Age for Otologically Normal Persons," Internactional Organization for Standardization, Geneva, Switzerland (1984).

[6] F. Toole, "Subjective Measurements of Loudspeakers: Sound Quality and Listener Performance," *J. Audio Eng. Soc*, vol. 33, pp. 2-32 (1985 Jan/Feb).

[7] V. Hansen, "Establishing a Panel of Listeners at Bang and Olufsen: a Report," in S. Bech and O. J. Pedersen (Eds.), *Symp. On Perception of Reproduced Sound* (Gammel Avernæs, Denmark, 1987), Ingeniørhøjskolen, Århus Teknikum, Århus, Denmark.

[8] David Clark, "High-Resolution Subjective Testing Using a Double-Blind Comparator," *J. Audio Eng. Soc*, vol. 30, pp. 330-338 (1982 May).

[9] Pan, Davis; *Multimedia*, IEEE , Volume: 2 Issue: 2, Summer 1995, Page(s): 60 –74

Tabela 1: Resultado das Avaliações Subjetivas

	TESTE 1 Voz PCM x Voz MP3 128 kbps	TESTE 2 Voz PCM x Voz MP3 92 kbps	TESTE 3 Música PCM x Música MP3 128 kbps	TESTE 4 Música PCM x Música MP3 92 kbps
% taxa de acerto	48,7	64,1	53,8	71,8

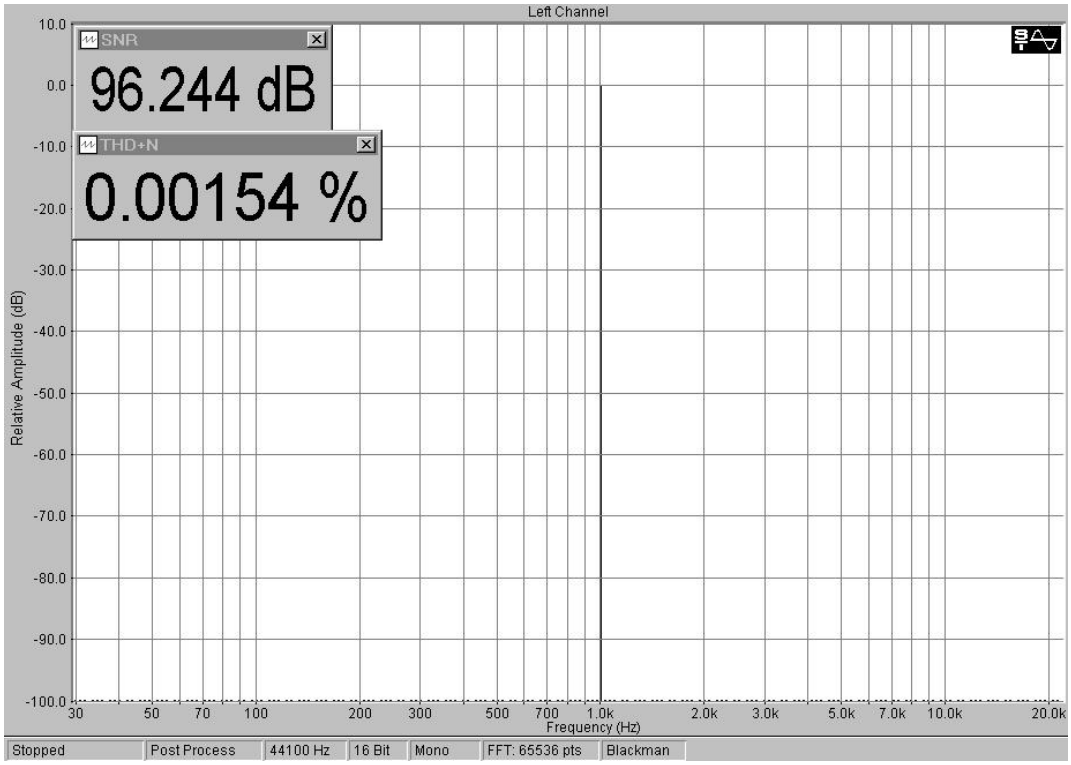


Figura 2: Análise FFT (64 k pontos) de um sinal senoidal de 1 kHz codificação PCM 16 bits/44.1 kHz.

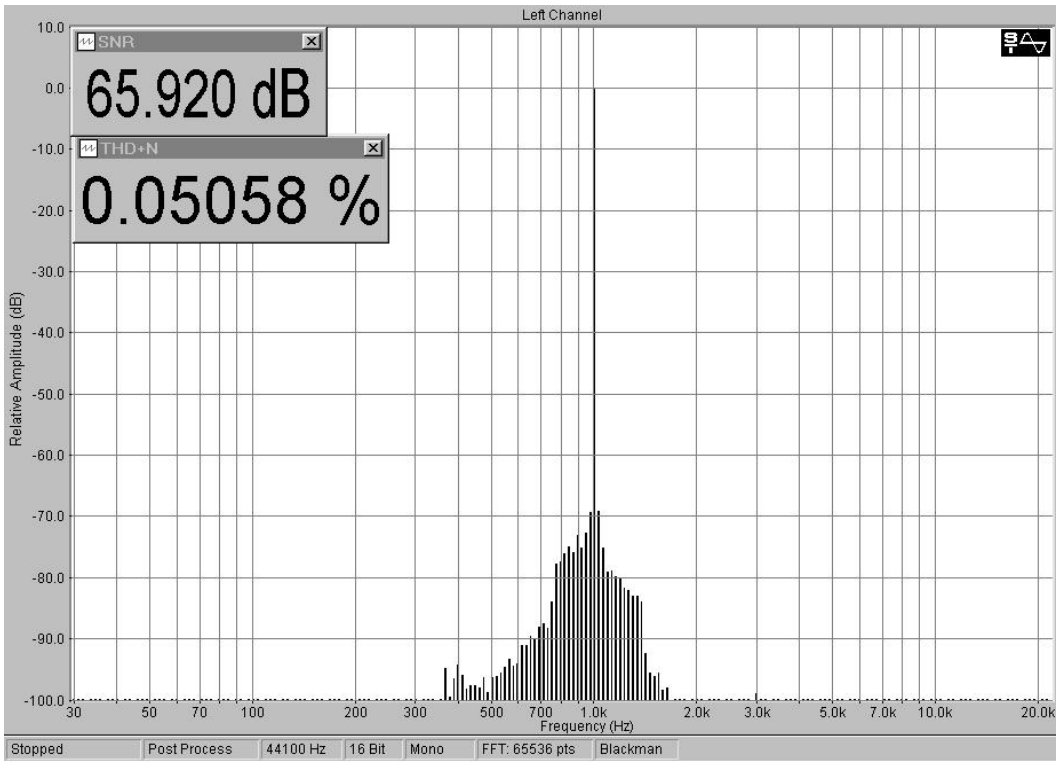


Figura 3: Análise FFT (64 k pontos) de um sinal senoidal de 1 kHz codificação MP3 a 128 kbps. Observar a redução na relação sinal ruído (SNR) e o aumento na taxa de distorção harmônica + ruído (THD+N).

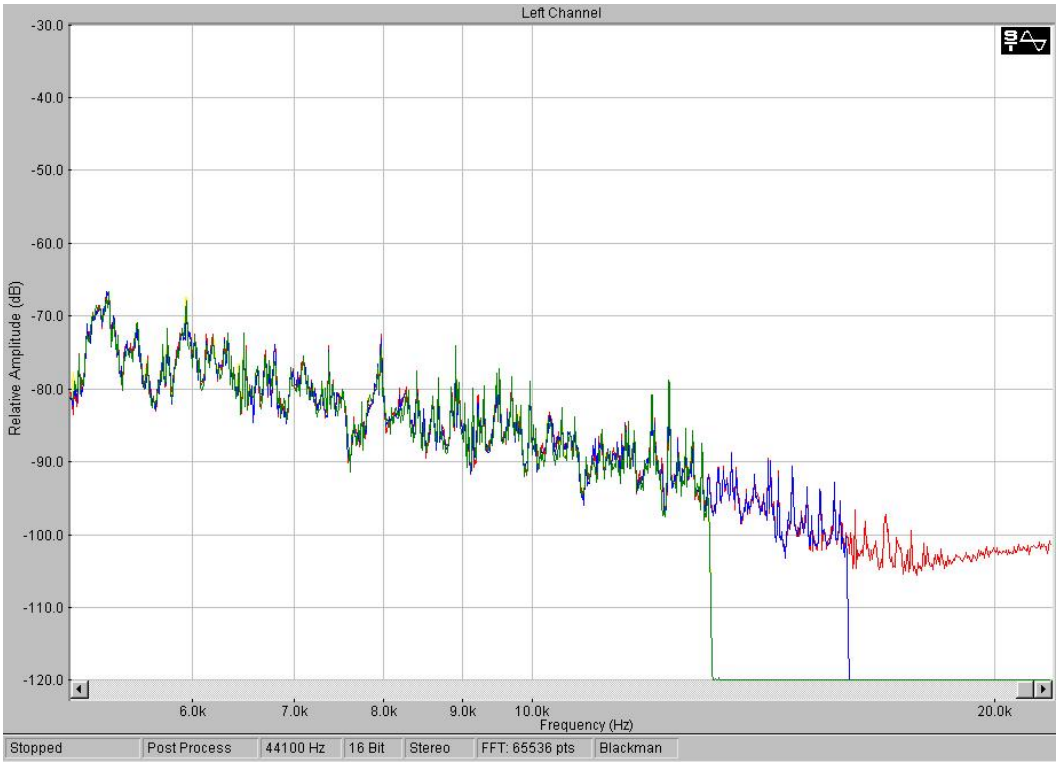


Figura 4: Detalhe da análise FFT (64 k pontos) de um trecho do programa musical utilizado (instrumentos de corda). O sinal com largura de banda mais limitada (em verde) refere-se à compressão MP3 à taxa de 92 kbps, enquanto que o sinal em azul refere-se à compressão MP3 com taxa de 128 kbps. O sinal em vermelho refere-se à codificação PCM.

ANEXO A

FORMULÁRIO PARA AVALIAÇÃO SUBJETIVA DE COMPONENTES DE ÁUDIO TESTE ABX

Localidade: CPDEE – UFMG – Belo Horizonte

Data: 20 de agosto de 2003

Horário: 12:00h

Nome: _____

Idade: _____

TESTE 1	TESTE 2	TESTE 3	TESTE 4
A – Voz PCM	A – Voz PCM	A – Música PCM	A – Música PCM
B – Voz MP3 128 kbps	B – Voz MP3 92 kbps	B – Música MP3 128 kbps	B – Música MP3 92kbps
X1 -	X1 -	X1 -	X1 -
X2 -	X2 -	X2 -	X2 -
X3 -	X3 -	X3 -	X3 -

Instruções:

1) Você está participando de uma avaliação subjetiva comparativa entre duas fontes sonoras.
2) O objetivo NÃO é avaliar qual das duas fontes soa melhor.
3) O objetivo é identificar DIFERENÇAS audíveis entre as duas fontes A e B.
4) Ouça com atenção a execução do programa A.
5) Ouça com atenção a execução do programa B.
6) Ouça as três execuções seguintes X1, X2 e X3 e marque a qual você julga se referir (A ou B).
7) Serão feitos ao todo quatro testes ABX.
8) A duração prevista é de oito minutos para todos os quatro testes.

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na VIII Convenção Nacional
Junho de 2004, 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 Adaptativo para Caracterização de Alto-Falantes

Rafael C. Haical, Fábio P. Itturriet, Márcio H. Costa, Delmar B. Carvalho
Universidade Católica de Pelotas
Pelotas, 96010-000, Rio Grande do Sul, Brasil
m.costa@ieee.org , delmar@phoenix.ucpel.tche.br

RESUMO

Sistemas para a avaliação e controle de qualidade de equipamentos de reprodução de som são de grande interesse para a indústria eletro-eletrônica, em especial na área de produção de alto-falantes. Neste trabalho é apresentado um sistema adaptativo, baseado no algoritmo *Normalized Least Mean Square*, para a avaliação de alto-falantes, que se caracteriza pela velocidade, confiabilidade e robustez à interferências externas.

1. INTRODUÇÃO

Nos tempos atuais de livre mercado, o controle de qualidade é um assunto de grande importância para as indústrias e em especial para as da área eletro-eletrônica. A partir disto, vários estudos vêm sendo realizados para desenvolver novas técnicas para a avaliação de produtos visando rapidez e confiabilidade.

De uma forma geral nem todos os dispositivos eletrônicos podem ser testados ao longo de uma linha de produção, portanto o procedimento de avaliação por amostragem é utilizado como forma de caracterização de cada lote produzido. Um exemplo desta prática é encontrado na produção de alto-falantes, importante setor tecnológico do país. O processo de controle de qualidade por amostragem em linhas de fabricação de alto-falantes decorre da necessidade de retirada do produto da linha de montagem para sua avaliação, devido ao elevado nível de ruído acústico encontrado no local. Como conseqüência, a avaliação em larga escala é prejudicada.

Diversas técnicas podem ser utilizadas para a identificação e modelagem de sistemas, cada uma apresentando diferentes propriedades [1-5]. Em [6] foi apresentado um sistema automático para teste de alto-falantes. Embora apresente resultados interessantes, este sistema requer o isolamento acústico do conjunto alto-falante – microfone de recepção de forma a diminuir a influência do ruído sonoro externo sobre os parâmetros obtidos. Em decorrência disto, aumenta o tempo necessário para a caracterização do produto.

Uma interessante alternativa para possibilitar a avaliação rápida e confiável *in loco*, ou seja, na linha de produção, é a utilização de filtros adaptativos, em decorrência de sua intrínseca robustez à interferências externas.

A família de algoritmos adaptativos baseada no método do gradiente estocástico é uma interessante possibilidade para a identificação de sistemas. Dentre seu inúmeros membros, o algoritmo *Normalized Least Mean Square* (NLMS) é amplamente utilizado em sistemas reais, possuindo como características um baixo custo computacional, simplicidade de implementação e robustez à variações da potência do sinal de excitação [7].

Este trabalho apresenta um sistema para a caracterização de sistemas de reprodução de som, em especial de alto-falantes, baseado em técnicas de filtragem adaptativa. Como resultado são obtidas a resposta ao impulso do sistema sob análise e sua resposta em freqüência. Em decorrência da elevada robustez à interferências não-correlacionadas com o

sinal de excitação (produzido artificialmente) o presente sistema é especialmente indicado para utilização na própria linha de produção, permitindo assim, a avaliação rápida e confiável do maior número possível de dispositivos.

2. ALGORITMO NLMS

Os filtros adaptativos são ferramentas extremamente úteis na área de processamento de sinais, principalmente em aplicações de tempo real onde pouca ou quase nenhuma informação dos sinais envolvidos é conhecida *a priori*. Suas áreas de aplicação abrangem desde o cancelamento de eco e processamento de sinais de radar ao controle de sistemas eletrônicos e/ou de potência [7].

O algoritmo NLMS ou LMS Normalizado é uma variação do bem conhecido algoritmo LMS. Ele aplica em sua fórmula uma normalização do erro, provocando a utilização de passos de convergência variáveis, de acordo com a potência do sinal de entrada $x(n)$.

O diagrama em blocos do algoritmo NLMS é apresentado na figura 1, onde $\mathbf{c^o} = \begin{bmatrix} c_0^o & c_1^o & \dots & c_{N-1}^o \end{bmatrix}^T$ é o sistema a ser identificado, $\mathbf{c}(n) = \begin{bmatrix} c_0(n) & c_1(n) & \dots & c_{N-1}(n) \end{bmatrix}^T$ o vetor de coeficientes adaptativos, $x(n)$ é o sinal de excitação, $d(n)$ o sinal desejado, $y(n)$ a saída do filtro adaptativo, $z(n)$ o ruído externo e $e(n)$ o sinal de erro definido por

$$e(n) = d(n) - y(n) + z(n) \tag{01}$$

a saída do filtro é determinada por $y(n) = \mathbf{c}^T(n)\mathbf{x}(n)$, onde $\mathbf{x}(n) = \begin{bmatrix} x(n) & x(n-1) & \dots & x(n-N+1) \end{bmatrix}^T$ é o vetor de amostras do sinal de excitação. A equação de atualização do algoritmo NLMS é dada por [7].

$$\mathbf{c}(n+1) = \mathbf{c}(n) + \beta\, e(n)\mathbf{x}(n) \Big/ \left[\mathbf{x}^T(n)\mathbf{x}(n) + \delta \right] \tag{02}$$

onde δ é uma constante pequena e positiva que evita *overflow* ou divisões por zero quando o produto $\mathbf{x}^T(n)\mathbf{x}(n)$ é muito pequeno ou nulo. O objetivo deste algoritmo é minimizar a potência do sinal de erro de tal forma que após um determinado tempo de adaptação obtenha-se $\mathbf{c}(n) \cong \mathbf{c^o}$, ou seja, os coeficientes do filtro adaptativo são uma estimativa do sistema sob análise, supondo-se este linear e invariante no tempo e que o número de coeficientes adaptativos seja suficiente.

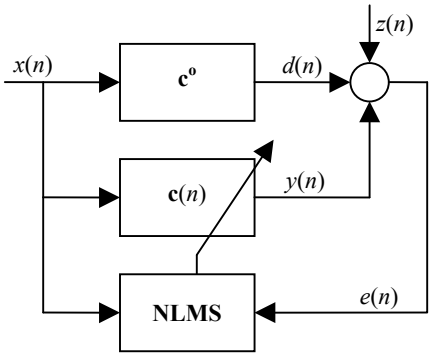


Fig. 1. Diagrama em blocos do algoritmo NLMS.

3. O SISTEMA DESENVOLVIDO

O sistema desenvolvido é baseado no processador digital de sinais ADSP-21061 [8] que possui uma arquitetura de 32 bits, ponto flutuante e ciclo de instrução de 20ns, operando em 40MHz. Com uma memória *cache* de instrução interna, este processador pode executar várias instruções em um único ciclo, acarretando uma elevada velocidade de processamento.

Para a implementação do sistema foi utilizada a placa comercial de aquisição de sinais Sharc Ez-Kit Lite [9], que além do processador ADSP-21061 possui um Codec com um par estéreo de conversão analógico-digital e porta serial de 16 bits para comunicação com um computador pessoal. A programação é realizada em linguagem C.

A figura 2 apresenta um diagrama em blocos simplificado do sistema implementado. Através de uma rotina de números pseudo-aleatórios, o sistema de processamento (processador ADSP-21061) gera um sinal de excitação elétrico de média nula, variância constante e característica Gaussiana não-correlacionada. Este sinal primeiramente é amplificado por um pré-amplificador baseado no circuito integrado TDA1013 (*Philips Semiconductors*) e logo após por um módulo de potência comercial marca *Cowatts*. Finalmente, o sinal é entregue ao alto-falante que se deseja caracterizar. O sinal acústico produzido pelo alto-falante é adquirido através de um microfone de eletreto e é amplificado de forma a compatibilizar o sinal com o canal de aquisição da placa EZ-Kit Lite.

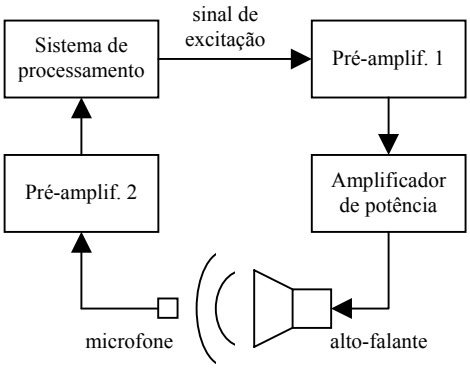


Fig. 2. Diagrama em blocos do sistema de identificação.

A figura 3 apresenta de forma detalhada o sistema implementado, sob o ponto de vista de um problema de identificação adaptativa. Verifica-se claramente que o sistema estimado, ao final do processo de adaptação, é na realidade o conjunto de subsistemas composto pelos amplificadores necessários ao condicionamento do sinal, alto-falante e microfone de aquisição. Desta forma, o projeto do sistema requer a correta escolha e dimensionamento dos amplificadores, microfone e freqüência de amostragem, de forma que sua influência em termos de banda passante seja desprezível frente às características do alto-falante em questão.

Na figura 4 é apresentado o protótipo implementado. À direita pode ser visualizado o computador responsável pela apresentação dos resultados e à esquerda o sistema desenvolvido. Neste, no canto superior direito podem ser

visualizados o alto-falante que se deseja caracterizar e o microfone de aquisição. Ao centro e acima o módulo de potência comercial e abaixo os pré-amplificadores e filtros necessários à compatibilização dos níveis de sinal entre a placa Ez-kit Lite e os transdutores envolvidos. À direita e acima encontra-se a fonte de alimentação e abaixo a placa Ez-Kit Lite.

O algoritmo NLMS foi implementado em linguagem C no processador ADSP-21061 que transfere os resultados do processo de adaptação ao computador pessoal.

Ao final do procedimento de identificação, os coeficientes do filtro adaptativo, que representam a resposta ao impulso do alto-falante, são utilizados para a avaliar a sua resposta em freqüência. No protótipo desenvolvido, foi utilizado o programa MatLab para esse propósito.

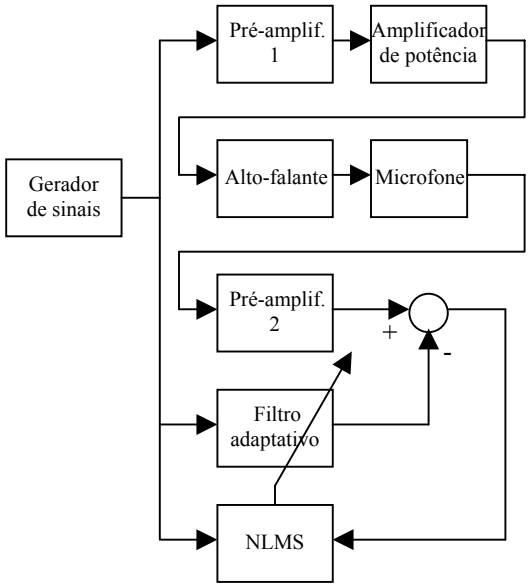


Fig. 3. Diagrama em blocos detalhado do sistema.

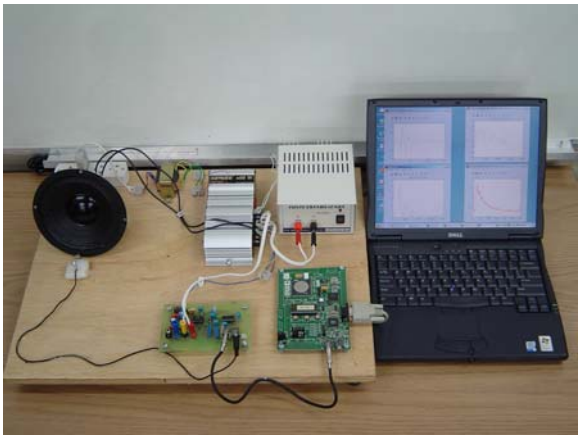


Fig. 4. Foto do sistema de identificação.

4. RESULTADOS EXPERIMENTAIS

De forma a demonstrar a capacidade do sistema implementado são apresentados alguns resultados representativos do desempenho do sistema, realizados em diferentes circunstâncias.

Os parâmetros utilizados foram: 150 coeficientes, freqüência de amostragem de 16 kHz, passo de convergência 0,15, sinal de excitação branco Gaussiano. O alto-falante utilizado foi do tipo *woofer* com potência de 140 W.

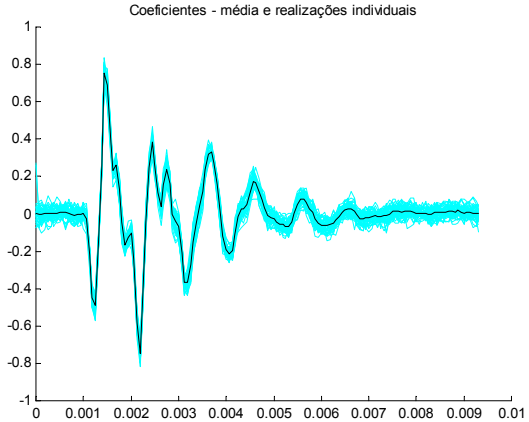


Fig. 5. Resposta ao impulso do sistema analisado (150 coeficientes). Em ciano as realizações individuais e em preto a média de 50 realizações do processo de identificação.

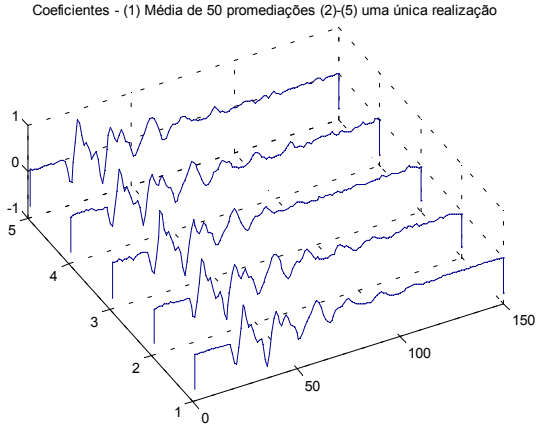


Fig. 6. Resposta ao impulso do sistema analisado. (1) média de 50 realizações, (2-5): realizações individuais.

Na figura 5 são apresentados os coeficientes do filtro adaptativo após 0,2 segundos de processamento, resultando em 3200 iterações do algoritmo NLMS. No eixo vertical encontra-se a amplitude e no horizontal o tempo de reverberação em segundos (150 coeficientes × 1/16000 Hz = 0,0094 segundos). Em preto pode-se visualizar a média de 50 realizações e em ciano cada uma das mesmas individualmente. Pode-se, portanto, concluir que a variabilidade dos diferentes procedimentos de identificação foi bastante pequena e portanto, uma única realização do procedimento de adaptação é suficientes para caracterizar a resposta impulsiva do alto-falante nas condições ambientais

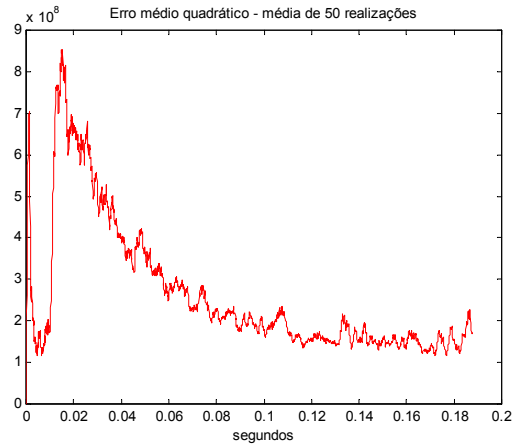


Fig. 7. Erro médio quadrático em função do tempo de adaptação.

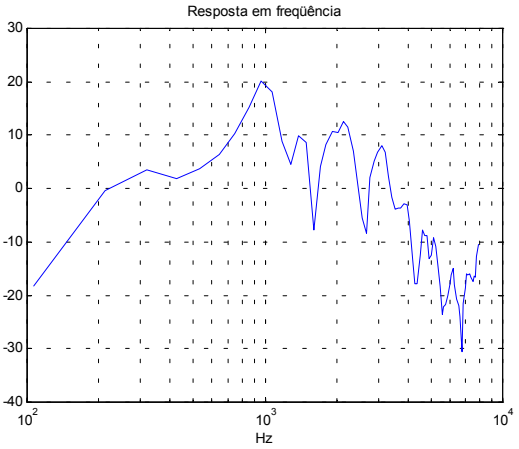


Fig. 8. Resposta em freqüência do sistema analisado.

A figura 7 apresenta o erro médio quadrático (avaliado como a média do sinal de erro $e(n)$ ao quadrado) para as 50 realizações. Este sinal foi processado por um filtro média móvel de norma unitária e 20 coeficientes para facilitar a visualização. Como resultado pode-se notar a consistente convergência do processo de adaptação através da diminuição da potência do erro em função do número de iterações.

A partir da resposta ao impulso obtida pode-se, através da utilização da transformada rápida de Fourier, obter a resposta em freqüência do alto-falante, apresentada na figura 8. Verifica-se o comportamento passa-faixa característico do alto-falante do tipo *woofer* analisado. O elevado decaimento em altas freqüências demonstra a adequada escolha da freqüência de amostragem para este tipo de alto-falante de forma a evitar a ocorrência de *aliasing*.

5. DISCUSSÃO E CONCLUSÕES

Nesse trabalho foram apresentadas as características principais da implementação de um sistema adaptativo real para aplicações de identificação. Através da determinação da resposta ao impulso do sistema analisado é obtida a sua resposta em freqüência.

O sistema caracteriza-se por um baixo custo e herda as propriedades de velocidade, confiabilidade e robustez à interferências externas pertinentes ao algoritmo adaptativo *Normalized Least Mean Square*. Embora possa ser utilizado em uma ampla gama de aplicações, encontra especial uso em sistemas de reprodução de som, mais especificamente na caracterização de alto-falantes.

Em trabalhos futuros pretende-se quantificar os erros nas estimativas obtidas em decorrência do ruído acústico de fundo e do *hardware* associado.

6. REFERÊNCIAS BIBLIOGRÁFICAS

[1] Ljung, L., System Identification – Theory for the User, segunda edição, Prentice-Hall, 1999.

[2] Aguirre, L.A., Introdução à Identificação de Sistemas – Técnicas Lineares e Não-Lineares Aplicadas a Sistemas Reais, Editora UFMG, Belo Horizonte, 2000.

[3] Heinle, F., Rabenstein, R. e Stenger, A., "A Measurement Method for the Linear and Nonlinear Properties of Electro-Acoustic Transmission Systems", Signal Processing, volume 64, número 1, páginas 49-60, 1998.

[4] Clark, D., "Precision Measurement of Loudspeaker Parameters", J. Audio Eng. Soc., volume 45, número 3, páginas 129-141, março, 1997.

[5] Scott, J. e Kelly, J., "New Method of Characterizing Driver Linearity", J. Audio Eng. Soc., volume 44, número 4, páginas 258-265, abril, 1996.

[6] Pereira, L.F.A., Fernandes, D., Manzoni, A., Lima, J.C.M. e Reis, F.S., "Sistema Automático para Teste de Auto-Falantes", XIII Congresso Chileno de Ingenieria Electrica, páginas 667-672, Santiago, Chile, novembro, 1999.

[7] Haykin, S. Adaptive Filter Theory, quarta edição, Prentice-Hall, 2002.

[8] ADSP 2106x Sharc User's Manual, second edition, Analog Devices, 1996.

[9] Ez-Kit Lite, ADSP 2106x Sharc, Reference Manual, Analog Devices, 1997.

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na VIII Convenção Nacional
7-9 de Junho de 2004, São Paulo, SP, Brasil

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 de viabilidade da aplicação de filme fino metálico em cones e domos de alto-falantes

Hugo D' Alessandro Jucosky[a], Roberto Katsuhiro Yamamoto[a,b], Walter Ullmann[c]
[a]FATEC-SP, [b]LSI-USP, [c]dBW LTDA.
São Paulo, SP, Brasil
hugo_j@pop.com.br, katsu@fatecsp.br, walter_ullmann@hotmail.com

RESUMO

Realizou-se um experimento com um tweeter com domo de tecido com o objetivo de alterar sua aparência sem que suas propriedades sonoras fossem modificadas. Para isto, foi aplicado um filme fino de alumínio depositado por método de evaporação, para tornar o domo de tecido com aparência de metal. Demonstrou-se a viabilidade de melhorar a aparência visual de alto-falantes por meio da deposição de filmes finos sem comprometimento da qualidade de reprodução.

INTRODUÇÃO

Muitas vezes não é possível aliar qualidade de reprodução com estética. Apesar da estética ser, teoricamente, pouco importante, muitas vezes é decisiva na escolha de equipamentos, principalmente no mercado de *home-theater*. Prova disto são os inúmeros cones de alto-falantes confeccionados em polipropileno, que fornecem a possibilidade de serem transparentes ou até mesmo multicoloridos. Porém, quando os cones são confeccionados com materiais como papel, cerâmica ou tecido (no caso de tweeters), não se torna viável sua alteração de cor. Isto porque seria necessário adicionar, na superfície do domo ou cone, algum material que forneça um visual mais agradável. Ao realizar este procedimento, acaba-se comprometendo a assinatura sônica do dispositivo. Isto ocorre devido a três fatores principais:

- *adição de massa no conjunto móvel;*
- *alteração da resistência mecânica do cone ou domo;*
- *diminuição ou fechamento dos poros do material;*

Estes três itens, considerados isoladamente ou em conjunto, acabam alterando a assinatura sônica do material utilizado para a construção do alto-falante. Existe uma maneira de se realizar a alteração estética sem envolver os itens anteriores: utilizar material com massa desprezível, como um filme fino.

Existem diversos métodos para tornar a quantidade de material depositado desprezível, ou seja, formação de filme fino. O método utilizado é conhecido como evaporação e é aplicado em diversos mercados. Trata-se de uma técnica simples onde o substrato é posicionado em um suporte dentro de uma câmara de vácuo. O substrato fica a uma certa distância de um filamento de tungstênio, onde ficam apoiados filetes de metal, normalmente o alumínio. Ao passar uma corrente através do filamento, provoca-se o aquecimento do mesmo e, conseqüentemente, o aquecimento do alumínio provocando sua evaporação. O vapor de alumínio tende a se espalhar pela câmara de vácuo de forma esférica atingindo o substrato que se deseja metalizar. O vapor de alumínio é depositado na superfície do material, formando um filme fino com espessura variável de acordo com os parâmetros utilizados.

PROCEDIMENTO EXPERIMENTAL

Foi realizado um experimento com um tweeter com domo de tecido, para verificar a viabilidade da técnica de deposição de filmes finos, pois o tweeter é bastante sensível à qualquer alteração em sua construção. Normalmente, sua fabricação é bem mais delicada e precisa, justamente devido à sua sensibilidade física e tamanho reduzido. Tipicamente, sua freqüência de trabalho está entre 2 e 20kHz, ou seja, de todos os tipos de alto-falantes, é o que possui maior abrangência de

trabalho. A utilização deste tweeter se deve às suas características de construção e funcionamento, ou seja, utiliza uma bobina móvel e possui o domo totalmente exposto, característica muito importante para a deposição de filmes finos.

O tweeter utilizado possui um domo de tecido impregnado com resina, diâmetro de 25,4 mm, potência máxima de 30W e impedância de 8Ω. A Fig. 1 mostra a foto do tweeter utilizado.

Fig. 1 – Foto do tweeter utilizado com domo de tecido.

Evaporação de alumínio.

O tweeter foi posicionado na evaporadora sobre um suporte a uma distância aproximada de 8 cm do filamento de tungstênio, sobre o qual filetes de alumínio foram carregadas. Não houve preocupação quanto à espessura (estimada em 500 nm) de alumínio a ser depositada sobre o domo, pois a uniformidade seria prejudicada devido a sua esfericidade. A única preocupação foi de obter deposição sobre todo o domo sem que houvesse danos provocados pelo calor. A pressão obtida foi de aproximadamente $6x10^{-4}$ Pa. Não foram utilizadas máscaras durante o processo. Logo, todo o domo foi recoberto com alumínio, inclusive as bordas plásticas, como pode ser visto na Fig. 2.

Fig. 2 – Foto do tweeter mostrando a aparência prateada após a metalização.

RESULTADOS E DISCUSSÕES

A curva de resposta em freqüência do tweeter foi realizada da seguinte maneira: aplicou-se um sinal senoidal no alto-falante com tensão eficaz de 2V (correspondente a 0,5W RMS). Com um microfone[1] a 3 cm do domo, captou-se o som reproduzido pelo tweeter para cada freqüência aplicada.

[1] Marca: LinearX, modelo M31, Diâmetro: 1/4 de polegada.

Devido às limitações de construção do alto-falante, haverá perdas de intensidade para determinadas freqüências. Idealmente, a curva de intensidade (dB SPL) versus freqüência (Hz) deve ser plana, ou seja, as respostas do falante devem ser de igual intensidade, independente da freqüência aplicada, garantindo assim uma boa fidelidade de reprodução do sinal aplicado. Porém, muitas vezes não é possível obter uma curva totalmente plana. Isto está relacionado à qualidade dos materiais utilizados na construção do alto falante, simetrias (de construção e campo magnético fixo), posicionamento correto da bobina e uniformidade do peso do conjunto bobina móvel e domo.

Para verificar a influência do filme fino de alumínio depositado na resposta do tweeter, foi realizada a medição antes e depois do processo, obtendo-se as curvas mostradas na Fig. 3.

Verifica-se que as curvas não são planas e restringem a operação de trabalho do tweeter de 2 a 8kHz (região mais plana do gráfico), embora esta região possa ser estendida até 20kHz. Pode-se verificar também que a deposição de alumínio não influenciou de forma significativa na resposta do tweeter, porém houve uma ligeira queda em seu desempenho, provavelmente devido a alguma deformação provocada durante o processo de deposição.

Análise da curva de impedância *versus* freqüência.

Em um alto-falante, a bobina vibra na mesma freqüência da corrente alternada que é aplicada em seus terminais. Isso indica que haverá diferentes valores de resistência elétrica da bobina de acordo com a variação da freqüência[2]. Logo, o termo "resistência" não pode ser aplicado neste caso, sendo substituído pelo termo "impedância" que considera o impedimento da passagem da corrente elétrica alternada [1]. No alto-falante não existe apenas a impedância elétrica da bobina, mas também outras impedâncias que variam com a freqüência:

- *Impedância mecânica:* envolve a dificuldade de movimento do cone devido à força restauradora (compliância) da aranha e da suspensão (borda);
- *Impedância acústica:* dificuldade de movimento devido ao ar.

A impedância total do alto-falante considera todos os itens mencionados acima. Por meio da curva "Impedância versus Freqüência" mostrada na Fig. 4, pode-se verificar se a bobina sofreu alteração após a deposição de alumínio. Esta curva indica que houve alterações nas propriedades elétricas e físicas do tweeter, justificando a diferença entre as curvas da Fig. 3.

Assim, através das curvas da Fig. 4, conclui-se que houve perda de ganho do tweeter devido às alterações mecânicas que ocorreram no momento da evaporação.

Teste de aderência.

Realizou-se um teste de aderência do filme fino de alumínio depositado sobre o domo, colocando-se o tweeter em funcionamento durante um certo período. Com a vibração do domo, o filme fino metálico também vibra, tendendo a descolar de sua superfície.

[2] Existem duas componentes da bobina que variam com a freqüência: 1 – componente resistiva que representa as perdas originadas pelas correntes induzidas na ferragem; e 2 – componente indutiva que representa a reatância indutiva da bobina (parte imaginária de sua impedância) [2].

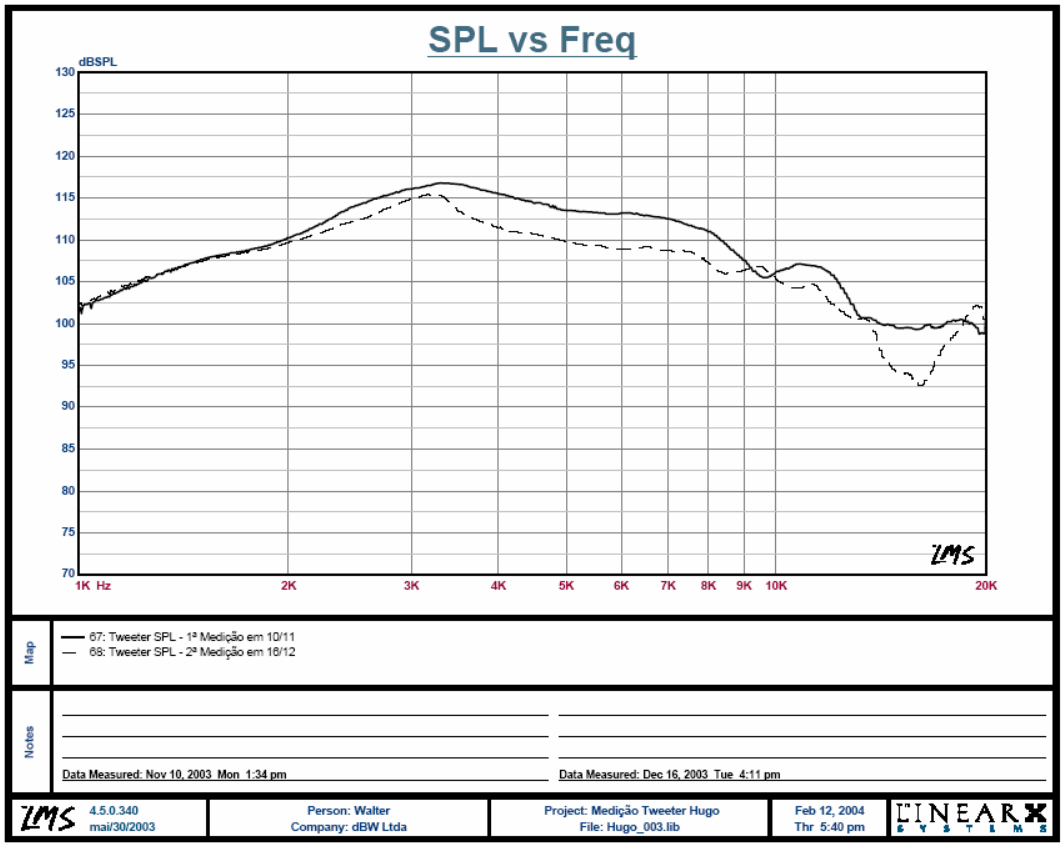


Fig. 3 –Curvas de resposta em freqüência do tweeter antes e depois da deposição do filme fino de alumínio.

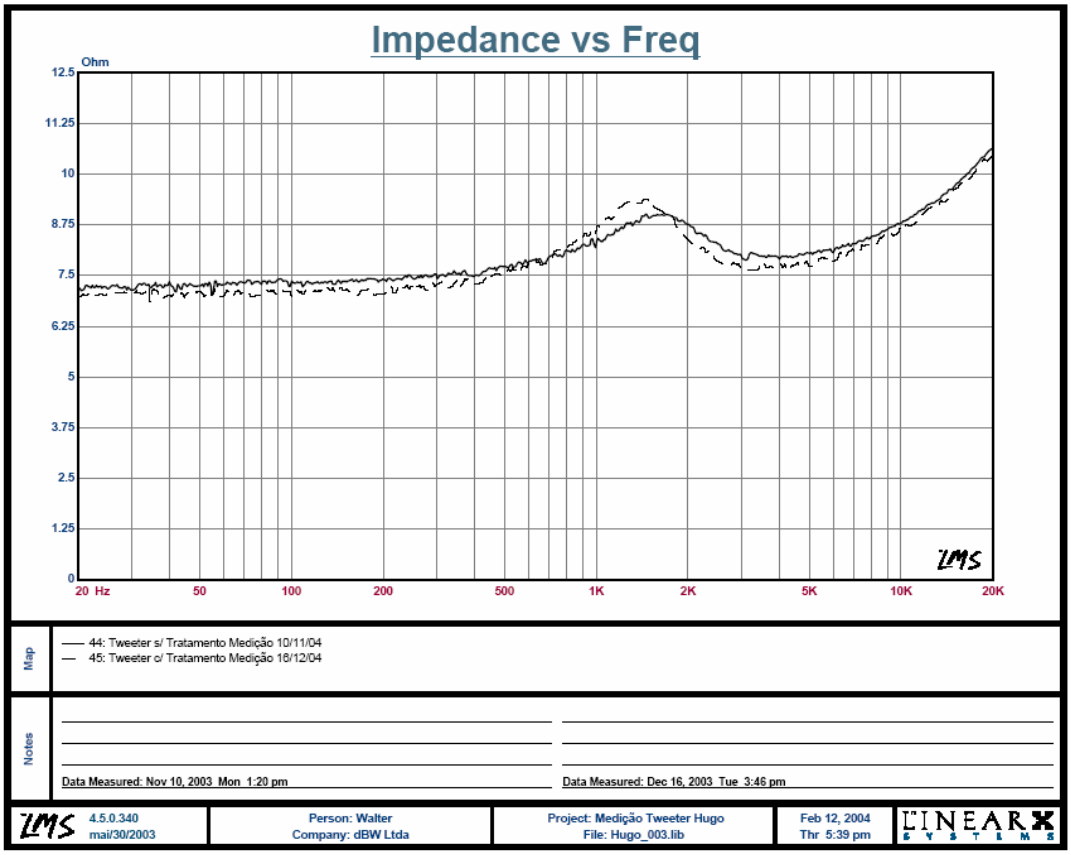


Fig. 4 - Curvas de impedância *versus* freqüência do tweeter antes e depois da deposição do filme fino de alumínio.

Caso o filme fino descolasse durante o funcionamento do tweeter, concluir-se-ia que a deposição de filmes finos sobre tecido ou papel não fornece aderência suficiente para aplicações em domos ou cones de alto-falantes.

Para realizar este teste, criou-se um *ruído rosa* com freqüência limitada entre 2 e 20kHz (região de trabalho do tweeter) juntamente com um sinal senoidal também variando de 2 a 20kHz. Desta maneira, evitou-se sons com baixas freqüências, com grandes amplitudes, que provocariam a queima do tweeter e dispensou o uso de divisor de freqüência. O teste foi realizado durante 200 horas ininterruptas e visualmente não se observou qualquer alteração na superfície do domo. Pode-se concluir que os filmes finos depositados sobre domos de tweeters resistem às vibrações mecânicas sem que descolem da superfície.

CONCLUSÃO

Foi possível verificar a viabilidade da deposição de filmes finos de alumínio para aplicação em alto-falantes, com o objetivo de proteção e melhoria estética.

Aplicações de filmes finos de qualquer espécie, desde que apresentem massa desprezível, sobre domos de tweeters não influenciam de maneira significativa o som reproduzido.

O teste de aderência mostrou que o filme fino de alumínio depositado por evaporação possui aderência necessária para trabalhar mesmo em altas freqüências.

Trata-se de um excelente resultado principalmente para mercados de alta fidelidade de reprodução sonora.

Um estudo mais detalhado, envolvendo-se variações de espessura, rugosidade e materiais utilizados, pode fornecer dados mais precisos. Variando-se o tipo de material, pode-se obter cores prateadas (ex.: alumínio) ou douradas (ex.: nitreto de titânio).

AGRADECIMENTOS

Ao Prof. Dr. Luis da Silva Zambom da FATEC-SP / LSI-USP pelo auxílio prestado no processo de evaporação.

CRÉDITOS DAS FOTOS

Carlos E. B. Nascimento

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Selenium Alto-falantes; “Para que serve o Alto-Falante? O que é som, intensidade, tom, timbre?”, www.selenium.com.br.

[2] V. Dickason; H. Sette, – “Caixas acústicas e alto-falantes” – Quinta edição – Ed. H. Sheldon – Capítulo 0 – págs.: 1-8.

[3] J. C. Fernandes: “Acústica e Ruídos”; 2002; Unesp – Bauru.

[4] S. Wolf and R.N. Tauber, “Silicon Fabrication for the VLSI Era”, Volume 1 - Silicon Processing, Lattice Press, 1986 – p. 161-195.

[5] “Antologia hi-fi estéreo” - Edições Monitor - 1ª. edição – 1960 – págs.: 99-108.

[6] Nova Eletrônica - Editora Técnica Eletrônica LTDA. Edições: 2, 3, 4, 5, 6, 7, 10, 11 (março de 1977 a janeiro de 1978). Matéria: "Curso de áudio".

[7] G. C. Johnston; “Measurement of the lowest resonance frequency of loudspeaker cones”; AES19-1992 (1998); (ALMA TM-100).

[8] V. E. P. Lazzarini: “Elementos de Acústica”; Universidade Estadual de Londrina; julho de 1998.

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na VIII Convenção Nacional
7-9 de Junho de 2004, São Paulo, SP, Brasil

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.

Impermeabilização de cones de papel de alto-falantes com plasma de CF_4/H_2

Hugo D' Alessandro Jucosky[a], Roberto Katsuhiro Yamamoto[a,b]
[a]FATEC-SP, [b]LSI-USP
São Paulo, SP, Brasil
hugo_j@pop.com.br, katsu@fatecsp.br

RESUMO

Visando obter proteção hidrofóbica e, conseqüentemente, um aumento da vida útil de cones de papel de alto-falantes, realizou-se experimentos com polimerização com plasma de CF_4/H_2. A maior vantagem do uso desta técnica é a possibilidade de fornecer a proteção impermeabilizante sem que o peso final do cone seja alterado de maneira significativa, não comprometendo o desempenho do alto-falante.

INTRODUÇÃO

A adsorção de produtos na superfície do cone de um alto-falante é um dos fatores que podem reduzir sua vida útil, pois o acúmulo de umidade e poeira em sua superfície pode provocar a degradação do material.

O objetivo deste trabalho consiste em depositar filmes finos poliméricos em cones de alto-falantes de papel, por meio do plasma de CF_4/H_2, como um meio de obter proteção hidrofóbica, ou seja, impermeabilização dos cones de papel, de maneira a aumentar a sua durabilidade ou a vida útil. Estes primeiros testes apresentam apenas este objetivo, pois a técnica de deposição por plasma em papel é pouco conhecida na literatura. Estes testes fornecerão dados que podem viabilizar, no futuro, estudos tímbricos (variação na resposta de freqüência por exemplo). A técnica permite a deposição de diversos materiais sobre a superfície (de polímeros a metais), permitindo desta maneira, expandir ainda mais as pesquisas futuras. Os cones utilizados no experimento são de alto-falantes de bobina móvel (eletrodinâmicos), que são os mais utilizados atualmente. O objetivo é obter uma impermeabilização total do cone (com boa aderência do material) e com menor quantidade de material possível depositado. Isto é importante para manter o peso praticamente inalterado sem prejudicar o desempenho do alto-falante.

No processo de deposição por plasma, o filme fino polimérico será depositado em ambos os lados do cone. Por ser um processo com rigoroso controle, é possível obter deposição mínima na superfície do cone. Portanto, pode-se considerar que o peso médio final será inferior ao processo manual (deposição líquida por exemplo) ou industrial (deposição por jato de *spray* ou colagem de material) de impermeabilização, melhorando sua performance ao mesmo tempo em que garante uma proteção igual ou superior aos métodos citados.

Há dois fatores que devem ser levados em consideração: a aderência do filme fino depositado na superfície e a uniformidade de deposição.

A aderência é importante uma vez que, os alto-falantes podem vibrar milhares de vezes por segundo, podendo soltar o filme fino depositado.

Por se tratar de um substrato não plano, a uniformidade deve ser tal que haja material depositado em toda a superfície do cone para garantir a proteção total.

Uma das técnicas de deposição de filmes finos é o PECVD (Plasma Enhanced Chemical Vapor Deposition) que é uma deposição química em fase vapor assistida por plasma. Sua maior vantagem é manter o substrato em baixa temperatura [1]. Neste trabalho, o processo foi realizado em um equipamento de corrosão do tipo RIE (Reactive Ion Etching).

Funcionamento da técnica PECVD.

O plasma é gerado dentro de um ambiente de baixa pressão (uma câmara de vácuo), através da aplicação de uma tensão de RF de 13,56 MHz de freqüência. Elétrons livres ganham energia suficiente (devido ao campo elétrico) e colidem com as moléculas do gás inserido, causando sua dissociação e ionização, formando o plasma. Os radicais são adsorvidos na superfície da amostra, que está sujeita ao bombardeamento de íons, promovendo reorganização e reações com outras espécies adsorvidas, estabelecendo novas ligações químicas, formando um filme fino. Os radicais formados no plasma são altamente reativos e os filmes obtidos, em geral, não são estequiométricos pois os mecanismos envolvidos na deposição são bastante complexos [2].

Neste trabalho, foram utilizados três gases: CF_4, H_2 e O_2. O oxigênio foi utilizado apenas para limpeza da câmara, após cada processo, para remover o polímero depositado nas paredes da câmara, de modo a manter a reprodutibilidade do processo. Os gases CF_4 e H_2 quando combinados em ambiente de plasma, tendem a formar compostos poliméricos que formarão o filme fino sobre as amostras [3]. Deve-se encontrar a concentração correta dos gases (inclusive fluxo) que combinado com outros fatores (pressão, potência, tempo, etc.) realizam a polimerização da superfície. O polímero obtido com plasma de CF_4/H_2 é do tipo teflon [2,4].

PROCEDIMENTO EXPERIMENTAL

O plasma de CF_4/H_2 foi gerado em um reator de placas paralelas com configuração RIE (Fig. 1), com gerador de RF de 13,56 MHz. A distância entre os eletrodos é de 5,0 cm e o catodo de cobre recoberto com alumínio possui diâmetro de 15,0 cm. Os gases de processo são inseridos pela parte superior da câmara através de um distribuidor do tipo "chuveiro" com diâmetro de 15,0 cm. O sistema de vácuo é constituído por uma bomba mecânica e uma roots. A pressão de trabalho foi de 10 a 300 mTorr, com potência de 50 e 100W e fluxo total dos gases de 20 sccm.

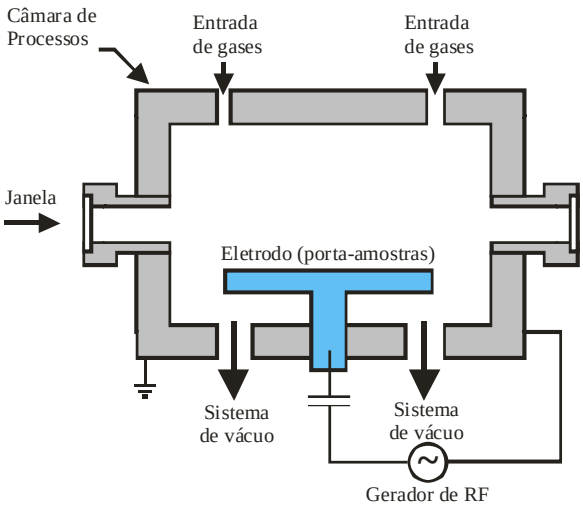


Fig. 1 - Representação esquemática do reator de plasma utilizado.

Para avaliar a hidrofobização do cone de papel foi utilizada a técnica de medição do ângulo de contato, que consiste em medir o ângulo que uma gota d´água forma sobre uma superfície, como mostra a Fig. 2. O processo de impermeabilização por plasma é bastante complexo, existindo uma janela de processo que fornece uma taxa de deposição maior em tempo menor. Vários parâmetros podem ser controlados, entre eles temos:

- **pressão de trabalho**: relacionada com a quantidade de moléculas de gás existente na câmara;
- **fluxo de gases**: relacionado com a quantidade de moléculas de gases reagentes injetada na câmara;
- **potência de RF**: relacionada com a taxa de ionização dos gases e conseqüentemente na formação de produtos;
- **proporção dos gases de processo**: no caso deste trabalho, a proporção entre o CF_4 e o H_2;
- **tempo de processo**: relacionado com a quantidade de filme fino formado.

Deve-se, portanto, encontrar os melhores parâmetros de processo que forneçam os melhores resultados. A medida do ângulo de contato qualificará esses parâmetros e que está relacionada com a tensão superficial do material em relação à solução empregada (água destilada) [5]. Quanto menor for o ângulo de contato, menor será a tensão superficial, portanto menor a aderência da solução ao substrato [6,7].

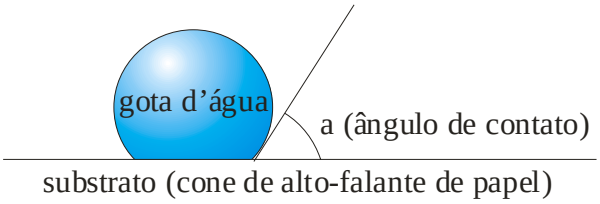


Fig. 2 – Método de obtenção do ângulo de contato.

O cone de papel foi cortado em pequenos retângulos de aproximadamente 2,5 x 3,5 cm^2 para serem colocados em porta-amostras de plástico (molduras plásticas utilizadas para projeção de diapositivos de 35 mm), como mostra a Fig. 3. Para cada processo, uma amostra foi colocada horizontalmente no centro do catodo. Algumas amostras foram colocadas perpendicularmente ao catodo, de modo a verificar a deposição de polímero fora da região da bainha do plasma.

As amostras foram pesadas antes e depois do processo por meio de uma microbalança, para verificar a variação da massa do papel ocorrida no processo de polimerização.

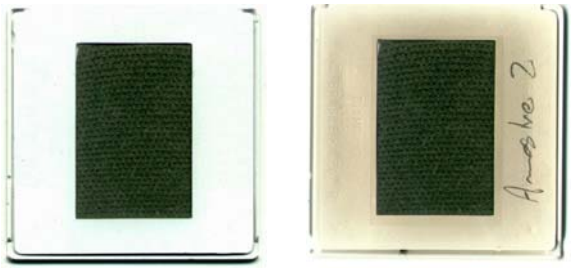


Fig. 3 – Porta amostra com pedaço de papel de cone. As fotos da esquerda e da direita correspondem às amostras antes e depois da polimerização, respectivamente. O amarelamento é devido à deposição de polímero na superfície.

RESULTADOS E DISCUSSÕES

Embora o cone de papel não apresentasse alteração visual devido a sua coloração escura, verificou-se que ocorreu impermeabilização da superfície em ambos os lados. Apesar do fluxo de gases estar direcionado apenas à sua face superior, os reagentes se deslocaram por baixo da amostra e formaram o filme fino polimérico nas costas da amostra. Isto indica que, se a técnica for aplicada em um processo industrial, será necessário apenas um processo para impermeabilizar todo o cone.

Na amostra deixada perpendicularmente ao catodo, também, ocorreu a polimerização, comprovando que, independentemente do ângulo, ocorre a deposição de polímero. Esse resultado garante que um cone inteiro pode ser impermeabilizado de uma só vez.

Variação de massa.

Foram processadas duas amostras de cones, com diferentes tamanhos e sem a utilização de molduras plásticas, com o objetivo de verificar a variação de massa das amostras no processo de polimerização, que estaria relacionada ao desempenho do alto-falante. Os resultados obtidos são mostrados na Tabela 1.

Tabela 1 – Massas das amostras antes e após o processo.

Amostra	antes do processo	após 5 min em vácuo (50mTorr)	após polimerização
1	0,3373g	0,3259g	0,3190g
2	0,3445g	0,3310g	0,3282g

Observa-se que as amostras perdem massa após ficarem em ambiente de baixa pressão devido à perda de umidade do papel (3% apenas) e que, após a polimerização há ligeira perda de massa (2%), ao contrário do que se esperava.

Esta perda de massa ocorre devido a redução da quantidade de água no material, causada pelo aumento da pressão de vapor em um ambiente de baixa pressão. Ou seja, quanto menor for a pressão de um ambiente, maior a proporção de água "evaporada".

Efeito da proporção de CF_4 e H_2.

Para obter a proporção de CF_4 e H_2 que resultasse na melhor impermeabilização do papel, realizou-se processos com variação desta relação. Fixou-se a potência em 50W, o fluxo total em 20sccm, a pressão em 100mTorr [8] e a concentração de H_2 em CF_4 foi variada entre 25% e 45%, com tempo de processo de 10 min. Os resultados obtidos estão mostrados na Fig. 4. Observa-se que o melhor processo ocorre para 35% de H_2 e 65% de CF_4, pois o ângulo de contato de 57,5° é o menor de todos.

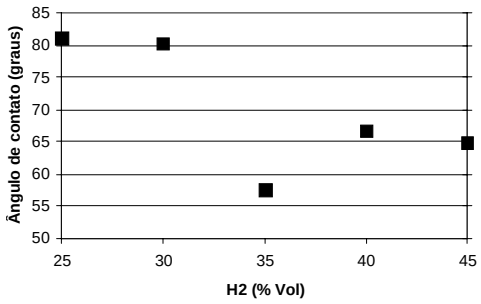


Fig. 4 – Ângulo de contato em função da porcentagem em volume de H_2 em CF_4 no processo de polimerização (100mTorr, 20sccm, 50W, 10 min.).

Efeito da pressão.

Fixando a proporção de 35% de H_2 e 65% de CF_4, a potência em 50W, o fluxo total em 20sccm e o tempo de processo em 10 min., foram realizados experimentos variando-se a pressão. A Fig. 5 mostra os resultados obtidos.

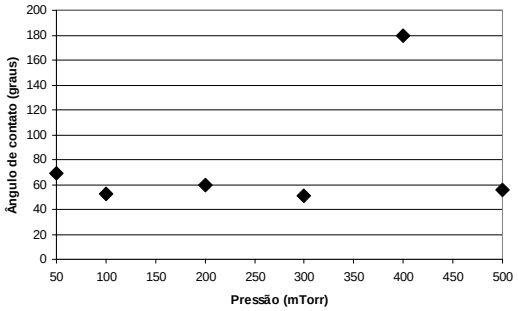


Fig. 5 – Ângulo de contato em função da pressão no processo de polimerização (35% H_2, 20sccm, 50W, 10min).

O menor ângulo de contato (melhor resultado) ocorre para a pressão de 100mTorr. Apesar da pressão de 300mTorr oferecer um valor de ângulo de contato menor, ela não oferece impermeabilização das costas das amostras, sendo portanto, desconsiderada. Para pressão de 400mTorr, a amostra processada absorveu água tanto na face superior, quanto inferior.

Pré-tratamento com plasma de oxigênio.

Algumas amostras passaram por um "pré-tratamento" com plasma de oxigênio (10 sccm, 50W e 100mTorr) para depois seguirem para o processo de polimerização (35% H_2, 20 sccm, 50W e 100mTorr). Isto forneceu os melhores resultados quando se realizou a medição do ângulo de contato (31,3º). Porém, verificou-se que o papel absorvia a água após algum tempo, apesar de seu ângulo de contato ser um dos melhores obtidos. Ao se realizar uma análise microscópica nas amostras, verificou-se que o plasma de oxigênio corroía todo o material entre as fibras do papel, variando de acordo com o tempo de processo: quanto maior o tempo de processo, maior a corrosão. Isto é bem caracterizado na literatura, onde o plasma de oxigênio é utilizado para corrosão de materiais orgânicos, pois reage com as moléculas de carbono do material formando CO_2 que é removido da câmara pelo equipamento de vácuo.

CONCLUSÃO

Os resultados demonstram que é possível realizar a impermeabilização dos cones de alto-falantes de papel com plasma de CF_4/H_2. Isto abre portas para novos estudos, relacionando, futuramente, diferentes materiais, texturas e espessuras para estudos sonoros (variação na resposta de freqüência do cone antes e depois do tratamento, por exemplo). A polimerização ocorre em todos os pontos do papel, apresentando a possibilidade de se realizar o processo em vários cones ao mesmo tempo com alteração não significativa da massa do cone. Isto é importante pois minimiza os agentes externos no momento da qualificação sonora, pois praticamente não haverá alterações no peso ou inércia do cone. Além disso, apresenta sua importância para uma possível realização deste processo em um ambiente industrial. Para pesquisas futuras, envolvendo qualificação sonora, recomenda-se utilizar domos de tweeters, pois são mais sensíveis à variações em sua construção física. Assim, garante-se maior confiabilidade nos estudos envolvendo filmes finos com áudio.

O melhor processo foi obtido nas condições de 35% de H_2, 65% de CF_4, 100W de potência, 100mTorr de pressão e 20sccm de fluxo total.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] S. Wolf and R.N. Tauber, "Silicon Fabrication for the VLSI Era", Volume 1 - Silicon Processing, Lattice Press, 1986 – p. 161-195.

[2] A. Bogaerts, E. Neyts , R. Gijbels , J. Mullen; "Gas discharge plasmas and their applications"; Spectrochimica Acta Part B 57 (2002) 609–658.

[3] E. J. Winder, K. K. Gleason; "Growth and characterization of fluorocarbon thin films grown from trifluoromethane (CHF_3) using pulsed-plasma enhanced CVD"; Insulating Biomaterials – NO1-NS-62350, Quarterly Progress Report 11; April-June 1999.

[4] E. Horanyi, Engineering Physics, University of Colorado, Boulder; P. Griffin, A. Agah, Electrical Engineering, Stanford University; "Teflon coatings for Biochips".

[5] P. M. Nogueira; "Estudo da viabilidade da utilização de filmes de carbono tipo diamante como camada de revestimento interno em dutos de transporte de petróleo"; Bol. téc. PETROBRÁS, Rio de Janeiro, 45 (2): abr./jun., 2002.

[6] A. Sklodowska, M. Wozniak, R. Matakowska; "The method of contact angle measurements and estimation of work of adhesion in bioleaching of metals" Biol. Proc. Online 1(3), 114-121; Vol. 1 No. 3 April 29, 1999 · www.biologicalprocedures.com.

[7] J. H. Choi, E. S. Lee , H. K. Baik, S. Lee, K. M. Song, M. K. Hwang , C. S. Huh; "Surface modification of natural leather using low-pressure parallel plate plasma"; Surface and Coatings Technology 171 (2003) 257–263.

[8] R. K. Yamamoto, F. R. de Almeida, H. S. Maciel; "Caracterização de filmes de polímero formado sobre o silício exposto ao plasma de CF_4/H_2"; Anais do 9º CBECIMAT – Laboratório de Sistemas Integráveis, Escola Politécnica da Universidade de São Paulo.

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na VIII Convenção Nacional
7 - 9 de Junho de 2004, 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.

Estado da Arte da Classificação de Sinais de Áudio

Jayme Garcia Arnal Barbedo, Amauri Lopes
Departamento de Comunicações - FEEC - UNICAMP
C.P. 6101, CEP: 13.083-970, Campinas - SP - Brasil, Tel: (19) 3788-3703;
jgab, amauri@decom.fee.unicamp.br

RESUMO

Este artigo apresenta uma compilação dos principais trabalhos desenvolvidos na área de classificação de sinais de áudio. São apresentadas algumas das técnicas mais conhecidas, bem como suas vantagens e pontos fracos. São sugeridas alternativas com potencial para superar alguns dos problemas enfrentados, visando fornecer subsídios para futuros projetos de pesquisa.

1. INTRODUÇÃO

Com o avanço da tecnologia da informação, a quantidade de dados gerada, armazenada e distribuída vem crescendo rapidamente. Essa nova realidade provoca uma demanda imediata por ferramentas capazes de tornar a manipulação desses dados mais simples e rápida. Nesse contexto, um novo campo de pesquisa, a Classificação de Sinais de Áudio (CSA), foi inaugurado. A CSA consiste na extração de parâmetros capazes de fornecer o máximo de informação a respeito de um sinal sonoro, e no uso de tais parâmetros para classificar o sinal de acordo com algum critério pré-estabelecido. Este é um campo de pesquisa amplo, composto por diversas sub-áreas, a maior parte das quais ainda pouco exploradas. Portanto, ainda são muitas as lacunas a serem preenchidas, o que representa uma ótima oportunidade para o desenvolvimento de novas tecnologias relevantes não apenas na esfera acadêmica, mas também na esfera comercial. Inúmeras aplicações podem se beneficiar diretamente de avanços alcançados em CSA, dentre as quais pode-se citar a criação de aparelhos auditivos seletivos, transcrição automática de música ou texto, remoção de ruído, sistemas especialistas de diagnóstico remoto, reconhecimento de voz e locutor, identificação de conteúdo de mídia, equalização automática de sinais musicais de acordo com o conteúdo, busca rápida em bancos de dados, entre outras. Muitas dessas aplicações têm demanda imediata.

Os primeiros trabalhos relevantes na área foram publicados em 1996 [1, 2], e desde então vem crescendo o número de pesquisadores envolvidos com o tema. Apesar disso, resultados satisfatórios só têm sido alcançados para aplicações muito específicas e restritas. As pesquisas que tentaram produzir classificadores mais gerais e robustos, capazes de enfrentar uma maior gama de situações, se depararam com dificuldades maiores que aquelas inicialmente previstas. Nesses casos, a taxa de acerto raramente atinge 60% [3].

A seguir será apresentada uma breve descrição das principais técnicas e estratégias utilizadas na classificação de áudio, bem como a bibliografia básica concernente a cada uma delas. O artigo está dividido da seguinte maneira: a Seção 2 apresenta as taxonomias (métodos de classificação) mais comumente adotadas; a Seção 3 apresenta alguns dos parâmetros utilizados na extração de informação dos sinais que se deseja classificar; a Seção 4 apresenta alguns dos mais métodos de classificação mais utilizados, cuja função é combinar os parâmetros extraídos de modo a resultar numa classificação confiável; a Seção 5 apresenta os resultados alcançados até agora pelas principais pesquisas na área e as perspectivas para pesquisas futuras; a Seção 6 faz uma análise crítica das estratégias propostas na literatura e sugere alternativas para que alguns dos problemas sejam superados; por fim, a Seção 7 apresenta as conclusões.

2. TAXONOMIA

A taxonomia é a maneira como se deseja classificar determinado conjunto de elementos. Ela pode ser composta de diversos níveis hierárquicos, onde os níveis mais baixos representam classificações mais específicas e restritas. No caso dos seres vivos, por exemplo, a classificação hierárquica começa, no seu nível mais elevado, com os reinos, descendo para filo, classe, ordem, família, espécie e sub-espécie. Na maior parte das vezes, é muito mais fácil distinguir entre um animal e uma planta, os quais pertencem a diferentes reinos, do que distinguir, por exemplo, entre duas sub-espécies de morcegos.

O mesmo fenômeno ocorre para sinais de áudio. Em geral, o primeiro nível hierárquico considerado neste tipo de classificação faz a distinção entre voz, música e ruído. Cada uma dessas classes possui peculiaridades que, se devidamente exploradas, permitem que elas sejam eficientemente diferenciadas. Apesar disso, mesmo neste nível hierárquico existem situações que representam verdadeiros desafios para os classificadores desenvolvidos. Um exemplo disso são músicas que muitas vezes contêm mais características de fala que de música, como é o caso do estilo musical conhecido como *rap*. No sentido oposto, tem-se o caso de sinais de voz que contêm características normalmente associadas à música, como o recital de uma poesia ou uma leitura dramática. Esses problemas vêm sendo especificamente abordados por alguns grupos de pesquisa [4, 5].

Na maior parte dos trabalhos, esta primeira divisão é apenas o ponto de partida para outros níveis de classificação com os mais diversos propósitos. Contudo, esta etapa tem, por si só, algumas aplicações importantes. Como exemplos, pode-se citar o monitoramento do conteúdo de rádios FM, codificação seletiva de voz e música e desativação de reconhecedores de voz em trechos não constituídos por voz [6]. Outros trabalhos que concentraram o foco neste nível de classificação incluem estudos para classificação diretamente no domínio codificado [7], bem como estudos visando fornecer requisitos para futuras pesquisas [8-11] (as referências [10] e [11] tiveram grande influência sobre trabalhos posteriores).

Na segunda camada hierárquica de classificação, os sinais já rotulados como voz ou áudio são divididos em sub-categorias. No primeiro caso, a divisão se limita a identificar vozes masculinas e femininas e, algumas vezes, voz com ruído. Este procedimento é útil sobretudo em sistemas para identificação de locutor e/ou segmentação de sinais de voz de acordo com diferentes locutores. Este tipo de classificação normalmente está inserido num contexto mais amplo, como pode ser visto em [3, 12]. Este tópico é também abordado em trabalhos que não se inserem no sistema hierárquico convencional, como será visto mais adiante.

No caso de sinais de música, a taxonomia varia consideravelmente, havendo desde divisões muito simples [13], onde há uma classificação em apenas três categorias, até taxonomias complexas, envolvendo grande número de classes e níveis hierárquicos [3, 14]. É claro que quanto mais complexo é o sistema de classificação, maior é o desafio, de modo que a taxa de acertos tende a diminuir. Contudo, apenas esses sistemas complexos têm aplicabilidade no mundo real. As taxonomias mais simples servem apenas ao propósito de acumular conhecimento para o desenvolvimento de sistemas mais gerais e robustos. A maior parte das pesquisas atuais apresenta classificações com grau intermediário de sofisticação [15, 16].

A Figura 1 mostra uma possível estrutura taxonômica, neste caso usando quatro níveis hierárquicos e dezesseis possíveis classificações finais para os sinais [14].

Um grande número de pesquisas tem se voltado para a classificação de sinais monofônicos, os quais são constituídos por apenas uma fonte acústica, seja ela um instrumento musical individual, a voz de uma pessoa, o som de algum animal, uma sirene, uma explosão, ou quaisquer outros efeitos sonoros. Este tipo de classificação tem aplicação direta na identificação do tipo de conteúdo de transmissões de rádio e televisão. Algumas pesquisas abordam apenas instrumentos musicais [17, 18], enquanto outras abordam uma grande variedade de efeitos [1, 19-27]. Além da aplicação acima citada, este tipo de pesquisa é uma boa fonte de idéias para o desenvolvimento de classificadores mais avançados.

Outra área que vem se beneficiando das pesquisas em CSA é o desenvolvimento de aparelhos auditivos mais avançados, nos quais eventos auditivos do dia-a-dia são identificados e classificados. Essa informação é usada pelo aparelho para decidir quais faixas de freqüência devem ser amplificadas ou atenuadas [28, 29].

A seção a seguir apresenta alguns dos parâmetros extraídos dos sinais de áudio.

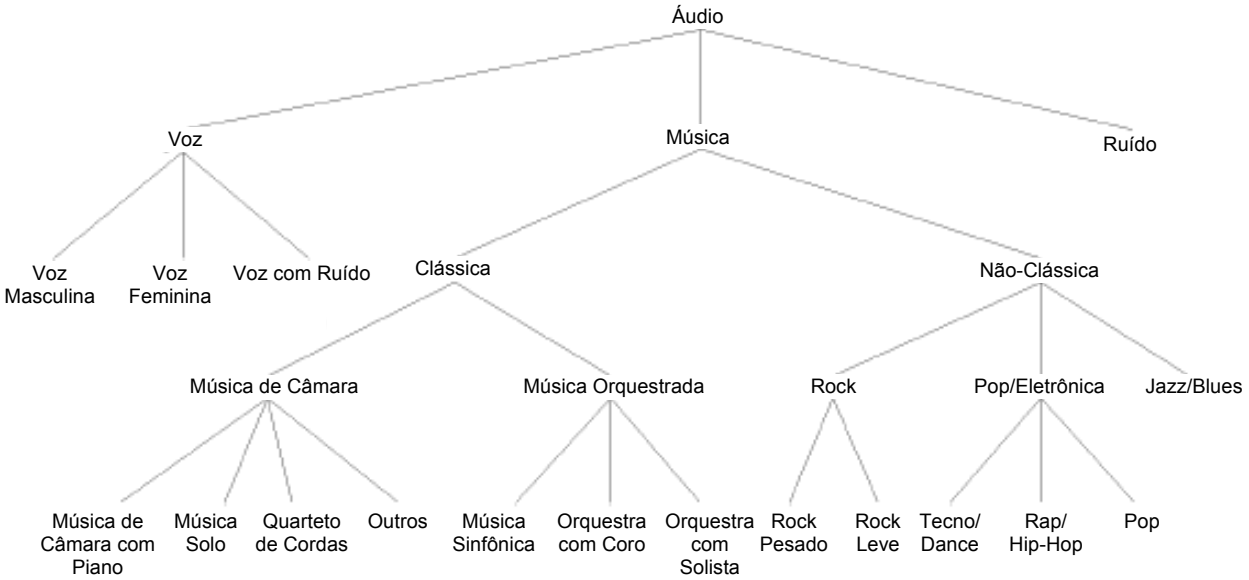


Figura 1 – Exemplo de estrutura taxonômica

3. EXTRAÇÃO DE PARÂMETROS

A primeira etapa em qualquer técnica de classificação de sinais é a extração de parâmetros capazes de fornecer o máximo de informação a respeito dos sinais que se deseja classificar. As pesquisas na área têm gerado uma grande quantidade de parâmetros descritores dos sinais, dentre os quais selecionou-se os mais utilizados para uma descrição resumida, como será visto a seguir. Não serão fornecidos detalhes matemáticos devido às limitações de espaço. Todos os parâmetros descritos a seguir são instantâneos, o que significa que são extraídos após a divisão dos sinais em quadros, a fim de se observar o comportamento dessas grandezas ao longo do tempo. O tamanho dos quadros varia de acordo com a pesquisa, mas raramente ultrapassa 100 ms.

3.1. Centróide Espectral

O centróide espectral vem sendo usado com sucesso em um grande número de trabalhos [1, 3, 12, 14, 15, 17, 30]. Ele representa o centro de gravidade da distribuição espectral de potência dos sinais. Altos valores para o centróide indicam texturas mais "brilhantes", com uma maior amplitude dos componentes de alta freqüência. Sinais de música, por serem espectralmente ricos, normalmente têm um valor de centróide mais elevado que voz.

3.2. Taxa de Cruzamentos por Zero

Este parâmetro foi adotado em [2, 3, 6, 9, 10, 11, 14, 15, 24]. Um cruzamento por zero ocorre quando as amplitudes de amostras temporais sucessivas têm sinais opostos. Em essência, a taxa de cruzamentos por zero fornece uma estimativa das propriedades espectrais dos sinais de áudio. Na diferenciação entre sinais de voz e música, é comum a utilização da variância desta grandeza como parâmetro.

3.3. Coeficientes Cepstrais

É, juntamente com o centróide e o número de cruzamentos por zero, o tipo de parâmetro mais usado [3, 10, 11, 14-16, 22, 23, 30]. As medidas cepstrais fornecem informação a respeito do formato do espectro. O termo cepstro é usado para designar a transformada de Fourier inversa do logaritmo do espectro de potência de um sinal. Em muitos casos, os próprios coeficientes cepstrais são usados como parâmetros, mas é comum a criação de variáveis derivadas, como o cepstro residual, o qual é obtido comparando-se a versão original do cepstro do sinal analisado com uma versão suavizada.

3.4. Pitch

A ANSI (*American National Standards Institute*) estabelece que *pitch* é o atributo auditivo de acordo com o qual os sons podem ser ordenados, em uma escala de freqüência, de baixo a alto. Este parâmetro perceptual está intimamente ligado ao conceito de freqüência fundamental, que é uma grandeza física absoluta. Apenas sinais periódicos, como música e voz, produzem a sensação de *pitch*. Esta grandeza é usada sobretudo para identificar sinais de voz, os quais possuem um valor de *pitch* característico. Esta abordagem é usada em [1, 11, 12, 15, 30].

3.5. Largura de Faixa

Este parâmetro representa a largura da faixa de freqüência em Hz que o sinal ocupa. Sinais de voz, em geral, possuem o espectro concentrado nas freqüências de 100 Hz a 7 kHz, enquanto a maior parte dos sinais de música está compreendida entre 15 Hz e 16 kHz [12]. Sinais de ruído geralmente possuem componentes espalhados ao longo de todo o espectro. A largura de faixa é usada em [1, 12, 15, 17, 30].

3.6. Fluxo Espectral

O fluxo espectral é definido como a diferença quadrática entre os espectros de potência de quadros consecutivos do sinal, de modo a caracterizar as mudanças no formato do espectro. Em geral, sinais de voz apresentam mudanças mais drásticas que sinais de música, devido à alternância entre fonemas surdos e sonoros. Assim, o fluxo espectral normalmente assume valores mais elevados para voz. Este parâmetro foi adotado em [3, 9, 10, 14].

3.7. Proporção de Quadros com Baixa Energia

Este parâmetro mede a proporção de quadros cuja potência RMS é menor que metade da potência RMS média ao longo de um quadro de 1 segundo. Normalmente, sinais de voz possuem uma maior proporção de quadros de baixa energia que sinais de música, situação em que esta variável assume valores mais elevados. Este parâmetro foi adotado em [3, 9, 10].

3.8. Sonoridade

Este parâmetro busca modelar a percepção humana em relação à sonoridade do sinal a ser classificado. Normalmente as formulações utilizadas nos métodos de classificação de áudio são muito simples, consistindo basicamente no cálculo do nível RMS de cada quadro do sinal. Obviamente, estas formulações não são suficientes para modelar de maneira precisa a percepção humana, mas fornecem estimativas suficientemente confiáveis para o propósito de classificação. Esta é uma grandeza muito útil na identificação do gênero musical, e é usada em [1, 14, 15].

3.9. Harmonicidade

Normalmente, a música possui um grau de harmonicidade maior que os demais sinais. Esta grandeza é também útil para diferenciar gêneros musicais. Este parâmetro foi adotado em [1, 12, 14, 17].

3.10. Ponto de Roll-Off

Este valor mede a freqüência abaixo da qual se encontra 95% da potência total do sinal. Sinais de música, devido a sons percursivos, transientes de ataque, etc., têm mais energia concentrada nas altas freqüências que sinais de voz. Portanto, este parâmetro apresenta valores mais elevados em sinais de música. O ponto de "rolloff" foi usado em [3, 10, 14, 15].

3.5. Outros

Pesquisas indicam que muitas vezes as estatísticas de segunda ordem são mais eficientes em caracterizar as diferenças entre voz e música que os próprios parâmetros que as originam [3, 15]. Assim, além dos parâmetros descritos, suas variâncias muitas vezes fornecem informação valiosa para uma classificação mais precisa dos sinais.

A lista de parâmetros usados para extrair informação dos sinais inclui ainda a taxa de silêncio [12], métrica de ritmo [3, 7], energia de modulação a 4 Hz [8, 10], entropia da modulação [8], energia das sub-bandas [15, 30], etc. Apesar de usados em menor escala, estes parâmetros possuem características que podem ser úteis em muitas aplicações.

4. MÉTODOS DE CLASSIFICAÇÃO

Os métodos de classificação têm como função básica combinar a informação contida nos parâmetros extraídos de modo a produzir uma classificação confiável para o sinal analisado. Como no caso dos parâmetros, as pesquisas têm

gerado um grande número de diferentes técnicas de classificação. A seguir são apresentadas as técnicas mais usadas. Novamente, os detalhes matemáticos serão omitidos devido às limitações de espaço.

4.1. K Vizinhos mais Próximos

O primeiro passo na implementação deste tipo de classificador é a construção de um dicionário de vetores contendo valores paramétricos que representem as categorias nas quais se deseja classificar os sinais. Assim, cada classe será representada por um determinado número de vetores, os quais contêm valores de parâmetros tipicamente encontrados para a classe em questão. Quanto ao processo de classificação, o vetor de parâmetros extraído do sinal é comparado com os vetores presentes no dicionário. Nessa comparação, utiliza-se uma medida de distância (geralmente euclidiana) para determinar os *k* vetores do dicionário mais próximos do vetor de parâmetros do sinal a ser classificado. Por fim, o sinal é classificado de acordo com a classe que mais vezes está representada entre os *k* vetores determinados na etapa anterior. Este é um classificador muito usado [6, 9, 10, 13, 17, 18], principalmente devido à facilidade de implementação. Por outro lado, ele possui a desvantagem de exigir que se armazene um grande número de vetores de treinamento, resultando em um grande número de computações. Tal fato torna seu uso desaconselhável em aplicações que precisem operar em tempo real.

4.2. Modelos de Mistura Gaussiana

Os modelos de mistura Gaussiana (MMG) têm sido usados com sucesso em muitos problemas de classificação de áudio [3,10,11,16,18]. A técnica é usada para modelar as diferentes classes e gêneros de áudio como uma função densidade de probabilidade (FDP), usando uma combinação ponderada de FDPs Gaussianas (misturas). Os parâmetros dos MMG são ajustados de acordo com um conjunto de treinamento. Esse é um processo iterativo, e a técnica utilizada para esse fim é conhecida como método de Maximização de Expectativa [31]. Após o treinamento, os MMGs resultantes são usados para determinar as probabilidades de um vetor de parâmetros pertencer a cada uma das classes. O maior valor de probabilidade identifica a classe à qual pertence o sinal.

4.3. Modelos Ocultos de Markov

Os modelos ocultos de Markov são muito usados na segmentação de sinais de áudio em diferentes eventos acústicos. Esses modelos são autômatos probabilísticos de estado finito com uma saída, e consistem de três componentes básicos: um estado inicial de distribuição, uma matriz de transição de estado e uma função densidade de observação. A saída do modelo é constituída por uma seqüência de estados e um valor de verossimilhança que mede a probabilidade que cada observação (vetor de parâmetros) tem de pertencer a cada modelo. Como no caso do MMG, o modelo que resulta na maior probabilidade revela a classe à qual pertence o sinal. Esta abordagem é adotada em [19, 21, 25].

4.4. Máquinas de Vetores Suporte (Support Vector Machines)

Este tipo de classificador foi desenvolvido recentemente como uma abordagem alternativa para o problema da aprendizagem. Seu objetivo é encontrar o hiperplano que melhor separa observações pertencentes a diferentes classes num espaço multidimensional. A teoria não garante que o melhor hiperplano possa ser sempre encontrado mas, na prática, uma solução heurística pode sempre ser alcançada. Este não é um classificador simples de ser implementado, mas possui diversas características desejáveis, como a capacidade de controlar a complexidade da aprendizagem, independente da dimensão do problema. Esse método foi analisado em [13, 17, 22].

4.5. Discriminador Binário Simples

Este tipo de classificação é realizado através da divisão sucessiva dos sinais em dois grupos. A decisão entre um ou outro grupo é tomada de acordo com os parâmetros considerados em cada divisão. Por exemplo, os sinais de áudio podem, numa primeira divisão, ser classificados em voz e silêncio, tendo como base os valores das variâncias do centróide espectral e do número de cruzamentos por zero. Numa segunda divisão, os sinais de voz podem ser classificados em masculinos ou femininos, de acordo com os valores de *pitch*, e os de música em populares ou clássicos, de acordo com a harmonicidade. As músicas populares, por sua vez, podem ser classificadas em agitadas ou calmas, e assim sucessivamente. Este tipo de classificador é muito simples, ideal para ser implementado em tempo real. Contudo, ele não é apropriado para a realização de classificações muito sofisticadas. Seu desempenho foi testado em [7, 8, 24].

4.6. Outros

Existem muitos outros sistemas de classificação além dos acima descritos. Dentre eles, pode-se citar o classificador Gaussiano Quadrático [6, 15], classificação nebulosa [4, 5], distância euclidiana simples [1], etc. A escolha do classificador a ser utilizado depende da aplicação pretendida, bem como das preferências pessoais do projetista.

5. ESTÁGIO DE DESENVOLVIMENTO DA CSA

Ao se fazer uma análise da qualidade dos resultados alcançados na área de CSA, deve-se tomar o cuidado de dividir os trabalhos de acordo com o grau de dificuldade da problemática abordada. Obviamente, pesquisas abordando simplesmente a distinção entre sinais de voz e música têm muito mais facilidade de alcançar resultados satisfatórios que aquelas que se propõem a fazer classificações mais complexas. Além disso, algumas pesquisas trabalham com a limitação de desenvolver estratégias que possam ser implementadas em tempo real, o que reduz significativamente a flexibilidade do projeto. Assim, a complexidade dos desafios enfrentados em cada pesquisa varia significativamente, tornando praticamente impossível uma comparação direta entre elas. O que se pretende nesta seção é fornecer uma visão geral do estágio de desenvolvimento das pesquisas na área, revelando assim quais desafios já foram vencidos e quais ainda estão por ser resolvidos.

Como comentado anteriormente, os resultados alcançados na distinção entre voz e música têm sido bastante satisfatórios. A taxa de acertos média neste tipo de aplicação freqüentemente alcança valores superiores a 95% [3, 7]. Exceções ocorrem para casos em que há uma preocupação maior com a execução em tempo real, onde a taxa de acertos pode cair para valores próximos a 80% [12]. O grande desafio neste nível de classificação é identificar corretamente sinais cuja estrutura mescla características tanto de voz como de música, como é o caso de gêneros musicais como o rap e a declamação de textos poéticos. Alguns trabalhos propõem que se criem classificações intermediárias entre música e voz. Esse é o caso de [4, 5], onde o autor usa técnicas de lógica nebulosa (fuzzy) para realizar uma classificação relativa, do tipo “30% voz e 70% música”. Contudo, este tipo de

estratégia ainda necessita de muitos ajustes até que atinja a maturidade.

Dentre as classificações realizadas no segundo nível hierárquico, a mais simples é a divisão em voz masculina e feminina. O índice de acertos para este tipo de classificação é também bastante elevado, muitas vezes alcançando valores próximos a 98% [3]. As estratégias utilizadas para este fim estão bem estabelecidas. Muitas das pesquisas neste nível de classificação incluem, além da distinção entre vozes masculinas e femininas, a identificação de outros eventos sonoros monofônicos, como sons produzidos por animais, por fenômenos meteorológicos, pela interação entre os seres humanos e o ambiente, etc. Apesar de representar um desafio maior, este tipo de classificação é facilitado pelo fato dos sinais monofônicos, em geral, apresentarem uma estrutura simples e previsível. Como resultado, os índices de acerto freqüentemente ultrapassam os 90% [3, 12].

Uma tarefa muito mais complicada é a classificação de sinais de música. Esse tipo de sinal é, em geral, polifônico, o que significa que ele é constituído por um grande número de diferentes fontes sonoras simultâneas. Sua complexidade torna difícil a distinção entre os diferentes gêneros musicais. Como descrito na Seção 2, a complexidade da classificação de sinais de música varia muito entre as pesquisas realizadas. Alguns trabalhos adotam taxonomias muito simples [13]; nesses casos, o objetivo principal é acumular conhecimento para o futuro desenvolvimento de estratégias mais complexas. Outras pesquisas adotam uma taxonomia complexa [3, 14], já visando o desenvolvimento de ferramentas que venham atender à demanda por novas tecnologias ocasionada pela evolução da tecnologia da informação. Nesses casos, o índice de sucesso ainda é baixo, raramente atingindo 60%. Como se pode observar, ainda há muito a ser feito até que esta se torne uma tecnologia consolidada.

6. ANÁLISE CRÍTICA E SUGESTÕES

A seguir será apresentada uma análise crítica das estratégias comumente utilizadas, além de se propor soluções para alguns de seus principais problemas.

6.1. Taxonomia

A classificação taxonômica adotada em cada pesquisa é fortemente subjetiva. Esta afirmação é particularmente verdadeira para música. Neste caso, não existe um consenso na maneira como os sinais devem ser classificados. Isto gera uma grande confusão de semântica e um grande volume de redundância dentro das próprias estruturas taxonômicas.

Esses fatos tornam claro que desenvolver uma taxonomia apropriada e eficiente não é uma tarefa trivial. A tentativa de se criar uma estrutura taxonômica universal é um desafio ainda maior. Apesar das dificuldades, um método de classificação de sinais de áudio realmente eficiente dependerá do sucesso de tal empreitada. Hoje, existem pesquisas voltadas exclusivamente para o desenvolvimento de técnicas capazes de padronizar a maneira como os sinais são classificados [32].

Um meio de evitar alguns dos problemas relacionados a este tópico é desenvolver uma taxonomia com diversos níveis hierárquicos, ou seja, as categorias são sucessivamente divididas em sub-categorias. Este procedimento confere uma boa flexibilidade à estrutura taxonômica, pois facilita a inserção de novos gêneros musicais à medida que o contexto musical evolui, além de permitir que modificações sejam facilmente introduzidas.

6.2. Parâmetros

Um dos principais problemas encontrados nas estratégias desenvolvidas até agora é o fato delas se basearem na extração de parâmetros de baixo nível. Tal fato ocorre tanto por questões de facilidade de implementação, quanto pelo compromisso que algumas pesquisas têm de desenvolver técnicas passíveis de serem implementadas em tempo real. Isso limita o sucesso a taxonomias muito pequenas (até 5 classes) e à discriminação de gêneros bastante distintos (como música clássica e tecno). Contudo, especialmente em classificações mais detalhadas, há a necessidade de se extrair parâmetros de nível mais elevado [32].

Os sistemas de classificação automática de sinais de áudio buscam reproduzir a maneira como ouvintes humanos processam e classificam os eventos auditivos que chegam a seus ouvidos. Assim, pode-se inferir que uma das possíveis estratégias para se extrair parâmetros de alto nível é submeter os sinais a processamentos que modelem alguns fenômenos auditivos importantes, como mascaramento e limiar de audibilidade. O conjunto de tais processamentos recebe o nome de modelos psico-acústicos. Os parâmetros extraídos a partir desses sinais perceptualmente modificados seriam uma fonte de informação muito mais confiável.

Outras alternativas em estudo para a extração de parâmetros de boa qualidade incluem a análise de ritmo da música e a busca por estratégias alternativas para a decomposição tempo-freqüência dos sinais.

6.3. Estratégias de Classificação

Como comentado na Seção 4, pode-se encontrar na literatura diversas estratégias de classificação. As técnicas tradicionalmente usadas para este fim podem ser divididas em dois tipos: 1) aquelas que usam abordagens muito simples a fim de possibilitar a execução em tempo real, em detrimento da qualidade; 2) aquelas que usam abordagens mais sofisticadas, as quais resultam em melhores resultados mas não podem ser implementadas em tempo real.

Uma alternativa com potencial de substituir com vantagem as técnicas clássicas é a classificação através das redes de Kohonen. Essa abordagem, usada com sucesso em pesquisas anteriores [33, 34], fornece resultados de boa qualidade aliados a um esforço computacional comparativamente pequeno. As novas estratégias de classificação a serem desenvolvidas no futuro deverão adotar a estratégia de aliar bons resultados a uma baixa complexidade computacional.

6.3. CONCLUSÕES

Este artigo apresentou o estado da arte da classificação de sinais de áudio. Foram apresentadas diferentes estratégias relacionadas à taxonomia, extração de parâmetros e métodos de classificação usados nesta área de pesquisa. As falhas concernentes a essas estratégias foram identificadas e alternativas com potencial de superar alguns dos problemas foram propostas.

A classificação de sinais de áudio é um campo de pesquisa em franca expansão. Por estar nos estágios iniciais de desenvolvimento, a CSA é uma área estimulante e ainda repleta de desafios a serem vencidos.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] E. Wold , T. Blum , D. Keislar and J. Wheaton, “Content-Based Classification, Search, and Retrieval of Audio”, *IEEE MultiMedia*, vol. 3, no. 3, pp. 27-36, September 1996.

[2] J. Saunders, “Real-Time Discrimination of Broadcast Speech/Music”, Proceedings of the IEEE Intern. Conf.

on Acoustics, Speech, and Signal Processing, vol. 2, 1996, pp 993-996.

[3] G. Tzanetakis and P. Cook, "Musical Genre Classification of Audio Signals", *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, July 2002.

[4] D. B. Gerhard, "A human vocal utterance corpus for perceptual and acoustic analysis of speech, singing and intermediate vocalizations", *Journal of the Acoustical Society of America*, vol. 112, no. 5, p. 2246, November 2002.

[5] D. B. Gerhard, *Computationally measurable differences between speech and song*, PhD Thesis, Computing Science, Simon Fraser University, Burnaby, Canada, April 2003.

[6] K. El-Maleh, M. Klein, G. Petrucci and P. Kabal, "Speech/Music Discrimination for Multimedia Applications", *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing* (Istanbul), pp. 2445-2448, June 2000.

[7] R. Jarina, N. O'Connor, and S. Marlow, "Rhythm Detection for Speech-Music Discrimination in MPEG Compressed Domain", *Proc. of the IEEE 14th International Conference on Digital Signal Processing 2002*, Santorini, Greece, July 2002, pp. 129-132

[8] J. Pinquier, J.-L. Rouas and R. André-Obrecht, "Robust speech / music classification in audio documents", *Proceedings of the International Conference on Spoken Language Processing (ICSLP'2002)*, Denver, USA, September 2002.

[9] L. Lu, H.-J. Zhang and Hao Jiang "Content Analysis for Audio Classification and Segmentation", *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 7, pp. 504-516, October 2002.

[10] E. Scheirer and M. Slaney: "Construction and Evaluation of a Robust Multifeature Speech/Music Discriminator", *Proceedings of ICASSP97*, pp. 1331-1334, April 1997, Munich, Germany.

[11] M. J. Carey, E. S. Parris and H. Lloyd-Thomas, "A comparison of features for speech, music discrimination", *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 149–152, April 1999.

[12] Lu, G.J. and T. Hankinson, "A Technique Towards Automatic Audio Classification and Retrieval", Proc. IEEE Intl. Conf. on Signal Processing, vol. 2, pp. 1142--1145, 1998.

[13] H. Deshpande, R. Singh, and U. Nam. "Classification of music signals in the visual domain", *Proceedings of the COST-G6 Conference on Digital Audio Effects*, Limerick, Ireland, December 2001.

[14] J. J. Burred and A. Lerch, "Hierarchical Approach to Automatic Musical Genre Classification", *Proceedings of the 6th International Conference on Digital Audio Effects DAFX03*, London, UK, September 2003.

[15] M. F. McKinney, "Features for Audio and Music Classification", *Proceedings of ISMIR*, Baltimore, USA, 2003.

[16] D. Pye, "Content-based methods for the management of digital music", *In ICASSP*, vol. IV, pages 2437-2440, 2000.

[17] G. Agostini, M. Longari and E. Pollastri, "Musical Instrument Timbres Classification with Spectral Features", *EURASIP Journal on Applied Signal Processing*, no. 1, pp. 1-11, 2003.

[18] G. Peeters and X. Rodet, "Automatically selecting signal descriptors for Sound Classification", *ICMC 2002*, Goteborg, Sweden, September 2002.

[19] M. Casey, "*General sound classification and similarity in MPEG-7*", *Organized Sound*, vol. 6, no. 2, pp. 153-164, Aug. 2001.

[20] Y.-C. Cho, S. Choi and S.-Y. Bang, "Non-negative component parts of sound for classification", *Proc. IEEE Int. Symp. Signal Processing and Information Technology*, Darmstadt, Germany, December 14-17, 2003.

[21] H.-G. Kim, E. Berdahl and T. Sikora, "Study of MPEG-7 Sound Classification and Retrieval", *5th International ITG Conference on Source and Channel Coding*, Erlangen, Germany, January 2004.

[22] D. Guo and S.Z. Li, "Content-based Audio Classification and Retrieval by Support Vector Machines". *IEEE Trans. on Neural Networks*. Vol.14, No.1, pp.209-215. January 2003.

[23] J. Foote, "Content-Based Retrieval of Music and Audio", in C.-C. J. Kuo et al., editor, *Multimedia Storage and Archiving Systems II, Proc. of SPIE*, Vol. 3229, pp. 138-147, 1997.

[24] T. Zhang and C.-C. J. Kuo, "Audio content analysis for online audiovisual data segmentation and classification", *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 4, pp. 441-457, 2001.

[25] A. A. Alatan, A. N. Akansu and W. Wolf, "Multi-modal Dialogue Scene Detection Using Hidden Markov Models for Content-based Multimedia Indexing", *Kluwer Acad., Int. Journal on Multimedia Tools and Applications*, vol. 14, pp. 137–151, 2001.

[26] N. Dimitrova, "Multimedia Content Analysis and Indexing for Filtering and Retrieval Applications", *Special Issue on Multimedia Technologies and Informing Systems, Part I*, vol. 2, no. 4, pp. 87-100, 1999.

[27] K. Minami, A. Akutsu, H. Hamada and Y. Tonomura, "Video Handling with Music and Speech Detection", *IEEE MultiMedia*, vol. 5, no. 3, pp.17-25, July 1998.

[28] V. Peltonen, J. Tuomi, A. Klapuri, J. Huopaniemi and T. Sorsa, "Computational auditory scene recognition", *Proceedings of ICASSP '02*, Florida, USA, May 2002.

[29] M. C. Büchler, *Algorithms for Sound Classification in Hearing Instruments*, Ph.D. Thesis, Zurich, Switzerland, 2002.

[30] S. Z. Li, "Content-based classification and retrieval of audio using the nearest feature line method", *IEEE Transactions on Speech and Audio Processing*, vol. 8, pp. 619–625, 2000.

[31] L. R. Rabiner, *Fundamentals of Speech Recognition*, PTR Prentice-Hall Inc., New Jersey, 1993.

[32] J.-J. Aucouturier and F. Pachet, "Representing Musical Genre: A State of the Art", *Journal of New Music Research*, vol. 32, no. 1, pp. 83–93, 2003.

[33] J. G. A. Barbedo, A. Lopes, "A New Cognitive Model for Objective Assessment of Audio Quality", *submitted to the Journal of Audio Engineering Society*, March 2004.

[34] J. G. A. Barbedo, A. Lopes, *Strategies to Increase the Applicability of Methods for Objective Assessment of Audio Quality*, to be presented at 116th AES Convention, Berlin, May 2004.

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na VIII Convenção Nacional
7 - 9 de Junho de 2004, São Paulo, SP, Brasil

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.

Interpolação de Sinais de Áudio Usando Polinômios do Par de Linhas Espectrais

Paulo A. A. Esquef[1]
[1]Helsinki Univ. of Technology, Lab. of Acoustics and Audio Signal Processing, P.O.Box 3000, FIN-02015 HUT, Espoo, Finland
paulo.esquef@hut.fi

RESUMO

Esse artigo aborda a interpolação de falhas longas em sinais de áudio através de técnicas basedas em modelos autorregressivos (AR). Demonstra-se que a utilização de modelos AR modificados, baseados na soma ponderada dos polinômios do par de linhas espectrais (PPLE), produz melhores resultados perceptivos do que o emprego de modelos convencionais de mesma ordem.

1 INTRODUÇÃO

A reconstrução de amostras perdidas em sinais de áudio é uma tarefa comumente necessária em aplicações de áudio. Exemplos de perda de sinal são encontrados em situações como erros em uma transmissão digital [1] e reprodução de matrizes degradadas de gravações históricas [2].

A modelagem autorregressiva (AR) encontra utilidade em várias aplicações no processamento de sinais de fala e música. Exemplos típicos são a codificação por predição linear [3], a deteção de erros [4] e a reconstrução de sinais de áudio [5, 6]. No último caso, interpoladores AR são adequados para reconstruir fragmentos muito curtos de sinais. Isso se deve, basicamente, ao fato de a suposição de estacionariedade ser válida somente para trechos curtos, i.e., aproximadamente 20 ms, de sinal.

Na interpolação de falhas longas, o desempenho de interpoladores AR é fraco, uma vez que a energia do sinal interpolado decai em direção ao meio do segmento a ser reconstruído. Isso se deve, principalmente, à minimização do erro de modelagem, empregada na estimação das amostras perdidas. Considerando uma situação limite em que o erro é nulo, o sinal interpolado dependerá basicamente da resposta impulsiva do modelo AR. Assim, para que a interpolação funcione, é preciso que o decaimento da resposta impulsiva do modelo seja lenta o suficiente para preencher a falha. Para modelos cujos pólos se situem afastados da borda do círculo unitário e, logo, possuam respostas impulsivas de decaimento rápido, o efeito de atenuação energética se torna crítico.

Uma opção simples para se contornar esse problema consiste em aumentar a ordem do modelo AR, visto que isso tende a produzir modelos com pólos mais próximos da circunferência unitária. Outras soluções incluem a imposição de um limite inferior para a minimização do erro de modelagem [7, 8].

Para fragmentos tão longos quanto o limite de estacionariedade de tempo curto é recomendável a utilização

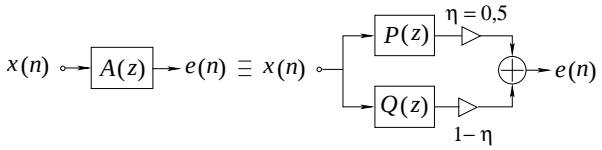


Figura 1: Equivalência entre $A(z)$ e os PPLEs.

de dois modelos AR: um estimado a partir do segmento que imediatamente precede a falha e outro a partir do segmento que imediatamente a sucede [9]. Abordagem similar foi também proposta em [10, 11, 12].

O interpolador investigado nesse trabalho é baseado na proposição descrita em [10, 11, 12]. Aqui, entretanto, ao invés de se utilizarem modelos AR convencionais, modelos AR modificados são empregados. A modificação consiste em computar os PPLEs associados ao modelo AR convencionalmente estimado e obter um modelo AR modificado através da soma ponderada daqueles [13]. O modelo modificado é, então, usado diretamente no algoritmo de interpolação.

A vantagem dessa proposta reside na possibilidade de, através do ajuste de um único parâmetro, se realocar os pólos do modelo modificado mais proximamente da circunferência unitária. Isso resulta num benefício claro ao problema de interpolação. Um efeito colateral da modificação proposta é que as freqüências dos pólos realocados se tornam polarizadas em relação aos valores originais. Conseqüentemente, distorções podem ser percebidas nas porções de sinal reconstruídas (ver mais detalhes na Seção 5). Apesar disso, o emprego dos modelos AR modificados produz resultados superiores aos obtidos com o uso de modelos convencionais de mesma ordem.

Este artigo é organizado como se segue. A Seção 2 define os polinômios do par de linhas espectrais e revê suas propriedades. A Seção 3 ilustra as propriedades dos modelos modificados. O método de interpolação utilizado é descrito na Seção 4. Os resultados experimentais são apresentados na Seção 5 e avaliados subjetivamente na Seção 6. As conclusões são apresentadas na Seção 7.

2 POLINÔMIOS DO PAR DE LINHAS ESPECTRAIS

Suponha a resposta de um filtro de predição linear de p-ésima ordem como $A(z) = 1 - \sum_{k=1}^{p} a(k) z^{-k}$, onde $a(\cdot)$ representam os coeficientes do filtro na sua forma direta. A função de transferência do filtro de síntese $1/A(z)$ é garantidamente estável se $A(z)$ possuir fase mínima.

A idéia por trás dos PPLEs é representar o polinômio $A(z)$ por meio de dois outros polinômios cujos zeros se encontram exatamente sobre a circunferência unitária. O mapeamento envolvido é dado por $P(z) = A(z) + z^{-(p+1)} A(z^{-1})$ e $Q(z) = A(z) - z^{-(p+1)} A(z^{-1})$, onde $P(z)$ e $Q(z)$ são as asssim chamados polinômios do par de linhas espectrais [14]. Ademais, $A(z)$ pode ser reconstruído a partir de $P(z)$ e $Q(z)$ através da relação $A(z) = \frac{1}{2}[P(z) + Q(z)]$. A equivalência entre $A(z)$ e a soma ponderada de $P(z)$ e $Q(z)$ é ilustrada na Fig. 1.

3 SOMA PONDERADA DE PPLES

Um esquema de predição linear baseado na soma ponderada dos PPLEs foi proposto em [13]. O filtro de análise é definido como

$$D(z, \eta) = \eta P(z) + (1 - \eta) Q(z), \tag{1}$$

onde, como antes, $P(z)$ e $Q(z)$ são, respectivamente, os PPLEs simétrico e assimétrico associados a $A(z)$. O peso η é um ganho de valor real. Como visto na Seção 2 e na Fig. 1, para $\eta = 0{,}5$ a equivalência $D(z, \eta = 0{,}5) = A(z)$ é verdadeira. Deve-se notar que, para $\eta \neq 0{,}5$, $D(z, \eta)$ tem ordem $p + 1$, enquanto $A(z)$ tem ordem p.

Para $\eta = 1$ e $\eta = 0$, o preditor modificado $D(z, \eta)$ se reduz a $P(z)$ e $Q(z)$, respectivamente. Logo, os pólos de $1/D(z, \eta)$ estão localizados na circunferência unitária. Além disso, mostra-se em [13] que $1/D(z, \eta)$ é estável se $\eta \in \,]\,0, 1\,[$. Por simplicidade, o modelo AR $1/D(z, \eta)$ será chamado, daqui em diante, de modelo modificado.

A figura 2 mostra o lugar-das-raízes de $1/D(z, \eta)$ em função do peso η, para um modelo AR sintético de quarta ordem. Uma inspeção visual da Fig. 2 revela que escolher η perto de 1 faz com que os ângulos dos pólos de $1/D(z, \eta)$ permaneçam (em baixas freqüências) mais próximos dos ângulos do modelo AR original do que no caso de se adotar η próximo de 0. Ademais, adotar $\eta = 0$ implica um pólo em DC, o que pode ser problemático em aplicações de síntese, enquanto que $\eta = 1$ implica um pólo em $z = -1$.

Dado o arrazoado anterior, é plausível esperar-se que a adoção de η tendendo a 1 seja mais vantajosa que escolher η perto de 0, como recurso para mover os pólos do modelo modificado mais para perto da circunferência unitária. Alternativamente, pode-se pensar em computar as raízes de $A(z)$ e artificialmente aumentar seus módulos (raios). Entretanto, o cálculo das raízes de polinômios de alta ordem, além de ser computacionalmente custoso, é sujeito a erros numéricos. Ao contrário, o custo computacional para se computar o modelo modificado é de apenas p multiplicações e p adições por modelo, ou seja, é negligível se comparado ao custo da estimação do modelo convencional.

4 MÉTODO DE INTERPOLAÇÃO MODIFICADO

Os modelos AR modificados podem ser utilizados em qualquer interpolador baseado em modelagem AR, e.g., o interpolador AR por critério de mínimos quadrados descrito em [6]. Como o presente trabalho lida com a interpolação de falhas tão longas quanto o máximo tempo de estacionariedade que se pode assumir em fragmentos

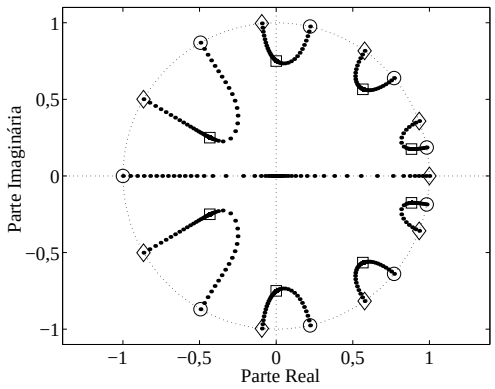


Figura 2: Lugar-das-raízes de $1/D(z,\eta)$ em função de $0 \leq \eta \leq 1$, para um modelo AR sintético de quarta ordem. Os símbolos $\square$, $\Diamond$ e $\bigcirc$ indicam a localização dos pólos para $\eta = 0{,}5$, $\eta = 0$ e $\eta = 1$, respectivamente.

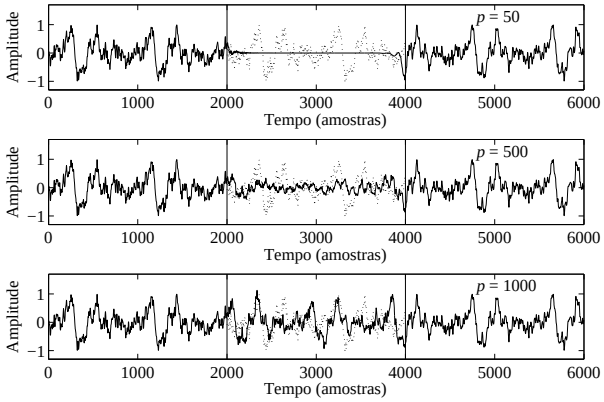


Figura 3: Desempenho do método de interpolação para diversos valores de ordem de modelo. A falha está presente entre as amostras 2001 e 4000. O sinal original na região da falha é mostrado em linha pontilhada.

de sinais de áudio, o esquema de interpolação proposto em [11] será ligeiramente modificado e utilizado nos experimentos.

Supondo uma falha de G amostras entre dois segmentos de N amostras cada um, os passos do método modificado de interpolação são os seguintes:

1. Estime um modelo AR a partir do segmento de N amostras que precede a falha e calcule o modelo modificado usando η perto de 1. A estimação do modelo AR é aqui realizada pelo método de Burg [15].
2. Extrapole para a frente o sinal através falha por meio da excitação do modelo modificado, como descrito em [12].
3. Repita os dois passos anteriores para o segmento que sucede a falha. Aqui, objetiva-se uma extrapolação para trás. Para isso, basta aplicar reversões temporais nos segmentos envolvidos.
4. Emende as duas seqüências extrapoladas através do esquema de janelamento[1] descrito em [11].

5 RESULTADOS EXPERIMENTAIS

O propósito desta seção é comparar o desempenho do método de interpolação proposto sob diversas configurações dos parâmetros de processamento. Primeiramente, ilustra-se o efeito de desvanecimento da energia do sinal interpolado e o papel da escolha da ordem do modelo AR nessa questão. Em seguida, demonstram-se os benefícios da utilização do modelo modificado dentro do método de interpolação.

[1]O parâmetro de *roll-off* foi ajustado em 3 em todas as simulações aqui apresentadas.

5.1 Estudo de Caso: tom grave de piano

Este estudo de caso se refere a um tom grave de piano com freqüência fundamental $f_0 \approx 50$ Hz, tocado *forte*. Sinais desse tipo apresentam centenas de parciais e, portanto, suas modelagens por processos AR requerem o emprego de ordens altas. A adoção de ordens insuficientes não só negligencia alguns modos de ressonância, mas também tende a alargar a banda das ressonâncias modeladas. Em outras palavras, os pólos associados a essas ressonâncias não estão tão próximos da circunferência unitária quanto deveriam estar.

Como mencionado na Seção 1, a localização dos pólos do modelo AR tem papel crucial no processo de interpolação de falhas longas. Para exemplificar o efeito da escolha da ordem do modelo, uma falha de 2000 amostras (@ 44,1 kHz de freqüência de amostragem) é artificialmente criada no sinal de teste, começando aproximadamente 5000 amostras depois do ataque da nota. A comparação entre os resultados de interpolação para vários valores de ordem de modelo é mostrada na Fig. 3. Dos resultados para esse sinal de teste, observa-se que ordens mais baixas que 500 são insuficientes para reconstruir o sinal adequadamente. Entretanto, a adoção de $p = 1000$ aparentemente é suficiente para esse exemplo.

Agora, os modelos convencionais são substituídos por seus respectivos modelos modificados. A ordem[2] dos modelos é fixada em $p = 150$ e $\eta \in \{0{,}5, 0{,}8, 0{,}99\}$. Os resultados do processo de interpolação são mostrados na Fig. 4. Nota-se que a adoção de η próximo de 1 parece melhorar o desempenho da intepolação, no que diz respeito à preservação da energia do sinal reconstruído ao longo da falha.

[2]O valor de p se refere à ordem do modelo convencional. A ordem do modelo modificado é, em geral, $p + 1$.

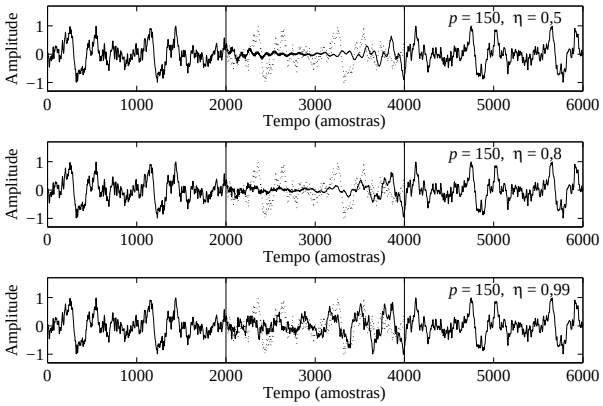


Figura 4: Desempenho do método de interpolação proposto para ordem de modelo $p = 150$ e vários valores de η. A falha é localizada entre as amostras 2001 e 4000. O sinal original é mostrado em linha pontilhada.

Os resultados anteriores contam a favor da utilização dos modelos modificados no método de interpolação investigado, uma vez que, com as mesmas ordens de modelo, o uso dos modelos modificados produz melhores resultados, pelo menos visualmente. Ademais, o custo computacional extra necessário ao cálculo dos modelos modificados é mínimo, se comparado àqueles relacionados com a estimação dos modelos convencionais e o procedimento de extrapolação.

O efeito colateral do uso dos modelos modificados reside no desvio observado nas suas freqüências de ressonância, em relação às freqüências do modelo original. Apesar disso, freqüências de ressonância ligeiramente desviadas serão provavelmente excitadas com um nível de energia similar ao das freqüências originais. Essa é a razão pela qual o método de interpolação ainda funciona. Freqüências de ressonâncias fortemente polarizadas serão provavelmente excitadas a baixos níveis de energia. Logo a contribuição destas na forma de distorções audíveis no sinal reconstruído pode ser menos crítica do que aquela causada por freqüências de ressonância ligeiramente desviadas.

A Fig. 5 confronta os espectros do sinal original na região da falha com os espectros dos sinais reconstruídos, dadas as configurações de processamento indicadas. Nota-se que o espectro do sinal original tem um número substancial de ressonâncias.

O espectro do sinal interpolado com $p = 150$ e $\eta = 0{,}5$ revela que somente as ressonâncias mais proeminentes foram modeladas. Além disso, suas bandas parecem mais largas do que as originais.

Por outro lado, a configuração $p = 150$ e $\eta = 0{,}99$ produz um sinal com um número maior de ressonâncias proeminentes. Tais ressonâncias têm bandas mais estreitas do que as obtidas com a configuração $p = 150$ e $\eta = 0{,}5$. Entretanto, é evidente que a estrutura intrincada do espectro do sinal original não é bem modelada através de modelos AR de ordem baixa.

Em relação às freqüências de ressonância, aquelas do sinal interpolado usando-se a configuração $p = 150$ e $\eta = 0{,}99$ mostram visualmente uma correlação relativamente alta com as ressonâncias do sinal original. A Fig. 6 ilustra melhor esse fato. Apesar disso, é possível observar desvios em algumas freqüências de ressonância. Por exemplo, a ressonância proeminente que ocorre no sinal original em torno de 1500 Hz aparece um pouco acima no sinal interpolado. Há que se considerar, também, que os desvios nas freqüências de ressonância são mais acentuados em altas freqüências. Logo, é plausível especular que a qualidade dos sinais interpolados possa ser aceitável perceptualmente, apesar de possíveis distorções espúrias.

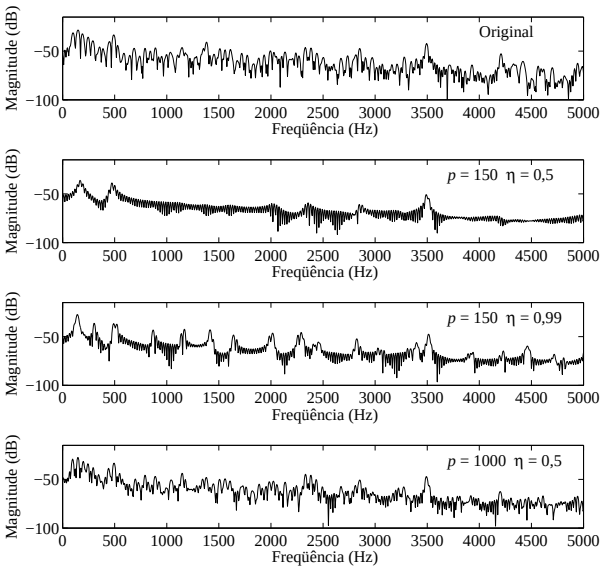


Figura 5: Detalhe (até 5 kHz) do espectro do sinal original e de sinais reconstruídos.

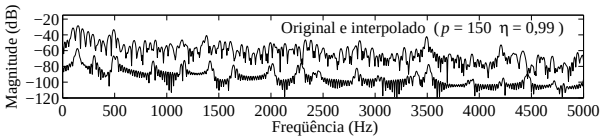


Figura 6: Detalhe (até 5 kHz) dos espectros do sinal original e do sinal reconstruído usando-se a configuração $\eta = 0{,}99$ e $p = 150$. O espectro do sinal interpolado é mostrado com uma compensação de -30 dB para melhor visualização.

6 DESEMPENHO QUALITATIVO

Essa seção é dedicada à avalição subjetiva do método de interpolação usando modelos modificados. Os sinais de teste utilizados, assim como as configurações de processamento empregadas, são descritos a seguir.

6.1 Sinais de teste

Nesse trabalho, por questões de praticidade, decidiu-se avaliar o desempenho do método de interpolação em quatro sinais de teste. Segue-se uma breve descrição desses sinais.

1. **Piano**: um tom de piano ($f_0 \approx 50Hz$) tocado *forte*;
2. **Pop**: um excerto de 14 s de música pop finlandesa;
3. **Clássico**: um fragmento de 13 s de música orquestral;
4. **Canto**: um segmento de 20 s de canto *a capella*.

6.2 Configuração de processamento

Os experimentos consistem em criar falhas periódicas nos sinais de teste e aplicar o procedimento de interpolação para cada uma delas, usando, para cada sinal de teste, diversas configurações de parâmetros de processamento. Os sinais reconstruídos são então submetidos à apreciação de ouvintes que julgaram a qualidade dos resultados.

O tamanho das falhas é fixado em 2000 amostras (@ 44,1 kHz), o que corresponde a uma duração de aproximadamente 45 ms. Essa escolha exige sobremaneira do procedimento de interpolação, uma vez que a duração das falhas está no limite máximo acima do qual a estacionariedade não pode mais ser assumida em sinais de áudio. O avanço entre uma falha e a ocorrência subseqüente é de 50000 amostras. Isso garante não só a existência de amostras suficientes antes e depois da falha para estimar os modelos AR, mas também que o interpolador será avaliado sob diferentes características espectro-temporais dos sinais de teste.

A idéia aqui é manter baixa a ordem dos modelos AR de modo que o resultado da interpolação seja insatisfatório quando usando modelos convencionais. Mantendo-se baixa a ordem, objetiva-se, então, verificar se ocorrem melhoras de desempenho devido ao uso dos modelos modificados. Nesse contexto, parece razoável adotar arbitrariamente $p = 150$ para a ordem dos modelos modificados. A investigação, então, fica focalizada no efeito do parâmetro η nos resultados. Os valores adotados para η na investigação são $\eta \in \{0{,}5,\ 0{,}8,\ 0{,}9,\ 0{,}99,\ 1\}$.

A título de referências para comparações subjetivas, versões restauradas dos sinais de teste corrompidos foram obtidas usando-se também modelos convencionais com ordens $p \in \{150, 300, 500, 1000\}$. Os parâmetros de processamento usados nas simulações são resumidos na Tabela 1.

Tabela 1: Parâmetros usados nas simulações.

Parâmetro	Valor
Extensão da falha	$G = 2000$ amostras
Periodicidade das falhas	50000 amostras
Ordem (AR convencional)	$p \in \{150, 300, 500, 1000\}$
Estimação AR	Burg, $N = 2000$ amostras
Ordem (AR modificado)	$p = 150$
Pesos nos PPLEs	$\eta \in [0{,}5,\ 1{,}0]$

6.3 Avaliação Subjetiva

A avaliação subjetiva foi realizada através de testes informais de audição. Quanto aos resultados relativos aos sinais de referência (usando modelos AR convencionais), observou-se que, como já se poderia antecipar, a qualidade da restauração aumenta com a ordem dos modelos AR, dentro da faixa investigada. Para todos os sinais de teste, a restauração usando modelos convencionais com ordem $p = 150$ já produz certa melhora na qualidade perceptual, haja vista que as transições nas bordas das falhas são suavizadas. Entretanto, ainda é clara a presença das falhas.

Quanto a avaliação dos resultados obtidos através da utilização dos modelos modificados, a impressão geral é de que a percepção de melhorias qualitativas ocorre mais claramente para valores de η bem próximos de 1. Para todos os sinais de teste utilizados, e para $p = 150$, os resultados de interpolação obtidos com $\eta = 1$ foram julgados como de qualidade superior aos correspondentes obtidos com $\eta = 0{,}5$, i.e., via modelos convencionais. Isso a despeito da ocorrência de algumas distorções nos sinais restaurados usando modelos modificados. Essas distorções aparecem como uma qualidade mais sintética e mecânica em certas falhas preenchidas.

O grau de melhora perceptiva varia de acordo com o sinal de teste. Por exemplo, para o sinal **Piano**, o procedimento de interpolação usando $p = 150$ e $\eta = 1$ produz um resultado que soa perceptivamente similar ao obtido com $p = 500$ e $\eta = 0{,}5$. Conclusões semelhantes foram tiradas com relação ao sinal **Clássico**.

Já para os sinais **Pop** e **Canto**, as restaurações obtidas usando $p = 150$ e $\eta = 1$ soam similares àquelas produzidas com $p = 1000$ e $\eta = 0{,}5$. Vale notar, entretanto, que os sinais **Pop** e **Canto** exigem mais do procedimento de interpolação, por serem menos estacionários do que os sinais **Piano** e **Clássico**. Assim, para aqueles, mesmo usando ordens altas para os modelos AR, como $p = 1000$, os resultados de restauração soam, em geral, inferiores qualitativamente a estes, quando comparados aos originais.

Pelos resultados supracitados, pode-se concluir que a utilização de modelos AR modificados ou, mais precisamente, a substituição de modelos AR convenci-

onais por seus correspondentes polinômios simétricos de linhas espectrais, dentro do procedimento de interpolação investigado, oferece um melhor balanço entre o custo computacional do método de interpolação e seu desempenho qualitativo. Exemplos dos sinais processados podem ser encontrados no URL: `http://www.acoustics.hut.fi/~esquef/aesbrasil/`.

7 CONCLUSÕES

Esse artigo apresentou uma aplicação dos polinômios do par de linhas espectrais associado a um modelo autorregressivo ao problema de reconstrução de sinais de áudio em falhas de longa duração. A modificação proposta no método de interpolação investigado consistiu na substituição de modelos AR estimados convencionalmente por uma soma ponderada de seus polinômios do par de linhas espectrais associados. A localização dos pólos desses modelos modificados pode ser controlada através de um único parâmetro de peso.

Foram realizados experimentos em um conjunto de quatro sinais de teste, degradados artificialmente e reconstruídos através do método proposto, sob diversas configurações dos parâmetros de processamento. A avaliação dos resultados obtidos, tanto visual como auditiva, revelou que adotar o parâmetro de peso igual a 1 favorece o desempenho qualitativo do método de interpolação. Essa escolha equivale à utilização do polinômio simétrico de linhas espectrais, cujos pólos se encontram exatamente na circunferência unitária.

Avaliações subjetivas através de testes de audição mostraram que, apesar de introduzir algumas distorções audíveis no sinais reconstruídos, a modificação proposta implica sinais restaurados com qualidade subjetiva superior aos obtidos através do método original, com mesma ordem (baixa) de modelo. Como a modificação proposta requer um custo computacional desprezível, é possível argumentar que a mesma melhora o método de interpolação, em termos de um melhor balanço entre complexidade computacional e desempenho qualitativo.

REFERÊNCIAS

[1] N. S. Jayant and S. Christensen, "Effects of Packet Losses in Waveform Coded Speech and Improvements Due to an Odd-Even Sample-Interpolation Procedure," *IEEE Trans. Communications*, vol. CAM-29, pp. 101–109, Feb. 1981.

[2] P. Wilson, "Record Contamination: Causes and Cure," *J. Audio Eng. Soc.*, vol. 13, pp. 166–176, Apr. 1965.

[3] J. Makhoul, "Linear Prediction: A Tutorial Review," *Proc. IEEE*, vol. 63, no. 4, pp. 561–580, 1975.

[4] S. V. Vaseghi and P. J. W. Rayner, "Detection and Suppression of Impulsive Noise in Speech Communication Systems," *IEE Proceedings*, vol. 137, pp. 38–46, Feb. 1990.

[5] A. J. E. M. Janssen, R. N. J. Veldhuis, and L. B. Vries, "Adaptive Interpolation of Discrete-Time Signals That Can Be Modeled as Autoregressive Processes," *IEEE Trans. Signal Processing*, vol. ASSP-34, pp. 317–330, Apr. 1986.

[6] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration — A Statistical Model Based Approach*, ch. 5. London, UK: Springer-Verlag, 1998.

[7] P. J. W. Rayner and S. J. Godsill, "The Detection and Correction of Artefacts in Degraded Gramophone Recordings," in *Proc. IEEE ASSP Workshop Applications Signal Processing Audio Acoustics*, pp. 151–152, Oct. 1991.

[8] M. Niedźwiecki, "Statistical Reconstruction of Multivariate Time Series," *IEEE Trans. Signal Processing*, vol. 41, pp. 451–457, Jan. 1993.

[9] W. Etter, "Restoration of a Discrete-Time Signal Segment by Interpolation Based on the Left-Sided and Right-Sided Autoregressive Parameters," *IEEE Trans. Signal Processing*, vol. 44, pp. 1124–1135, May 1996.

[10] I. Kauppinen, J. Kauppinen, and P. Saarinen, "A Method for Long Extrapolation of Audio Signals," *J. Audio Eng. Soc.*, vol. 49, pp. 1167–1180, Dec. 2001.

[11] I. Kauppinen and J. Kauppinen, "Reconstruction Method for Missing or Damaged Long Portions in Audio Signal," *J. Audio Eng. Soc.*, vol. 50, pp. 594–602, July/Aug. 2002.

[12] I. Kauppinen and K. Roth, "Audio Signal Extrapolation – Theory and Applications," in *Proc. 5th Int. Conf. on Digital Audio Effects*, (Hamburg, Germany), pp. 105–110, Sept. 2002. URL: http://www.unibw-hamburg.de/EWEB/ANT/dafx2002/papers.html.

[13] P. Alku and T. Bäckström, "All-pole Modeling Technique based on the Weighted Sum of the LSP Polynomials," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2002)*, vol. 1, (Orlando, Florida, USA), pp. 665–668, May 2002.

[14] F. K. Soong and B. W. Juang, "Line Spectrum Pair (LSP) and Speech Data Compression," in *Proc. Int. Conf. Acoust., Speech, Audio Signal Procesing*, (San Diego, CA), pp. 1.10.1–1.10.4, 1984.

[15] M. H. Hayes, *Statistical Signal Processing and Modeling*, ch. 6. John Wiley & Sons, Inc., 1996.

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na VIII Convenção Nacional
7 - 9 de Junho de 2004, São Paulo, SP, Brasil

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.

Interpolação Bilinear Generalizada de HRTFs para Geração de Som Tridimensional

Fábio P. Freeland[1], Luiz W. P. Biscainho[1], Paulo S. R. Diniz[1]
[1]LPS – PEE/COPPE & DEL/Poli, UFRJ
freeland@lps.ufrj.br

RESUMO

A geração de som tridimensional em fones de ouvido pode ser feita por interpolação de Funções de Transferência Relativas à Cabeça (HRTFs). Nesse trabalho, explicita-se a forma de cálculo dos ponderadores usados nessa interpolação por analogia com um método multicanal conhecido como Variação Gradual da Amplitude Baseada em Vetores (VBAP). Por fim, algumas simplificações são realizadas, obtendo-se expressões práticas para casos particulares usuais.

INTRODUÇÃO

Já faz algum tempo que se começou a pensar em uma forma de se recriar de maneira controlada a sensação de imersão em um ambiente virtual. Já na época da criação do estéreo o que se pretendia era tornar o som mais envolvente. O cinema também foi um grande impulsionador do aparecimento de novas técnicas de atribuição de ambiência ao som, já que uma de suas características mais importantes para o público é a diferença entre o ambiente da sala de cinema e o doméstico; e isso se liga não só à diferença no tamanho da tela, mas também à imersão sonora no ambiente do filme.

As técnicas de geração de som tridimensional evoluíram desde o estéreo até o som *surround* (que domesticamente já conta com sete canais mais o canal de graves) passando pelo som quadrafônico. A grande revolução ocorrida nessa área se deu com a possibilidade de se fazer o processamento digital dos sinais, viabilizado pela velocidade elevada dos processadores atuais. E com isso, surgiu um novo conceito: a qualidade controlada; o grau de fidelidade desejado na representação digital para um sinal analógico é determinado pelo número de bits nela utilizados. Portanto, a qualidade fica dependente apenas do quanto se está disposto a gastar na representação do sinal antes, durante e depois do processamento.

Uma das aplicações que mais se beneficiaram desse salto tecnológico foi a recriação de ambiente virtual por simuladores e até jogos eletrônicos, onde se quer o maior grau possível de realismo. As mesmas técnicas de processamento digital neles empregadas podem servir para gerar efeitos especiais nas trilhas sonoras de filmes.

A criação de um ambiente sonoro virtual tenta produzir o som que se ouviria no ambiente em questão, isto é, dar noção de tridimensionalidade ao som. Isso corresponde a simular as alterações sofridas pelo som desde sua emissão pela fonte até chegar a cada um dos ouvidos. Deve-se notar que a mera utilização de uma configuração com diversos alto-falantes ao redor do ouvinte, embora produza uma sensação de ambientação, não é suficiente

para localizá-lo na cena sonora; para tal, é necessário processar convenientemente o sinal.

Existem vários métodos diferentes para se gerar o som tridimensional [1, 2]. Os métodos diferenciam-se não só pelo modo como são aplicados, mas também pelo modo de apresentação do som ao ouvinte. Alguns deles utilizam caixas acústicas, geralmente com mais de dois canais independentes, cada um relacionado com uma direção, sendo por isso chamados de multicanal. Outros, chamados binaurais [3], fornecem o som de forma individual através de fones de ouvido.

A produção binaural do som 3D se baseia, em geral, em funções de transferência chamadas Funções de Transferência Relativas à Cabeça (HRTFs—*Head-Related Transfer Functions*), previamente medidas [4]. As HRTFs descrevem as mudanças sofridas pelo som entre a fonte emissora e cada ouvido. A filtragem de um som monaural por um par dessas funções recria uma fonte virtual na posição para a qual elas foram medidas. Claro que para se posicionar a fonte virtual em qualquer ponto, é necessário interpolar HRTFs não medidas.

Já os métodos multicanal, a princípio, não empregam esse tipo de função de transferência no processamento. Eles se valem do posicionamento de várias caixas acústicas em torno do ouvinte e, com uma combinação do som emitido por um subconjunto delas, dão a impressão de posicionamento da fonte virtual.

Neste artigo, parte-se de um método multicanal conhecido como Posicionamento por Variação da Amplitude Baseada em Vetores (VBAP—*Vector-Based Amplitude Panning*) [5] para se obter uma formulação para a interpolação de HRTFs utilizadas na geração binaural de som 3D. Deve-se ressaltar que o objetivo deste trabalho não é a comparação entre métodos de interpolação, mas sim o detalhamento da formulação de um método particular.

Inicialmente, discute-se o princípio do posicionamento de uma fonte virtual por variação de amplitude. A seguir, mostra-se que a interpolação de HRTFs pode ser descrita da mesma forma. Na seqüência, revisa-se o método multicanal VBAP. Então, derivam-se as expressões que realizam a interpolação de HRTFs com base no VBAP. Por fim, apresentam-se as conclusões.

POSICIONAMENTO POR VARIAÇÃO DE AMPLITUDE, COM TRÊS CANAIS

Todos os métodos de geração de som tridimensional têm como objetivo básico reproduzir, física ou virtualmente, as funções de transferência $H_{PE}(\omega)$ e $H_{PD}(\omega)$ que descrevem os caminhos desde uma posição P até os ouvidos esquerdo e direito, respectivamente. Assim, tenta-se reproduzir o som de uma fonte virtual localizada em P filtrando-se o som $S_P(\omega)$ por ela emitido de forma que no domínio da freqüência se tenham

$$\begin{cases} S_E(\omega) = S_P(\omega) H_{PE}(\omega) \\ S_D(\omega) = S_P(\omega) H_{PD}(\omega). \end{cases} \tag{1}$$

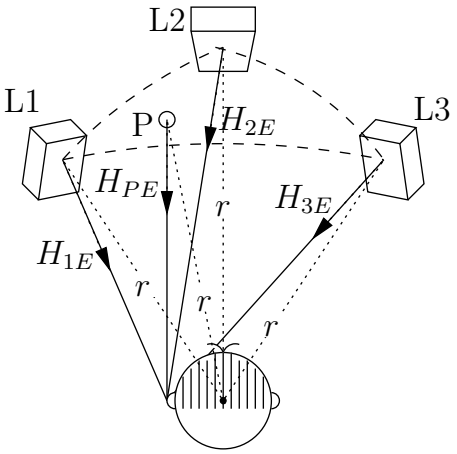


Figura 1: Posicionamento baseado na variação de amplitude com 3 caixas acústicas–ouvido esquerdo.

Considere-se o caso de três canais apresentados através de caixas acústicas direcionadas a um ouvinte situado à distância r, como mostra a Fig. 1. O sinal recebido por cada um dos ouvidos é uma combinação dos sinais emitidos por cada uma das caixas, podendo ser descrito como

$$\begin{cases} \hat{S}_E(\omega) = \sum_{i=1}^{3} S_i(\omega) H_{iE}(\omega) \\ \hat{S}_D(\omega) = \sum_{i=1}^{3} S_i(\omega) H_{iD}(\omega), \end{cases} \tag{2}$$

onde $S_i(\omega)$ é o espectro do sinal emitido pela caixa Li, e $H_{iE}(\omega)$ e $H_{iD}(\omega)$ são as funções de transferência da caixa Li aos ouvidos esquerdo e direito, respectivamente. Potencialmente, esta configuração posiciona uma fonte virtual em qualquer ponto P sobre o triângulo esférico determinado pelas três caixas. Um sistema de som 3D eficaz deve garantir que

$$\begin{cases} \hat{S}_E(\omega) \approx S_E(\omega) \\ \hat{S}_D(\omega) \approx S_D(\omega). \end{cases} \tag{3}$$

Uma solução intuitiva consiste em aplicar o sinal que se deseja ouvir vindo do ponto P a cada uma das caixas L_i ponderado por uma constante g_i [5], ou seja, $S_i(\omega) = g_i S_P(\omega)$. Nesse caso, a aproximação (3) se resume a fazer

$$\begin{cases} H_{PE}(\omega) \approx \sum_{i=1}^{3} g_i H_{iE}(\omega) \\ H_{PD}(\omega) \approx \sum_{i=1}^{3} g_i H_{iD}(\omega). \end{cases} \tag{4}$$

INTERPOLAÇÃO BASEADA EM TRÊS HRTFS

Um sistema típico de geração binaural de som tridimensional se baseia em um conjunto de HRTFs (*Head-Related Transfer Function*) medidas para um número

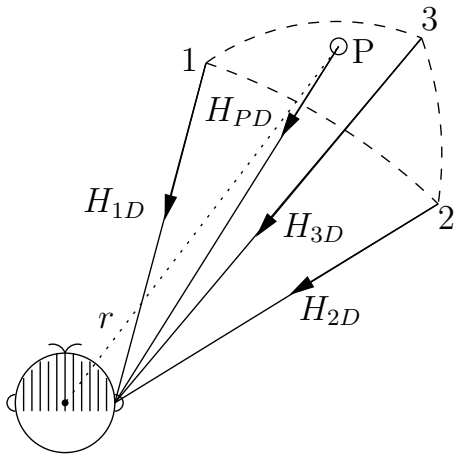


Figura 2: Interpolação de HRTF a partir de 3 funções conhecidas–ouvido direito.

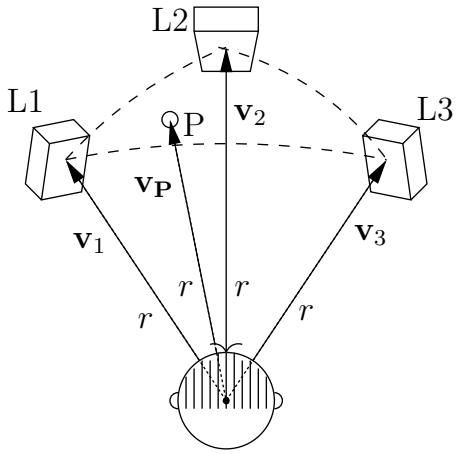


Figura 3: VBAP com três canais.

finito de posições sobre uma esfera de raio r com o ouvinte no centro. Para se posicionar uma fonte virtual em qualquer ponto P sobre a esfera, uma solução simples é atribuir à HRTF para o ponto P uma combinação linear das HRTFs medidas para os pontos mais próximos de P, isso para cada ouvido.

A Fig. 2 mostra um esquema de interpolação baseado em três HRTFs conhecidas.

Nesse caso, a aproximação realizada é regida pelas mesmas Eqs. (4), onde as funções $H_{PD}(\omega)$ e $H_{PE}(\omega)$ são estimadas diretamente pela combinação das funções $H_{iD}(\omega)$ e $H_{iE}(\omega)$, com $i = 1, 2, 3$, respectivamente, para os ouvidos direito e esquerdo. Por esse motivo, tendo as funções de transferência a mesma significação física nos casos multicanal e binaural, espera-se que uma solução para os pesos g_i adequada ao caso multicanal também seja apropriada para o caso binaural.

MÉTODO VBAP

A técnica conhecida como VBAP (*Vector-Based Amplitude Panning*) [5] é um método multicanal para posicionamento de uma fonte virtual que se baseia na descrição vetorial das posições dos alto-falantes e da fonte virtual em relação ao ouvinte. No caso de três canais, ilustrado na Fig. 3, os coeficientes g_i que satisfazem a aproximação da Eq. (4) são calculados através de

$$\mathbf{v_P} = \sum_{i=1}^{3} g_i \mathbf{v}_i. \tag{5}$$

Essa equação descreve o vetor associado à posição P como uma combinação linear dos vetores $\mathbf{v}_i$ associados às posições das caixas L_i com $i = 1, 2, 3$ em coordenadas cartesianas. Admite-se, assim, que a mesma combinação é válida para as funções de transferência correspondentes. Explicitamente,

$$\begin{bmatrix} x_P \\ y_P \\ z_P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} \tag{6}$$

onde (x_i,y_i,z_i) são as coordenadas relacionadas com a posição da caixa L_i.

INTERPOLAÇÃO DE HRTFS BASEADA NO VBAP

Como foi mencionado anteriormente, a interpolação de HRTFs pode utilizar os mesmos ponderadores calculados para um método multicanal com três caixas acústicas. Aqui, propõe-se usar como base a formulação do VBAP.

Considerando a geometria definida na Fig. 2, é vantajoso escrever os vetores (x,y,z) em função das coordenadas esféricas:

$$\begin{cases} x = r\cos\phi\cos\theta \\ y = r\cos\phi\, sen\,\theta \\ z = r\, sen\,\phi, \end{cases} \tag{7}$$

onde ϕ é o ângulo de elevação relativo ao plano horizontal que passa pelos ouvidos, θ é o ângulo de azimute relativo ao plano vertical que corta perpendicularmente a linha que une os ouvidos e r é o raio da esfera.

Escolhendo como referência o ponto 1, conforme a Fig. 4, as coordenadas angulares dos outros pontos ficam iguais a

$$(\theta_2, \phi_2) = (\theta_1 + \Delta\theta_2, \phi_1 + \Delta\phi_2) \tag{8}$$

$$(\theta_3, \phi_3) = (\theta_1 + \Delta\theta_3, \phi_1 + \Delta\phi_3) \tag{9}$$

$$(\theta_P, \phi_P) = (\theta_1 + \delta\theta, \phi_1 + \delta\phi). \tag{10}$$

Substituindo as Eqs. (8)–(10) na Eq.(7), obtêm-se as coordenadas dos pontos

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} r\cos\phi\cos\theta_1 \\ r\cos\phi\, sen\,\theta_1 \\ r\, sen\,\phi_1 \end{bmatrix} \tag{11}$$

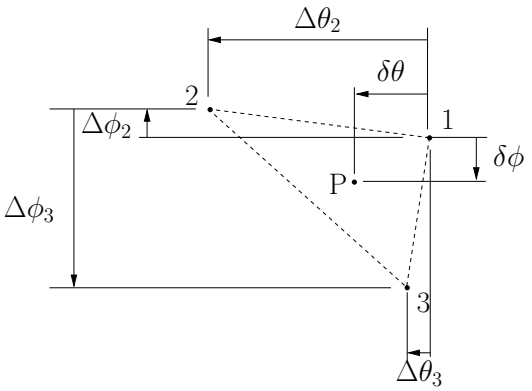


Figura 4: Geometria para interpolação triangular de HRTFs.

$$\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} r\cos(\phi_1+\Delta\phi_2)\cos(\theta_1+\Delta\theta_2) \\ r\cos(\phi_1+\Delta\phi_2)\,sen(\theta_1+\Delta\theta_2) \\ r\,sen(\phi_1+\Delta\phi_2) \end{bmatrix} \tag{12}$$

$$\begin{bmatrix} x_3 \\ y_3 \\ z_3 \end{bmatrix} = \begin{bmatrix} r\cos(\phi_1+\Delta\phi_3)\cos(\theta_1+\Delta\theta_3) \\ r\cos(\phi_1+\Delta\phi_3)\,sen(\theta_1+\Delta\theta_3) \\ r\,sen(\phi_1+\Delta\phi_3) \end{bmatrix} \tag{13}$$

$$\begin{bmatrix} x_P \\ y_P \\ z_P \end{bmatrix} = \begin{bmatrix} r\cos(\phi_1+\delta\phi)\cos(\theta_1+\delta\theta) \\ r\cos(\phi_1+\delta\phi)\,sen(\theta_1+\delta\theta) \\ r\,sen(\phi_1+\delta\phi) \end{bmatrix}. \tag{14}$$

Agora, considerando que as HRTFs medidas formam um conjunto denso (i.e., foram medidas para pontos bastante próximos[1]), pode-se considerar que $\Delta\theta_2$, $\Delta\theta_3$, $\Delta\phi_2$, $\Delta\phi_3$, $\delta\theta$ e $\delta\phi$ são pequenos. Sendo $sen(\alpha) \approx \alpha$ e $\cos(\alpha) \approx 1$ para $\alpha \approx 0$, podem-se obter as aproximações para as coordenadas

$$\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} \approx \begin{bmatrix} r(\cos\phi_1 - \Delta\phi_2\,sen\,\phi_1)(\cos\theta_1 - \Delta\theta_2\,sen\,\theta_1) \\ r(\cos\phi_1 - \Delta\phi_2\,sen\,\phi_1)(sen\,\theta_1 + \Delta\theta_2\cos\theta_1) \\ r(sen\,\phi_1 + \Delta\phi_2\cos\phi_1) \end{bmatrix} \tag{15}$$

$$\begin{bmatrix} x_3 \\ y_3 \\ z_3 \end{bmatrix} \approx \begin{bmatrix} r(\cos\phi_1 - \Delta\phi_3\,sen\,\phi_1)(\cos\theta_1 - \Delta\theta_3\,sen\,\theta_1) \\ r(\cos\phi_1 - \Delta\phi_3\,sen\,\phi_1)(sen\,\theta_1 + \Delta\theta_3\cos\theta_1) \\ r(sen\,\phi_1 + \Delta\phi_3\cos\phi_1) \end{bmatrix} \tag{16}$$

$$\begin{bmatrix} x_P \\ y_P \\ z_P \end{bmatrix} \approx \begin{bmatrix} r(\cos\phi_1 - \delta\phi\,sen\,\phi_1)(\cos\theta_1 - \delta\theta\,sen\,\theta_1) \\ r(\cos\phi_1 - \delta\phi\,sen\,\phi_1)(sen\,\theta_1 + \delta\theta\cos\theta_1) \\ r(sen\,\phi_1 + \delta\phi\cos\phi_1) \end{bmatrix}. \tag{17}$$

Por fim, substituindo-se as Eqs. (11) e (15)–(17) na Eq. (6), podem-se calcular os ponderadores g_i em função das distâncias angulares:

$$\begin{cases} g_3 = \dfrac{\delta\phi\Delta\theta_2 - \delta\theta\Delta\phi_2}{\Delta\theta_2\Delta\phi_3 - \Delta\theta_3\Delta\phi_2} \\ g_2 = \dfrac{\delta\theta\Delta\phi_3 - \delta\phi\Delta\theta_3}{\Delta\theta_2\Delta\phi_3 - \Delta\theta_3\Delta\phi_2} \\ g_1 = 1 - g_2 - g_3. \end{cases} \tag{18}$$

[1] Em [4], por exemplo, tem-se algo em torno de 700 posições distribuídas de forma aproximadamente uniforme sobre a esfera de referência.

Nesse caso, o número de operações necessárias para se obter os coeficientes é de 11 somas/subtrações e 8 multiplicações/divisões por nova posição.

Deve ser notado que no caso geral, as coordenadas angulares dos pontos para os quais as HRTFs foram medidas são quaisquer. Entretanto, na prática, as medidas são realizadas com pelo menos o passo de elevação fixo. Nesse caso particular, pode-se fazer $\Delta\phi_2 = 0$ nas Eqs. (18), o que as torna

$$\begin{cases} g_3 = \dfrac{\delta\phi}{\Delta\phi_3} \\ g_2 = \dfrac{\delta\theta}{\Delta\theta_2} - g_3\dfrac{\Delta\theta_3}{\Delta\theta_2} \\ g_1 = 1 - g_2 - g_3. \end{cases} \tag{19}$$

Uma simplificação adicional ainda pode ser feita no caso de as HRTFs serem medidas com ambos os passos, de azimute e de elevação, constantes, o que permite fazer $\Delta\theta_3 = 0$ nas Eqs. (19), o que resulta em

$$\begin{cases} g_3 = \dfrac{\delta\phi}{\Delta\phi_3} \\ g_2 = \dfrac{\delta\theta}{\Delta\theta_2} \\ g_1 = 1 - g_2 - g_3. \end{cases} \tag{20}$$

Assim, quando a geometria da grade de pontos o permite, a complexidade do cálculo dos ponderadores g_i é consideravelmente reduzida.

CONCLUSÕES

Neste artigo, deduziu-se uma formulação para o cálculo dos ponderadores usados para geração de som tridimensional utilizando-se HRTFs. Chegou-se às equações dos ponderadores através da analogia com um sistema multicanal baseado em três caixas acústicas, chamado de Posicionamento por Variação de Amplitude Baseado em Vetores (VBAP). Os resultados bem sucedidos do VBAP são conhecidos da literatura. Como a interpolação de HRTFs é simplesmente uma combinação linear de funções de transferência medidas para pontos sobre uma esfera, a formulação do VBAP pode ser reutilizada. Já que ambos os métodos atuam sobre o mesmo tipo de função, a expectativa é de que os resultados dos dois métodos sejam equivalentes. Além da formulação para o caso geral, citaram-se dois casos particulares de menor complexidade, aplicáveis na prática. Um dos casos particulares fora anteriormente aplicado, sem demonstração, pelos autores em [6] a um método de interpolação eficiente de HRTFs, com sucesso.

REFERÊNCIAS

[1] D. R. Begault, *3D Sound for Virtual Reality and Multimedia*. Cambridge, MA: Academic Press, 1994.

[2] J.-M. Jot, V. Larcher, and O. Warusfel, "Digital signal processing issues in the context of binaural and

transaural stereophony," in *Proc. 98th AES Convention*, (Paris, France), AES, Feb. 1995. (Preprint 3980).

[3] C. I. Cheng and G. H. Wakefield, "Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space," *J. Audio Eng. Soc.*, vol. 49, pp. 231–249, Apr. 2001.

[4] B. Gardner and K. Martin, "HRTF measurements of a KEMAR dummy-head microphone," Technical Report 280, MIT Media Lab., Cambridge, MA, May 1994.

[5] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *J. Audio Eng. Soc.*, vol. 45, pp. 456–466, Jun. 1997.

[6] F. P. Freeland, L. W. P. Biscainho, and P. S. R. Diniz, "Efficient HRTF interpolation in 3d moving sound," in *The Proceedings of The AES 22nd International Conference*, (Espoo, Finland), pp. 106–114, AES, June 2002.

Sociedade de Engenharia de Áudio
Artigo de Convenção

Apresentado na VIII Convenção Nacional
7-9 de Junho de 2004, São Paulo, SP, Brasil

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

Controle em Tempo Real da Temperatura da Bobina e do Deslocamento do Cone de Alto-falantes para Operação em Alta Potência

Constâncio Bortoni, Sidnei Noceti Filho, Rosalfonso Bortoni, Rui Seara
LINSE – Laboratório de Circuitos e Processamento de Sinais
Depto. de Engenharia Elétrica, Universidade Federal de Santa Catarina
88040-900, Florianópolis, Santa Catarina, Brasil
cbortoni@linse.ufsc.br

RESUMO

Este trabalho apresenta um procedimento de medida e controle digital de temperatura da bobina e deslocamento do cone de alto-falantes para operação em alta potência. O controle é realizado em tempo real, usando para tal um processador de sinal digital (DSP). As informações de temperatura da bobina e de deslocamento do cone são obtidas pela variação da resistência para corrente contínua (DC) da bobina e através da medida de aceleração do cone, respectivamente.

1. INTRODUÇÃO

Alto-falantes são transdutores que convertem sinais elétricos em ondas sonoras. Entretanto, a maior parte da energia elétrica de excitação não é transformada em energia acústica. A eficiência de conversão de energia dos alto-falantes é muito baixa, tipicamente de 1 a 5% para sistemas de radiação direta e de 10 a 40% para sistemas tipo corneta [1].

Para se obter altos níveis de pressão sonora em sistemas de radiação direta, deve-se aplicar altos níveis de potência elétrica, visto que cerca de 95 a 99% dessa potência é transformada em calor na bobina móvel do alto-falante [2], [3]. Assim, quando aplicamos um sinal elétrico em um alto-falante, mesmo que sejam respeitados todos os valores máximos especificados pelo fabricante, pode-se causar um dano irreversível devido ao excesso de temperatura em sua bobina. Na realidade, para uma operação segura sob o ponto de vista térmico, apenas a informação da potência máxima admitida não é suficiente, pois a temperatura depende também da capacidade de dissipação de calor do sistema. O supequecimento da bobina é uma das principais razões de danos e falhas em alto-falantes.

Uma outra causa de danos irreversíveis em alto-falantes é o deslocamento excessivo do cone, alterando as características de funcionamento ou até mesmo levando à completa paralisação de funcionamento do alto-falante.

Em [4]-[7] são apresentados alguns procedimentos de medida de temperatura da bobina em alto-falantes. Tais abordagens não consideram a medida de deslocamento do cone do alto-falante. Além do mais, as abordagens [4]-[6] não usam a informação da temperatura para o controle de potência no alto-falante. Em [7], o sistema de controle de temperatura é efetuado considerando apenas o valor de potência máximo, o que nem sempre é muito eficaz como anteriormente mencionado. Em [8], é apresentada uma abordagem que faz uso de um processador de sinal digital (DSP) para a medida de temperatura, porém não

é proposto um método de controle de tal temperatura. Em [9], [10] é apresentado um sistema de medição e controle da temperatura e do deslocamento do cone. Tal sistema é relativamente complexo e de alto custo comercial. Nessa abordagem, o sistema de controle estima os parâmetros livres do modelo eletroacústico estendido do alto-falante, considerando a escala de operação segura, protegendo o alto-falante de superaquecimento da bobina e de deslocamento excessivo do cone.

Este trabalho propõe um sistema de controle em tempo real no qual a potência fornecida ao alto-falante é limitada tanto pela temperatura da bobina quanto pelo deslocamento do cone. Dessa forma, é garantido o rendimento máximo do sistema de sonorização sem a necessidade de uso de grandes margens de segurança, que é um procedimento muito empregado atualmente. A principal diferença entre as abordagens discutidas em [9], [10] e a proposta, neste trabalho, reside principalmente na simplicidade de concepção e no custo final de implementação.

2. DESCRIÇÃO DO MÉTODO

A abordagem em questão pode ser divida em duas etapas, tais como:

i) determinação da temperatura da bobina e do deslocamento do cone do alto-falante;
ii) controle da temperatura da bobina e do deslocamento do cone do alto-falante.

Para determinar a temperatura da bobina, utilizou-se a relação direta de variação entre a resistência para corrente contínua (DC) da bobina e sua variação de temperatura [6], [8], [11]. Para obter o deslocamento do cone é utilizado um acelerômetro.

O modo para determinar a resistência DC da bobina é através da medida da tensão e da corrente nos terminais do alto-falante em condições normais de operação, sendo esse excitado apenas pelo sinal de áudio original. A Fig. 1 ilustra o esquema utilizado para as medidas de tensão e corrente nos terminais do alto-falante.

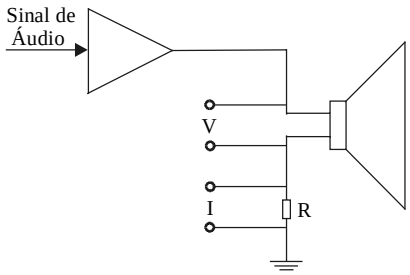


Fig. 1 – Esquema para determinar a tensão e a corrente nos terminais do alto-falante.

A resistência DC é calculada como sendo a parte real da impedância do alto-falante. Essa impedância é obtida através do quociente entre a tensão e a corrente medidas.

Uma vez determinada a resistência DC da bobina, pode-se então calcular a temperatura fazendo uso da relação entre a variação da resistência DC da bobina e a sua temperatura [11], através da expressão (1). Assim,

$$T(n)=T(n-1)+\left(\frac{R(n)}{R(n-1)}-1\right)\left(T(n-1)-25+\frac{1}{\alpha_{25}}\right) \quad (1)$$

onde $T(n)$ caracteriza a temperatura da bobina do alto-falante no instante atual; $T(n-1)$ é a temperatura no instante anterior ao atual; $R(n)$ é a resistência DC da bobina do alto-falante no instante atual; $R(n-1)$ é a resistência DC no instante anterior ao atual; α_{25} é o coeficiente de temperatura do material do fio da bobina a 25 ºC.

A Fig. 2 mostra a curva de variação da resistência DC da bobina em função da temperatura.

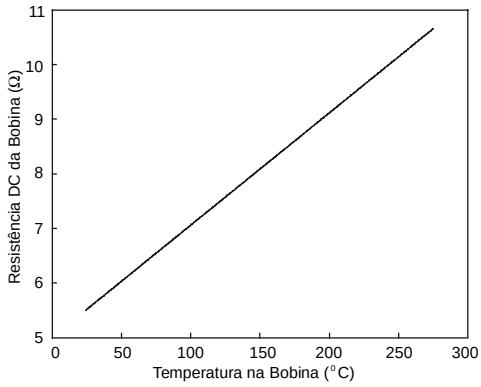


Fig. 2 – Curva da resistência DC da bobina *versus* temperatura para o alto-falante modelo 18SPW1P-p.

Para obter o deslocamento do cone do alto-falante nós utilizamos um acelerômetro. O acelerômetro deve ser colocado preferencialmente na calota protetora da bobina localizada no centro do alto-falante, pois é necessário, para uma correta medida, que o acelerômetro esteja perpendicular ao movimento do objeto para qual se quer obter o deslocamento. A superfície que estará em contato com o acelerômetro deve ser a mais plana possível.

O sinal obtido através do acelerômetro representa a aceleração do cone. Para obter o deslocamento do cone, é necessário que esse sinal seja integrado duas vezes.

O controle da temperatura da bobina e do deslocamento do cone do alto-falante é realizado digitalmente através de um DSP. Esse controle atua sobre a amplitude do sinal de áudio entregue ao alto-falante. A Fig. 3 mostra o diagrama em blocos do sistema proposto.

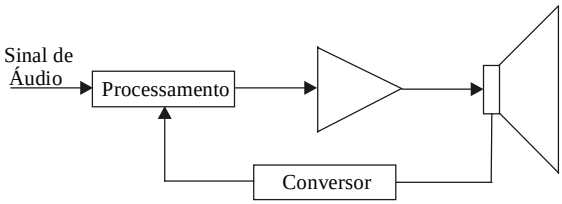


Fig. 3 – Diagrama em blocos do sistema de controle.

No bloco *Conversor* é feito a aquisição dos dados, que são: a aceleração do cone, e a tensão e corrente nos terminais do alto-falante. Após a aquisição, esses dados são digitalizados e enviados ao bloco *Processamento*, onde é efetuado o cálculo da temperatura da bobina e do deslocamento do cone. O bloco *Processamento* também realiza o controle digital.

2.1. Determinação da resistência DC da bobina do alto-falante

Devido à impossibilidade de se obter diretamente a resistência DC do alto-falante, por limitação do conversor A/D, é necessário se fazer uma aproximação deste parâmetro. Considera-se então o módulo da impedância da bobina, para uma freqüência abaixo de 10 Hz, como sendo igual à sua resistência DC.

A impedância da bobina do alto-falante, discutida em [6], [12], [13], é modelada por

$$Z = R_e + R_{ed} + j\omega L + Z_{EMA}\,, \tag{2}$$

onde

$$R_{ed} = K_{rm}\omega^{E_{rm}}\,, \tag{3}$$

$$L = K_{xm}\omega^{(E_{xm}-1)}\,, \tag{4}$$

e

$$Z_{EMA} = \frac{j\omega L_{ces}}{(1-\omega^2 L_{ces}c_{mes}) + \dfrac{j\omega L_{ces}}{r_{es}}}\,. \tag{5}$$

Em (2)-(5), R_e representa a resistência DC da bobina; R_{ed} caracteriza as perdas devido às correntes induzidas no entreferro do conjunto magnético, que são dependentes da freqüência; L é a indutância da bobina, dependente também da freqüência; Z_{EMA} representa a resultante das impedâncias refletidas do lado mecânico e acústico para o lado elétrico do alto-falante; K e E são coeficientes obtidos a partir da medição do módulo e do ângulo de fase da impedância em duas freqüências distintas. Tais coeficientes são geralmente fornecidos pelo fabricante do alto-falante. L_{ces} representa a compliância mecânica da suspensão do alto-falante; C_{mes} representa a massa móvel do diafragma, incluindo a massa de ar, nos dois lados do cone, variando com a freqüência; r_{es} denota a resistência da suspensão incluindo a resistência de radiação do ar nos dois lados do cone.

A Fig. 4 mostra a curva de módulo da impedância de um alto-falante (modelo 18SPW1P-p, fabricado pela Eletrônica Selenium S. A.), como também a respectiva resistência DC de sua bobina.

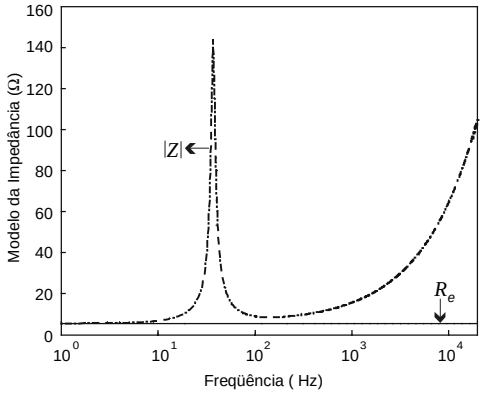


Fig. 4 – Curva da resistência DC da bobina para a temperatura de 24ºC e do módulo da impedância do alto-falante modelo 18SPW1P-p, desconsiderando sua impedância acústica.

2.2. Controle de temperatura da bobina e deslocamento do cone do alto-falante

O sistema de controle de temperatura da bobina e de deslocamento do cone do alto-falante não deve causar alterações audíveis às características originais do sinal reproduzido. O resultado tem que ser perceptualmente aceitável, não comprometendo a capacidade do alto-falante em reproduzir sinais de alta potência.

A Fig. 5 ilustra o esquema básico de controle de temperatura da bobina e deslocamento do cone do alto-falante. Nesta estrutura, G denota o ganho do sinal de áudio controlado pela temperatura da bobina, T; L representa um limitador de deslocamento do cone, x; AD representa o conversor analógico-digital; e DA o conversor digital-analógico.

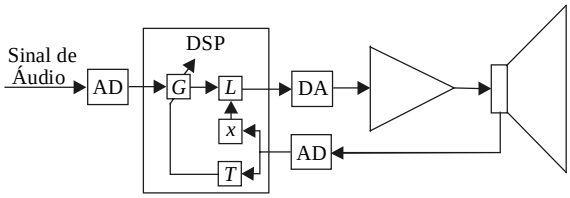


Fig. 5 – Esquema básico do controle de temperatura da bobina e deslocamento do cone do alto-falante.

O presente trabalho sugere o controle de temperatura da bobina levando em consideração as características do alto-falante. Tal controle é efetuado através de uma característica de ganho temporal que deve ser inversa à variação da temperatura da bobina relativa ao tempo.

O controle de temperatura é feito através do controle de ganho do sinal de áudio que vai ser amplificado e enviado ao alto-falante [14].

Um exemplo ilustrativo da curva de variação de temperatura da bobina com a atuação do controle de temperatura é mostrado na Fig. 6. A Fig. 7 ilustra a respectiva curva de ganho para o controle de temperatura da Fig. 6.

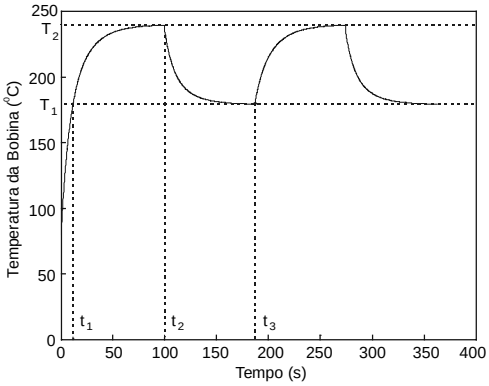


Fig. 6 - Exemplo de curva de variação da temperatura da bobina com a atuação do controle de temperatura.

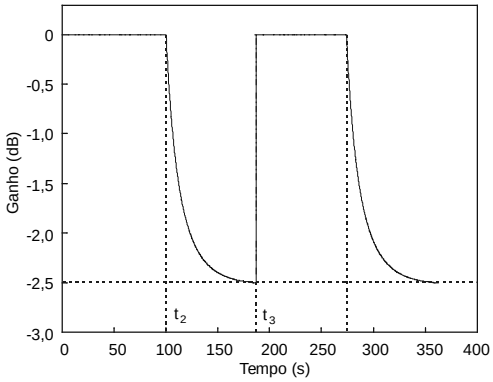


Fig. 7 - Curva de ganho para o controle da temperatura.

Observando a Fig. 6, verifica-se que a curva de variação de temperatura da bobina é dependente não só da potência aplicada como também do tempo. Assim, para uma dada potência constante aplicada ao alto-falante, mede-se no instante t_1 a temperatura da bobina, que é T_1. Mantida a potência, em um segundo instante t_2, a temperatura da bobina será T_2, sendo maior do que T_1. Considerando agora T_1 e T_2 como limiares inferior e superior, respectivamente, pode-se fazer atuar o controle para estes níveis de temperatura. Dessa forma, uma vez que seja acionado o controle de temperatura em T_2, uma redução de temperatura se inicia, decrescendo a temperatura para o nível T_1, no instante t_3. Cessada a atuação do controle no instante t_3, o processo de aumento da temperatura é reiniciado, e um novo ciclo tem origem, conforme mostrado na Fig. 6. As temperaturas T_1 e T_2 são pré-determinadas na fase de projeto do sistema de controle, sendo $T_1 < T_2$, e T_2 dependente da temperatura máxima da bobina. A Fig. 7 ilustra a variação de ganho (com conseqüente variação de potência), visando o controle de temperatura da bobina.

A partir do instante que a temperatura da bobina é maior do que a temperatura T_1, os valores de temperatura em instantes sucessivos de tempo, entre T_1 e T_2, são armazenados. Tais valores são interpolados, gerando uma curva de variação da temperatura entre os limiares T_1 e T_2. Com auxílio dessa característica, uma curva de ganho *versus* tempo para o controle de temperatura pode ser obtida. Tal curva deve ter uma característica inversa da curva de aumento de temperatura com o tempo. Ela deve ser ajustada em 0 dB para os instantes de tempo em que a temperatura está situada entre T_1 e T_2, e segue a expressão (6), que modela a característica inversa da curva de temperatura função do tempo, para os instantes referentes ao intervalo de temperatura entre T_2 e T_1.

$$G(n) = 20\log\left(\sqrt{\frac{Ta(n)}{Ta(n-1)}}\right), \qquad (6)$$

onde $G(n)$ representa o ganho entre T_2 e T_1, e $Ta(n)$ e $Ta(n-1)$ denotam a temperatura armazenada no instante atual e anterior, respectivamente, considerando o intervalo entre T_1 e T_2.

O controle do deslocamento do cone, como mostrado na Fig. 5, é feito através de um limitador de deslocamento. O limite é dado como uma porcentagem do deslocamento máximo do cone, fornecido pelo fabricante do alto-falante. Assim, o deslocamento do cone está mantido dentro dos limites dado pelo fabricante.

Como o deslocamento do cone e a potência entregue ao alto-falante são obtidos em tempo real, programa-se que quando o deslocamento do cone for maior do que o limite pré-definido, pela primeira vez, esse valor de potência é então armazenado. Dessa forma, quando o deslocamento do cone tender a ser maior do que tal limite, a potência entregue ao alto-falante será limitada ao valor da potência armazenada, evitando, assim, que o alto-falante seja danificado.

3. RESULTADO DE SIMULAÇÕES

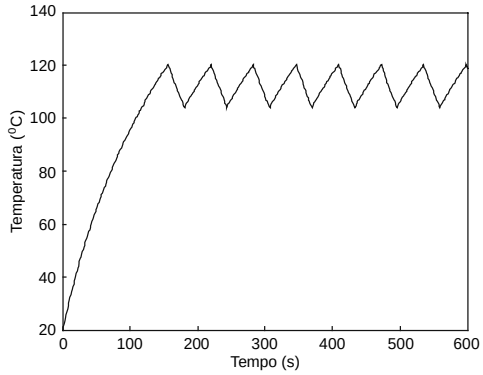
Para avaliar o funcionamento dos sistemas de controle de temperatura da bobina e deslocamento do cone do alto-falante, são realizadas simulações considerando o uso de modelos lineares tanto para o deslocamento do cone quanto para o modelo térmico do alto-falante [1], [13].

Os dados usados nas simulações são os de um alto-falante profissional típico. São considerados o efeito do aumento da resistência DC da bobina com a temperatura, um deslocamento máximo do cone de 12 mm, e os limiares de temperatura inferior $T_1 = 110$ °C e superior $T_2 = 120$ °C.

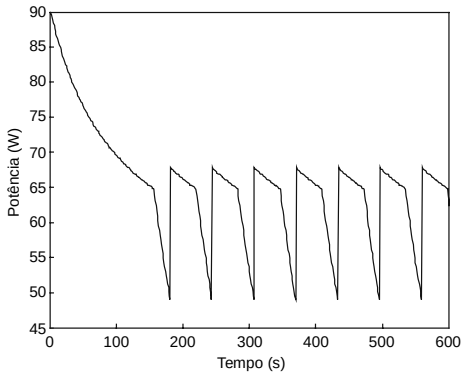
A Fig. 8 ilustra a atuação do sistema de controle de temperatura da bobina e as Figs. 9 e 10, a atuação do sistema de controle de deslocamento do cone, ambos

operando separadamente. A Fig. 9 mostra a potência entregue ao alto-falante e o correspondente deslocamento do cone sem considerar a atuação do controle, e a Fig. 10 mostra os mesmos parâmetros com o sistema de controle atuando.

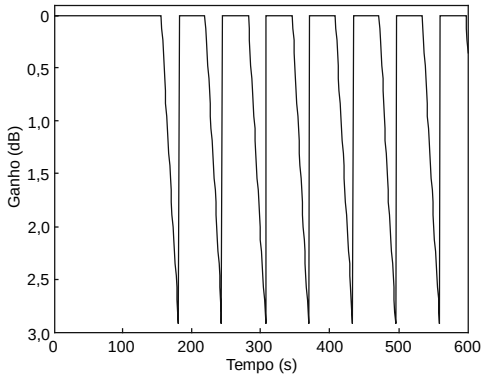
Após a simulação dos sistemas de controle em separado, realizamos uma simulação considerando a atuação simultânea de ambos os sistemas de controle (Fig. 11). Assim, pode-se inferir dos resultados obtidos via simulação numérica a eficácia do sistema de controle quando implementado com um processador de sinais digital.



(a)

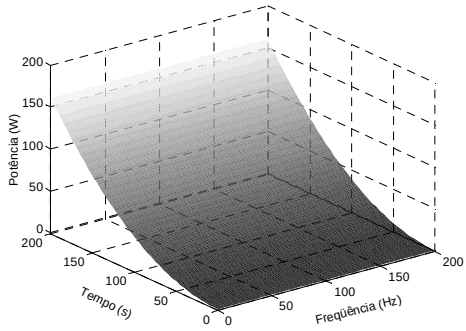


(b)

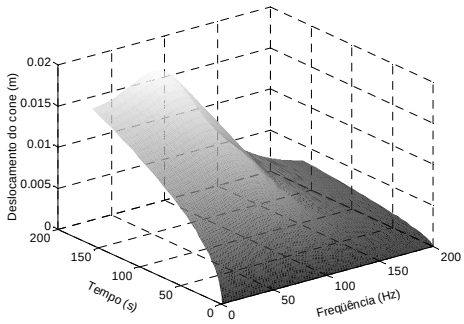


(c)

Fig. 8 - Atuação do sistema de controle de temperatura. (a) Curva da temperatura; (b) curva da potência entregue ao alto-falante; (c) curva do ganho para o controle de temperatura.

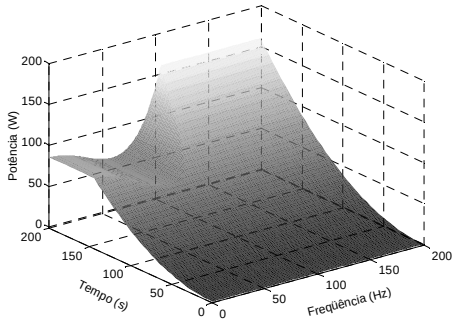


(a)

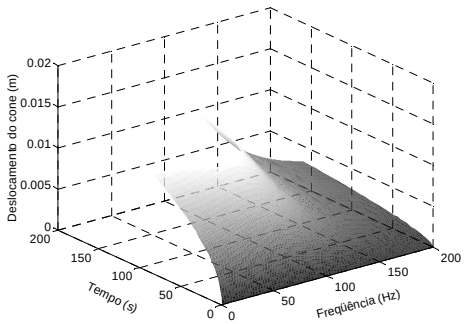


(b)

Fig. 9 – Desconsiderando o sistema de controle para o deslocamento do cone. (a) Curva da potência entregue ao alto-falante; (b) curva do deslocamento do cone.

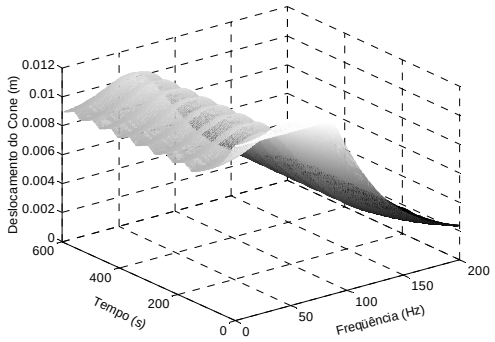


(a)

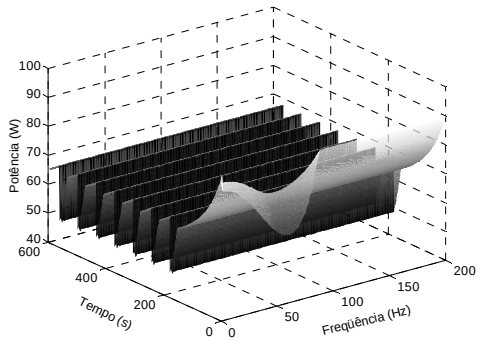


(b)

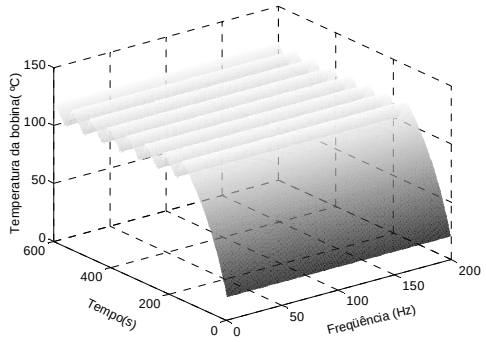
Fig. 10 – Atuação do sistema de controle para o deslocamento do cone. (a) Curva da potência entregue ao alto-falante; (b) curva do deslocamento do cone.



(a)



(b)



(c)

Fig. 11 – Atuação simultânea de ambos os controles. (a) Curva do deslocamento do cone; (b) curva da potência entregue ao alto-falante; (c) curva da temperatura da bobina.

4. CONCLUSÕES

Este artigo propõe um sistema de controle de temperatura da bobina móvel e de deslocamento do cone de alto-falantes para operação em sistemas de sonorização de alta potência, visando uma maior eficácia de controle e o máximo aproveitamento do sistema. A abordagem proposta usando DSP mostrou-se eficaz, sendo relativamente simples e de baixo custo de implementação.

5. AGRADECIMENTOS

Agradecemos ao Eng. Walter Antônio Gontijo pela ajuda referente a parte digital deste trabalho e ao Eng. Homero Sette Silva pelas muitas discussões envolvendo o presente trabalho, como também alguns esclarecimentos sobre alto-falantes.

6. REFERÊNCIAS BIBLIOGRÁFICAS

[1] C. Zuccatti, “Thermal Parameters and Power Ratings of Loudspeakers,” *J Audio Eng. Soc*., vol. 38, no. 1/2, pp. 34-39, Janeiro/Fevereiro 1990.

[2] D. J. Button, “Heat Dissipation and Power Compression in Loudspeakers,” *J Audio Eng. Soc.*, vol. 40, no. 1/2, pp. 32-41, Janeiro/Fevereiro 1992.

[3] P. J. Chapman, “Thermal Simulation of Loudspeakers,” 104th *AES Convention*, AES Preprint no. 4667, 1998.

[4] M. Buck, “Measuring Loudspeakers Voicecoil Temperature,” 106th *AES Convention*, AES Preprint no. 4969, 1999.

[5] C. A. Henricksen, “Heat-Transfer n Loudspeakers: Analysis, Measurement, and Design,” *J Audio Eng. Soc*., vol. 35, no. 10, pp. 778-791, Outubro 1987.

[6] H. S. Silva, “O Alto-falante em Regime de Grandes Sinais,” 1º SEMEA - Seminário de Engenharia de Áudio, Belo Horizonte, MG, CD-ROM, Junho 2002.

[7] G. Penkov e P. Valtchev, “Overheat Protection Circuits for Moving-Coil Loudspeakers,” 90th *AES Convention*, AES Preprint no. 3036, 1991.

[8] I. G. Behler, “Measuring the Loudspeaker’s Impedance During Operation for the Derivation of the Voice Coil Temperature,” 98th *AES Convention*, AES Preprint no. 4001, 1995.

[9] W. Klippel, “Measurement of Large-Signal Parameters of Electrodynamic Transducer,” 107th *AES Convention*, AES Preprint no. 5008, 1999.

[10] W. Klippel, “Nonlinear Adaptive Controller for Loudspeakers with Current Sensor,” 106th *AES Convention*, AES Preprint no. 4864, 1999.

[11] H. S. Silva, “Variação da Resistência da Bobina Móvel em Função da Temperatura,” 1º SEMEA - Seminário de Engenharia de Áudio, Belo Horizonte, MG, CD-ROM, Junho 2002.

[12] J. R. Wright, “An Empirical Model for Loudspeakers Motor Impedance,” *J Audio Eng. Soc.*, vol. 38, no. 10, pp. 749-754, Outubro 1990.

[13] H. S. Silva, “Modelo não Linear do Alto-Falante para Pequenos Sinais,” 1º SEMEA - Seminário de Engenharia de Áudio, Belo Horizonte, MG, CD-ROM, Junho 2002.

[14] P. Chapman, “Complete Protection of an Active Loudspeaker,” 108th *AES Convention*, AES Preprint no. 5112, 2000.

CLICDATA
MULTIMÍDIA

Projeto e produção do CD

fone/fax: (48) 232.1357

clicdata@clicdata.com.br

www.clicdata.com.br