Anais do I Congresso Brasileiro de Engenharia de Áudio da AES-Brasil apresentado na VII Convenção da AES-Brasil

AUDIO AES

LINSE
Laboratório de Circuitos e Processamento de Sinais

Expo-Center Norte
São Paulo, 26 a 28 de maio de 2003

sumário

Apresentação

Créditos

Trabalhos

Autores

apresentação

Sejam todos bem vindos à VII Convenção 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 apoia 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 têm sido compostas de exposições e palestras convidadas. Neste ano de 2003, pela primeira vez, está sendo realizado um Congresso Científico associado à VII Convenção da AES Brasil. Esse Congresso visa dar também à Convenção um caráter mais científico/acadêmico, apresentando trabalhos técnicos relacionados aos avanços da área de engenharia de áudio. A organização técnica desse Congresso contou com a participação ativa de três importantes universidades brasileiras (UFSC, UFMG e UFRJ). 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. Sidnei Noceti Filho - Coordenador Técnico

créditos

I Congresso Brasileiro de Engenharia de Áudio da AES Brasil

São Paulo, 26 a 28 de maio de 2003.

Coordenação Geral:

Joel Brito (Presidente-AES Brasil)

Coordenação Técnica:

Sidnei Noceti Filho (UFSC)

Comissão Técnica:

Hani Camille Yehia (UFMG)

Luiz Wagner P. Biscainho (UFRJ)

Pedro Donoso Garcia (UFMG)

Rosalfonso Bortoni (Studio R)

Rui Seara (UFSC)

Sergio Lima Netto (COPPE/UFRJ)

Sidnei Noceti Filho (UFSC)

créditos

Relação de Revisores:

Acir Edvam Ozelame

Antônio Carlos Moreirão de Queiroz

Christian Gonçalves Herrera

Fábio Violaro

Fernando Antônio Pinto Barúqui

Fernando Santana Pacheco

Hani Camille Yehia

Hermes Aguiar Magalhães

Izabel Christine Seara

Joarez Bastos Monteiro

Marcello Luiz Rodrigues de Campos

Paulo Fernando Seixas

Pedro Donoso Garcia

Regis Rossi Alves Faria

Rosalfonso Bortoni

Rui Seara

Rui Seara Junior

Sergio Lima Netto

Sidnei Noceti Filho

Solimar de Souza Silva

Análise do Comportamento de Alto-falantes Excitados por Fonte de Corrente para Pequenos e Grandes Sinais

Rosalfonso Bortoni; Sidnei Noceti Filho; José Antônio Justino Ribeiro; Rui Seara

Considerações Sobre o Uso de Funções de Transferência Clássicas no Projeto de Divisores de Freqüência

André Luís Dalcastagnê; Sidnei Noceti Filho; Homero Sette Silva

Conversão de arquivos WAVE em MIDI

F. Paiva; G.C.R. Abrahão; R.J.R. Cirigliano; R.S. Maia; F.G.V. Resende Jr.

Difusores Acústicos III

Alejandro Bidondo

Equalização de Áudio: Considerações Relevantes

Phabio Junckes Setubal; Sidnei Noceti Filho; Rui Seara

Implementação de Efeitos em Sinais Digitais de Áudio

Bernardo Machado da F. e Silva; Vicente Coelho Ferreira

Innovations on the Objective Assessment of Audio Quality

Jayme Garcia Arnal Barbedo; Amauri Lopes

Modulação Sigma Delta em Áudio

Christian Gonçalves Herrera

Sistema de Efeitos para Guitarra Aplicados em Tempo Real Utilizando DSP

Rodrigo Coura Torres; José Manoel de Seixas

Transcrição Musical Automática com Bancos de Filtros

Cristiano N. dos Santos; Luiz W. P. Biscainho; Sergio L. Netto

autores

Alejandro Bidondo

Difusores Acústicos III

Amauri Lopes

Innovations on the Objective Assessment of Audio Quality

André Luís Dalcastagnê

Considerações Sobre o Uso de Funções de Transferência Clássicas no Projeto de Divisores de Freqüência

Bernardo Machado da F. e Silva

Implementação de Efeitos em Sinais Digitais de Áudio

Christian Gonçalves Herrera

Modulação Sigma Delta em Áudio

Cristiano N. dos Santos

Transcrição Musical Automática com Bancos de Filtros

F. Paiva

Conversão de arquivos WAVE em MIDI

F.G.V. Resende Jr.

Conversão de arquivos WAVE em MIDI

G.C.R. Abrahão

Conversão de arquivos WAVE em MIDI

autores

Homero Sette Silva

Considerações Sobre o Uso de Funções de Transferência Clássicas no Projeto de Divisores de Freqüência

Jayme Garcia Arnal Barbedo

Innovations on the Objective Assessment of Audio Quality

José Antônio Justino Ribeiro

Análise do Comportamento de Alto-falantes Excitados por Fonte de Corrente para Pequenos e Grandes Sinais

José Manoel de Seixas

Sistema de Efeitos para Guitarra Aplicados em Tempo Real Utilizando DSP

Luiz W. P. Biscainho

Transcrição Musical Automática com Bancos de Filtros

Phabio Junckes Setubal

Equalização de Áudio: Considerações Relevantes

R.J.R. Cirigliano

Conversão de arquivos WAVE em MIDI

R.S. Maia

Conversão de arquivos WAVE em MIDI

autores

Rodrigo Coura Torres

Sistema de Efeitos para Guitarra Aplicados em Tempo Real Utilizando DSP

Rosalfonso Bortoni

Análise do Comportamento de Alto-falantes Excitados por Fonte de Corrente para Pequenos e Grandes Sinais

Rui Seara

Análise do Comportamento de Alto-falantes Excitados por Fonte de Corrente para Pequenos e Grandes Sinais

Equalização de Áudio: Considerações Relevantes

Sergio L. Netto

Transcrição Musical Automática com Bancos de Filtros

Sidnei Noceti Filho

Análise do Comportamento de Alto-falantes Excitados por Fonte de Corrente para Pequenos e Grandes Sinais

Considerações Sobre o Uso de Funções de Transferência Clássicas no Projeto de Divisores de Freqüência

Equalização de Áudio: Considerações Relevantes

Vicente Coelho Ferreira

Implementação de Efeitos em Sinais Digitais de Áudio

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na VII Convenção Nacional
26-28 de maio de 2003, São Paulo, 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.

Innovations on the Objective Assessment of Audio Quality

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

SUMMARY

This paper presents new features for objective assessment of audio quality that were incorporated to the PEAQ method, which is currently adopted as standard by the International Telecommunication Union (ITU). These modifications lead to a new procedure named Objective Measure of Audio Quality (Medida Objetiva da Qualidade de Áudio - MOQA). The performance of the proposed method was measured over databases and is compared here to that obtained by PEAQ method.

1. INTRODUCTION

The digital transmission and storing of audio signals have been strongly based on algorithms for data reduction, which are adapted to several peculiarities of human auditory system, as the masking effects. Such algorithms do not necessarily aim the minimization of distortions. They intend some manipulations of the audio signal, in such a way that the users minimally perceive them. Therefore, the quality of the so-called perceptual coders cannot anymore be assessed by the traditional methods based on the global value of distortion, such as the signal-to-noise ratio (SNR) and total harmonic distortion (THD). In certain cases, the noisy structures are so effectively masked by the signal that they become nearly inaudible, even when the signal has a SNR as low as 13 dB.

In this way, the use of subjective tests is necessary to perform confident quality assessments of perceptual codecs. Nevertheless, such tests are expensive in terms of time and cost. So, the development of objective measures able to replace efficiently the subjective tests is highly desirable.

Some methods were proposed at the late seventies, but the first perceptual codecs (MPEG and Dolby) at the late eighties turned such measures obsolete. Then, in 1994, the ITU-R (*International Telecommunication Union - Radiocommunication*) performed an open call of proposals, in order to establish a standard for objective audio quality measurement. Six methods were proposed [1, 2, 3, 4, 5], none of them reaching the minimum acceptable performance. After that, the proponents concentrated their efforts in the development of a single method composed by the best former proposals, originating the method Perceptual Evaluation of Audio Quality (PEAQ) and a new recommendation, the ITU-R BS-1387 [6]. This method presents a clearly better performance than its predecessors. Nevertheless, it is not good enough for the most part of practical conditions. Such situation has motivated the search for new methods capable to overcome those limitations. In that context, a new method (MOQA), the object of this paper, has been developed. More details about its implementation can be found in [7].

Section 2 presents the main characteristics common to every objective audio quality measure. Section 3 resumes the structure used on the MOQA method. Section 4 discusses with details the modifications and innovations proposed. Section 5 presents the tests, their results and a comparison with the performance reached by the PEAQ method. At last, Section 6 presents the conclusions and final considerations.

This work was supported by Fapesp, Proc. n. 01/04144-0.

2. PERCEPTUAL MEASURES

Figure 1 shows the basic structure common to all objective audio quality measures. Each block is briefly explained in the following.

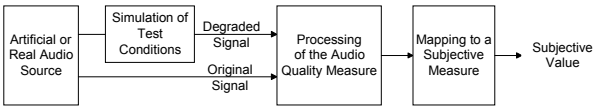


Fig. 1 - General structure of perceptual measures

- Artificial or real audio source: the test signals to be used are usually the same musical excerpts used in the subjective assessment of codecs. However, in principle any kind of audio signal, including the artificial ones, can be used.

- Simulation of test conditions: here, the test signal is submitted to conditions that may potentially introduce degradations, as several kinds of codification, bit errors, noise, or any other situation desired to be assessed; at same time, a unaltered version of the signal is kept for later comparison with the degraded version.

- Audio quality measure: this stage is the most important of any method for audio quality assessment; here are included the time-frequency decomposition, the modelling of the human hearing features (among them, the masking, briefly described in section 2.1) and the cognitive subtraction, which produces the perceptual difference among the signals. As result, a quality measure of the tested signal is obtained.

- Mapping to a subjective measure: this stage transforms the objective measure, represented in a particular objective scale, into a standard ITU subjective scale. This stage is optional and can be performed by polynomials or artificial neural networks.

2.1. Masking Modeling

Masking is the most important phenomenon in the quality perception of a signal. For that reason, its correct modelling is an essential factor in the performance of an objective method for audio assessment.

The masking phenomenon is due to ear limitations in terms of temporal, spectral and amplitude resolution, combined to an also limited dynamic range. When two signals are close enough to each other, in time or frequency domain, the weaker signal may become inaudible due the presence of the stronger one.

The modelling of masking effects is a feature common to all perceptual methods. The simultaneous (spectral) masking is always modelled by applying a spreading function, which corresponds to the shape of an average masking curve. Temporal masking effects are frequently implicitly modelled in the expressions of the model, but in a crude way, due to the limited temporal resolution of the time-frequency decomposition normally used.

3. THE MOQA METHOD

In this first version, the MOQA method borrowed several characteristics from the PEAQ method, as, for instance, its basic structure. As the research evolves, it is expected that both methods become more unrelated, since several new features must be implemented in next versions. Nevertheless, it is important to note that the version presented here has its own implementation, which has enough peculiarities and innovations to be considered as an original method. Furthermore, those new features represent important contributions towards a more efficient audio assessment methods. Such new features will be detailed explored in Section 4.

The general structure of MOQA method is shown in Figure 2. As can be seen, the input signals correspond to the original signal, which will be taken as reference, and the degraded signal, which is the original signal submitted to some kind of condition capable to insert distortions.

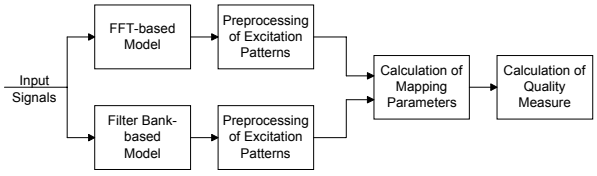


Fig. 2 - General structure of MOQA method

As in PEAQ, two different models for the ear were implemented. The main distinctive characteristic of the MOQA models is the strategy adopted to perform the time-frequency decomposition (Fast Fourier Transform or Filter Bank). The models will be described with more details in the following, as well the processings indicated in Figure 2.

3.1. FFT-Based Model

The main feature of this model is the low computational burden. Its basic scheme is shown in Figure 3.

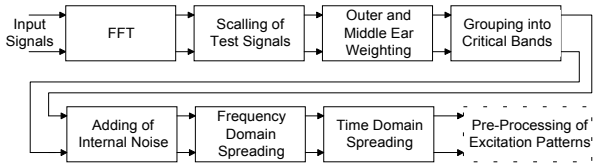


Fig. 3 - Basic scheme of the FFT-based model

The inputs for this model, which are the original and degraded signals aligned in the time domain and sampled at a rate of 48 kHz, are divided into 42 milliseconds blocks (2048 samples), with a 50% superposition. After that, a Hanning window is applied.

Each windowed block is transformed to the frequency domain by a FFT algorithm. At last, each block is scaled to the playback level (if such level is unknown, it is recommended the adoption of 92 dB_{SPL}). A weighting function is applied to the spectral coefficients in order to model the frequency response of outer and middle ears.

The weighted spectral coefficients are grouped into critical bands and an offset is added to simulate the internal noise of the auditory system. The next step is to submit the signals to two spreading functions, the first one modeling the frequency domain masking and the second one modeling the time domain masking (see Figure 3). Such processing results in the so-called excitation patterns, which are submitted to some additional processing, as described latter.

3.2. Filter Bank-Based Model

The main feature of this model is its good temporal resolution, which allows one to obtain, theoretically, more precise results. On the other hand, the computational effort demanded is higher. Figure 4 shows the basic scheme adopted for this model.

The original and degraded signals at the input of this model are adjusted to the playback level and are sent through a high-pass filter to remove DC and subsonic components. Then, the signals are decomposed into 40 bands by linear-phase FIR filters, which are equally distributed across the perceptual scale. A frequency-dependent weighting is applied

to the decomposed signal, in order to model the spectral features of outer and middle ears. The level-dependent spectral resolution of the input components to the auditory filters is modeled by a frequency-domain convolution of the outputs with a level-dependent spreading function.

The envelopes of the signals are calculated using the Hilbert-transform of the band pass signals (rectification) and a time domain convolution with a window function is computed in order to model backward masking. Then, a frequency dependent offset is added to take into account the internal noise in the auditory system and to model the threshold in silence. Finally, a second time-domain convolution is carried out using an exponential spreading function that take into account the forward masking. The resulting vectors are the so-called "excitation patterns".

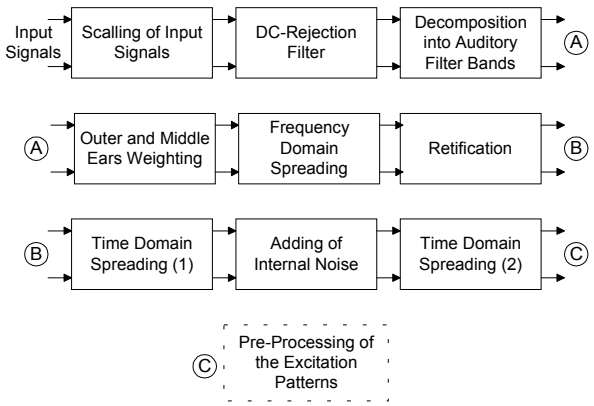


Fig. 4 - Basic scheme of the FFT-based model

3.3. Pre-Processing of Excitation Patterns

This stage consists of four procedures aiming to prepare the excitation patterns for an adequate extraction of the output parameters:

1- Level and pattern adaptation: the average levels of the original and degraded signals are adapted to each other by filters and correction factors, in order to compensate level disparities and linear distortions.

2- Modulation: filters and weighting factors are applied in order to calculate a measure for the modulation of the envelope at each filter output. The resulting patterns are used to calculate some output parameters.

3- Loudness: this processing aims to determine the loudness of the resulting excitation patterns, in agreement to Zwicker's expression for the specific loudness [8]. The resulting patterns are also used in the calculation of some output parameters.

4- Masking threshold: it is obtained by the appropriate weighting of the excitation patterns, and it is used in the calculation of one output parameter.

3.4. Output Parameters

The model output parameters consist are submitted to an artificial neural network that produces a quality measure to the analyzed signal. Those parameters that were inspired in the PEAQ method are described in the following, divided into groups in agreement to their purpose; the new ones will be presented in Section 4.

1- Modulation difference: it is calculated from the temporal envelopes of original and degraded signals. This group is composed by four parameters, three related to the FFT-based model and one related to the filter bank-based model.

2- Noise loudness: the parameters belonging to this group estimate the partial loudness of distortions added to the original signal. This group is composed by three output parameters, two from the filter bank-based model and one from the FFT-based model.

3- Bandwidth: the two parameters resulting from this stage provide an estimation of the average bandwidth of the original and degraded signals, in terms of FFT lines.

4- Noise-to-mask ratio: this group is composed by two parameters, one from each model, consisting on the relationship between the noise and masking patterns levels, in dB.

5- Relative number of disturbed frames: it is composed by only one output parameter deriving from the FFT-based model, and is given by the number of frames whose mask-to-noise ratio exceeds determined value in dB.

6- Detection probability: this group estimates the probability that a listener will detect a given disturbance. In PEAQ, it is composed by two parameters, both related to the FFT-based model. One of them was eliminated because its results are very poor. Furthermore, the other parameter was modified, leading to much better results.

The mapping of all those parameters to a subjective quality estimation was performed using a multi-layer perceptron neural networks (MLPNN) with one hidden layer. The activation functions used for the hidden layer were hyperbolic tangents. For the output layer, the activation function was linear. The training was carried out using a Levenberg-Marquardt second-order optimization method [9], with an optimization criteria based on the least squares.

4. NOVEL FEATURES

The new features presented next are divided into sub-sections according to the stages they were implemented inside the algorithm.

4.1. FFT-Based Model

Two innovations were introduced for this model, as described next.

1- The calculations involved in the time-frequency decomposition using the FFT would demand a high quantity of values, specially for audio signals. Some signals contain more than 3,000,000 samples by channel, what corresponds to 3,000 frames of 2,048 samples. At the end, almost 100 Mbytes will be needed to store all the variables. Many computers do not have enough Random Access Memory to deal with such demand; even if this amount of memory is available, the execution of the program will become too slow.

To solve this problem, two solutions were adopted. In the first one the total number of frames is split into 64 separated groups before the application of the FFT; in this way, the FFT is calculated for a few frames at each time, and then the temporal samples are immediately eliminated. This procedure itself reduces the storage needs to the half. The second solution consists in discarding the spectral components that are not used in the subsequent processing. The band used in all calculations is limited to 18 kHz and the sample rate is 48 kHz. Then, as the number of samples used in the calculation of the FFT is 2,048, this implies that only the first 768 samples are useful; the remaining samples are discarded. This procedure reduces the storage needs in about 60 %. Figure 5 illustrates this last procedure, where the first plot shows the entire amplitude spectrum of a given frame, the second plot shows only the remaining components after the discard of unneeded lines and the third one presents a zoom of the resulting spectrum. Both techniques combined represent a reduction in almost 90% of the time required for this stage.

2- In the frequency domain spreading performed for the PEAQ method, a normalization factor is used in order to keep the frame energies constant. However, such factor not played its role efficiently. For that reason, it was replaced by a simpler procedure, where the relation between the energies before and after the spreading is computed for each frame. Then, the frames submitted to the spreading are multiplied by this value. The shapes of the curves obtained by this procedure are very close to that ones obtained by the other approach, but with more adjusted levels. Besides, this procedure is computationally simpler.

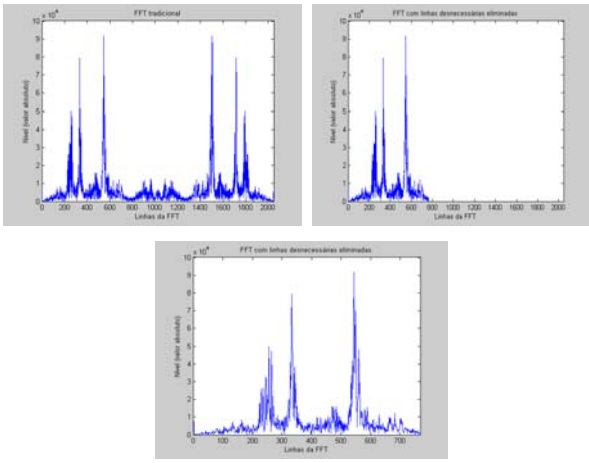


Fig. 5 - Elimination of unneeded spectral lines

4.2. Filter Bank-Based Model

This model suffered several modifications compared to that one implemented in the PEAQ. Most of them are the replacement of some processing for simpler ones, which demand less computational resources keeping the same effectiveness [10].

The most important innovation introduced in this model is related to the filter implementation. In PEAQ, the FIR filters are implemented recursively. This approach inserts a pole in the equations of the filters that must be canceled by the correct allocation of zeros. Then, although the filters still present a finite impulse response, its implementation is quite related to that ones used to IIR filters, what reduces considerably the computational burden required.

Due to particular features of the environment where the program was implemented (Matlab), such approach led to very slow runs. Thus, the development of a structure able to limit the use of loops and better adapted to the faced peculiarities was strongly recommended. After several attempts, a very efficient structure, which uses only matrix operations, was created [10]. Such procedure is composed by seven stages, as described next.

1- 1,424 zeros are added to the beginning of each channel, in order to allow the filtering to be applied to the correct samples.
2- The coefficients of each filter are grouped into a 40 by 1,456 matrix. The number of lines corresponds to the number of coefficients of the filter with the longest impulse response (the filters with less coefficients are completed by zeros). The number of columns is the number of filters.
3- The order of the filter coefficients is inverted, which means that the last coefficient becomes the first one and vice-versa. Such procedure aims to facilitate the implementation of the convolution that is performed next.
4- The signals are divided into frames of 20,000 samples; such division aims to reduce the storage needs. The frames are superposed by 1,440 samples in order to apply the convolution to the correct samples, avoiding future problems when concatenating the results. A decimation by a factor of 32 is automatically performed, since the filtering is applied only for samples multiple of 32. Figure 5 shows more details related to this division.
5- Each frame is divided into smaller frames of 1,456 samples, with a 1,424 samples superposition, which are then concatenated into a matrix of 625 by 1,456 samples. Such division allows the filter coefficients be matricially applied to the whole signal , as showed in Figure 6. As a consequence, all the 40 filters can be applied at once. In other words, each line of that matrix is the prior line shifted by 32 samples, modelling the fact that the filtering is applied only for the samples whose index is multiple of 32. The superposition of 1,440 samples guarantees that the first sample for which the filtering is applied is located exactly 32 places after the last sample considered in the previous frame, as showed in Figure 6.
6- Next, a matrix-based filtering is performed, where the matrices of filter coefficients, which dimension is 40 by 1,456, are multiplied by the transposed matrix of the shifted samples of the signal, resulting in a 40 by 625 filtered matrix. Such procedure is repeated for each frame resulting from the first division of the signal.
7- The results obtained for each frame are concatenated into a single matrix composed by 40 lines and a number of columns 32 times smaller than the total length of the signals.

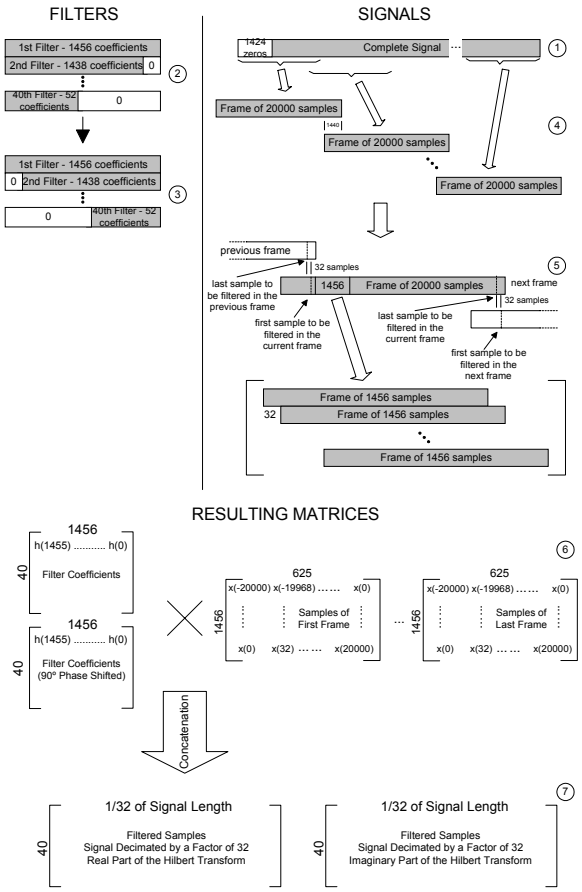


Fig. 6 - Signal-filtering scheme

4.3. Model Output Parameters

Some new parameters were introduced to extract the most information from the signals; some of them were never used in any previous method for audio quality assessment. Such parameters are described next.

1- Detection probability: this is not exactly a new parameter. However, it was observed that the implementation suggested in the PEAQ led to very poor results, since the detection probabilities returned by the algorithm were almost invariably very close to 1. This behaviour turns this variable useless to the neural network, since it carries very little information about the signal. To face this problem, a modification in the averaging of the probability found for each component of the signal was introduced. Such modification led this parameter alone to reach excellent results, even better than the ones obtained by most of previous methods. This innovation was one of the most important reasons for the good results reached by the MOQA method, as presented in Section 5.

2- Channel correlation: this is a new parameter and yet unpublished. It is calculated only for the patterns resulting from the filter bank-based model. Its motivation lies on the observation that eventual phase shifts between the channels can be extremely annoying to the listener. The strategy to quantify such phenomenon is simple: the correlation between the channels is calculated; as closer to 1 is its value, less is the disturbance (the channels are in phase); on the other hand, if its value is close to –1, it is very likely that a severe shift between the channels has happened (channels in quadrature), causing a significant annoyance to the listener.

3- Perceptual streaming: this concept was published in [11] and was not used in the PEAQ algorithm. It is a central cognitive feature of the human auditory system that separates different auditory events into distinct streams. If the codec distorts the input signal in such a way that the output signal is split by ear into two parts by the auditory system, the original signal and the distortion, then the disturbance caused by such distortion is more intense than when both parts (signal and distortion) are integrated into a single perception.

   The modelling of the perceptual streaming is complicated. The adopted approach assumes that when the codec rarely will introduce a new time-frequency component that perfectly integrates with the input signal. In this way, the output signal will be decomposed into the two different perceptions previously described. However, when the codec eliminates a component, the output signal cannot be decomposed in the same way, implying in a less severe distortion. This effect is quantified by a correction factor for the noisy disturbance, which is the perceptual difference between the signals. The correction factor is based in the relation between the powers of the degraded and original signals in a certain point of the time-frequency plane, indicating how much such they are different.

4- Informational masking: this concept also was published in [11] and was not used in the PEAQ algorithm. It is a central cognitive feature of the human auditory system where distortions that must be audible, since they are above the audibility threshold, become inaudible due to the informational content (complexity) of the masker signal. The perceptual streaming effect can diminish the informational masking effect. When a signal can be decomposed in the terms of perceptual streaming, then the informational masking effect will be smaller than in cases where such decomposition is not possible. For that reason, both effects must be modeled together.

   The informational masking effect is implemented based on the variation of power in the time domain, for each frequency band of the original signal. The variance is taken into account in the calculation of the noisy disturbance for each temporal frame, in such a way that complex signals with a larger power variation produce a more pronounced masking effect than simpler signals.

5- Difference signal loudness using L_p norms: this parameter uses a strategy adopted by the Perceptual Evaluation of Speech Quality (PESQ) method [12], where the average calculation is performed using different norms, in order to emphasize certain characteristics of the difference between the signals. Firstly, a L_3 norm is calculated, meaning that the components are raised to 3 and summed, and then the cube root is extracted. Such procedure provides a slight emphasis to the signal energy peaks. The same procedure is conducted in the time domain, now using a L_6 norm, which emphasizes even more the peaks. As result, a single value representing the loudness of the difference signal is obtained.

5. TESTS AND RESULTS

The features of the tests and the results obtained are presented in the following.

5.1. Databases

The databases containing audio files and respective subjective measures are not public available, so they are very difficult to be obtained. Among the ten databases used in the validation of the PEAQ method [6], three were exceptionally discharged to this research, resulting in a total of 239 pairs of files.

The files present in the databases have a large number of features in terms degradation types and levels, as well in terms of content. Therefore, despite this is not a large set of files, it is representative enough to allow the extraction of consistent results and conclusions.

5.2. Tests Description

The parameters whose individual results were inappropriate to supply the neural network were eliminated before the tests. From this selection, seven parameters from the FFT-based model and four parameters from the filter bank-based model remained.

Several configurations for the neural network were tested. The configurations were obtained changing two parameters: number of inputs for the neural network and number of neurons in the hidden layer, as described next.

- Parameters used as inputs to the neural network: the strategy to test the importance and contribution of each parameter consisted, initially, in performing tests using all the eleven parameters remaining from the selection stage as input to the net; then, they were gradually eliminated and, after each removal, the performance was computed. The parameters with lower correlation with the subjective scores were eliminated first. Tests showed that for four inputs or less, the performance of the method drops quickly.

- Number of neurons in the hidden layer: the number of neurons was varied from 2 to 25; such tests revealed that, above six neurons, the correlations do not present a significant improvement.

Finally, two-thirds of the files were used in the trainings and one-third in the tests.

5.3. Results

The criterion used to validate the method was the correlation between the objective and subjective parameters. The average correlation obtained for the three databases was 0.86 (1 is the best correlation value). This can be considered an excellent result, especially if one considers that the best mean correlations reached by the PEAQ did not exceed 0.84 [6]. The Figure 6 illustrates the performance of the MOQA method. A higher concentration of points around the mapping line indicates good results.

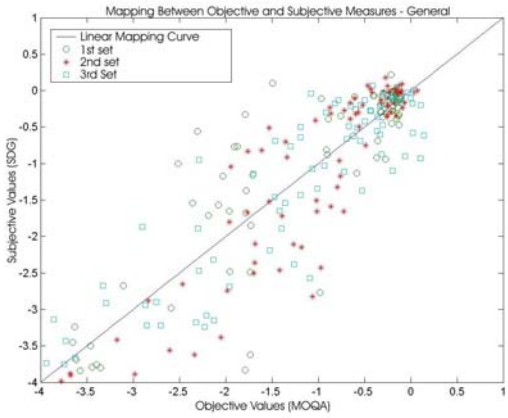


Fig. 7 - Signal-filtering scheme

This good performance is due to the extra information extracted by the new parameters. Some of those parameters showed a good individual performance. In particular, the "detection probability" parameter, modified as shown in Section 4.3, showed higher correlations with the subjective values (0.71) than any other method prior to PEAQ. This is the case of the DIX (Distortion Index) method [1], whose structure was used in the development of the filter bank-based model of the PEAQ method [6], and whose correlations barely reached 0.7. Such performance becomes even more impressive when one takes into account that this parameter was implemented for the FFT-based model, theoretically inferior to the filter bank-based model. Therefore, it is likely that this parameter can reach an even better performance if extracted after the application of a filter bank. This implementation will be performed in the next versions of the program.

Unfortunately, it is not possible to directly compare the performances of PEAQ and MOQA, since the training and test sets used in both cases are different. However, there are two facts that turn the results presented earlier very significant:

- The databases used in the tests with the MOQA have a range of conditions almost as large as that one found in the ten databases used for the PEAQ tests, what means that the MOQA was tested with the same hard circumstances faced by the PEAQ method.

- The set of data used to train the artificial neural network of the MOQA method was significantly smaller than that one available for the PEAQ tests, what means that the PEAQ method had the opportunity to be much better trained, and then to generate a better mapping surface.

Therefore, as the set used in this work is limited but wide, what turns more difficult to reach good results, it is possible to say with a high degree of confidence that the MOQA reached a better performance than PEAQ. Additionally, if the complete set of data was available to the MOQA, it is very likely that it would reach even better results, since it could be better trained.

It is important to detach that the improvement reached by the MOQA is very significant, despite the little difference between the correlations of both methods. Most of the effort spent in the last years resulted in modest improvements, but still deserved distinction [6]. Moreover, as high are the correlations, more difficult is to reach results numerically much superior. In this context, even the slightest improvements are relevant.

5. CONCLUSIONS

The proposed method performed better than the PEAQ method, which is currently adopted as standard by the ITU. Such result is a consequence of the extra information extracted from some new parameters. In particular, the "detection probability" parameter supplied high quality information, allowing the artificial neural network to generate a mapping surface better fitted with the actual subjective values.

Despite the promising results, the performance of the MOQA is still under the minimum desirable. On the other hand, these are only preliminary results of a doctoring project that intends to overcome many of the limitations exhibited by the current methods.

Therefore, it is very likely that modifications to be introduced in future versions of the program will improve its functioning and increase the correlations. The good results currently obtained indicate that this research can successfully originate a totally new and efficient method.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Thiede, T.; Kabot, E. *A New Perceptual Quality Measure for Bit Rate Reduced Audio*, Contribution to the 100th AES Convention, preprint 4280, Copenhagen, 1996.

[2] Brandenburg, K. *Evaluation of Quality for Audio Encoding at Low Bit Rates*, Contribution to the 82nd AES Convention, preprint 2433, London, 1987.

[3] Beerends, J.G.; Stemerdink, J.A. *A Perceptual Audio Quality Measure Based on a Psychoacoustic Sound Representation*, J. Audio Eng. Soc., vol. 40, pp. 963-978, Dec. 1992.

[4] Paillard, B.; Mabilleau, P.; Morisette, S.; Soumagne, J. *Perceval: Perceptual Evaluation of the Quality of Audio Signals*, J. Audio Eng. Soc., vol. 40, pp. 21-31, Jan. 1992.

[5] Colomes, C.; Lever, M.; Rault, J.B.; Dehery, Y.F. *A Perceptual Model Applied to Audio Bit-Rate Reduction,* J. Audio Eng. Soc., vol. 43, pp. 233-240, April 1995.

[6] ITU-R Recommendation BS-1387, *Method for Objective Measurements of Perceived Audio Quality*, 1998.

[7] Barbedo, J.G.A.; Lopes, A. *A New Method for Objective Assessment of Audio Quality*, submitted to the XX Simpósio Brasileiro de Telecomunicações.

[8] Zwicker, E.; Fastl, H. *Psychoacoustics, Facts and Models*, Springer Verlag, Berlin, 1990.

[9] Bazaraa, M.S.; Sherali, H.D.; Shetty, C.M. *Nonlinear programming*, John Wiley & Sons, New York, 1993.

[10] Barbedo, J.G.A. *1º Relatório Técnico Fapesp - Processo nº 01/04144-0*, Campinas, julho de 2002.

[11] Beerends, J.G.; van den Brink, W.A.C. *The Role of Informational Masking and Perceptual Streaming in the Measurement of Music Codec Quality*, Contribution to the 100th Convention of the Audio Engineering Society, Preprint 4176, Copenhagen, May 1996.

[12] ITU-T Recommendation P.862, Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, 2001.

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na VII Convenção Nacional
26-28 de maio de 2003, São Paulo, 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.

Implementação de Efeitos em Sinais Digitais de Áudio

Bernardo Machado da F. e Silva, Vicente Coelho Ferreira
Universidade do Estado do Rio de Janeiro
Niterói, Rio de Janeiro, Brasil
mailto:vicente@cerj.com.br

RESUMO

O objetivo deste trabalho é implementar efeitos em sinais digitais de áudio através de algoritmos capazes de realizar operações matemáticas que modifiquem as formas de onda destes sinais no domínio do tempo e no domínio da frequência. No domínio do tempo implementaremos os efeitos: eco, delay, reverb, flanging e chorus. Para o domínio da frequência implementaremos um equalizador de 6 bandas de frequência.

INTRODUÇÃO

A implementação dos efeitos no domínio do tempo (efeito eco, efeito delay, efeito reverb, efeito flanging e efeito chorus) e do equalizador de bandas para o domínio da frequência baseia-se na modificação da forma de onda dos sinais digitais de áudio.

Para o desenvolvimento deste trabalho, utilizamos algoritmos compatíveis com um software de modelagem matemática capaz de mudar a estrutura dos sinais de áudio, gerando os efeitos desejados. Porém estaremos apresentando aqui apenas algumas operações matemáticas necessárias na implementação de cada efeito, relacionando as características naturais de cada um deles. Inicialmente, trabalhamos com um arquivo de áudio num formato padrão com a extensão *.wav*, em seguida aplicamos os algoritmos responsáveis pela implementação de cada efeito que realizam todas as operações necessárias, em função de alguns parâmetros inseridos. Alguns detalhes sobre a implementação de cada efeito serão apresentados ao longo deste trabalho.

Mostraremos ainda um pouco da teoria que envolve os efeitos que foram propostos para o domínio do tempo.

Apresentaremos alguns conceitos teóricos necessários para a implementação do efeito equalizador proposto para o domínio da frequência. Falaremos um pouco sobre a teoria de Banco de Filtros e mostraremos a técnica do Banco de Filtros Modulados por Cosseno, mostrando como o equalizador pode ser obtido aplicando estes conceitos.

Faremos ainda alguns comentários sobre a implementação e a aplicação dos efeitos deste trabalho, comparando os resultados obtidos com os processadores de efeitos já existentes no mercado.

IMPLEMENTAÇÃO DOS EFEITOS

O diagrama da Figura 1 mostra todas as etapas necessárias para a implementação de um efeito. O que muda de um efeito para o outro é somente a etapa do processamento.

O primeiro passo consiste em fazer a aquisição digital de um sinal de áudio analógico [1].

Após a aquisição, utilizaremos os algoritmos capazes de modificar a estrutura de um sinal de áudio produzindo o efeito desejado.

Antes de realizarem as operações matemáticas na etapa

Figura 1 - Diagrama Geral.

do processamento, referentes a cada efeito, estes algoritmos transformam um arquivo de áudio *.wav* em uma matriz de *m x n* elementos, onde m representa o número de amostras do sinal digital e n representa o número de canais. Supondo que estejamos trabalhando com um arquivo de áudio estéreo (2 canais) gravado no formato *.wav* a uma taxa de amostragem de 44.1kHz (qualidade de CD) [2] com uma duração de 10 segundos. Com a taxa de amostragem a 44.1kHz para 10 segundos de música, teremos 441.000 amostras do sinal de áudio gravado em cada canal. Sendo assim, após a conversão do arquivo *.wav* para uma matriz *m x n*, esta mesma terá 441.000 linhas e 2 colunas.

A primeira coluna desta matriz representa o canal 1, onde cada elemento desta coluna representa o valor normalizado de amplitude referente à cada amostra do sinal. A segunda coluna representa o canal 2. Sabemos então que o elemento a_{11} representa a primeira amostra do canal 1, o elemento a_{32} representa a terceira amostra do canal 2 e assim por diante. A normalização das amostras é feita convertendo o maior valor positivo de todas as amostras para 1 e o maior negativo para –1. Sendo assim, o sinal normalizado terá amplitude variando entre +1 e -1.

A etapa final consiste na conversão do sinal processado da forma matricial para o formato .wav que posteriormente poderá ser reproduzido pelo PC.

EFEITOS NO DOMÍNIO DO TEMPO

Após a conversão do arquivo *.wav* em uma matriz (etapa comum em todos os efeitos), faremos uso de algumas operações matemáticas bastante simples nas amostras do sinal que desejados modificar, produzindo os efeitos propostos para o domínio do tempo.

Efeito Eco

O efeito eco é o mais simples de todos os efeitos. Ele é obtido atrasando-se o sinal original de um número *d* de amostras, multiplicando este sinal atrasado por um ganho *A* e somando este sinal resultante ao sinal original [3].

O diagrama de blocos da Figura 2 ilustra melhor as etapas deste processo. O sinal original é representado por *X[n]* e o trecho selecionado do sinal é representado por $X_1[n]$. Note que o trecho selecionado do sinal onde se quer aplicar o efeito é definido inserindo valores em unidades de tempo (início do intervalo e fim do intervalo). Estes valores em tempo são convertidos, definindo assim, o intervalo de amostras selecionado.

O eco é um efeito bem simples que pode ser bem observado em quase todos os sons musicais. A repetição atrasada do sinal selecionado é facilmente percebida pelo ouvido humano e interpretada pelo cérebro.

Efeito Delay

O efeito *delay* é similar ao eco, porém com um número *n* de repetições atrasadas, onde *n* depende da variável de entrada *delay feedback p*, *a* é o ganho aplicado na repetição e são relacionados pela seguinte equação [3]:

$$a = 1 - q\,(100 - p) / 100 \qquad (01)$$

onde q = 1,2,3,4,...,n enquanto a > 0 e *p* pode variar entre 0 e 100%.

A variável *q*, que inicialmente possui valor igual a 1, é incrementada de uma unidade a medida em que uma nova repetição é gerada. Quando a variável *a* assume um valor negativo, o algoritmo interrompe a geração de novas repetições, interrompe a incrementação da variável *q* e finaliza o processamento.

A Figura 3 apresenta um diagrama de blocos explicativo descrevendo passo a passo as etapas deste processo. O tempo de eco é o tempo de atraso em segundos entre as repetições geradas pelo efeito. Essas repetições além de estarem igualmente defasadas, ainda sofrem um decaimento de amplitude linear produzido por este efeito. Vale a pena esclarecer que este decaimento linear de amplitude acaba soando como um decaimento exponencial para os nossos

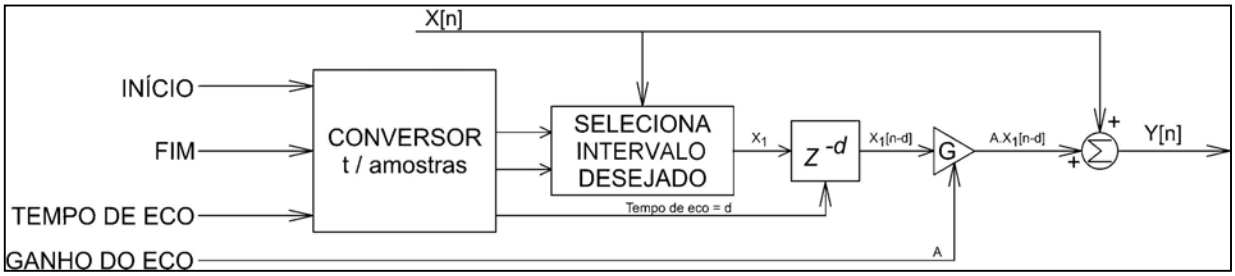


Figura 2 - Diagrama de Blocos do Efeito Eco.

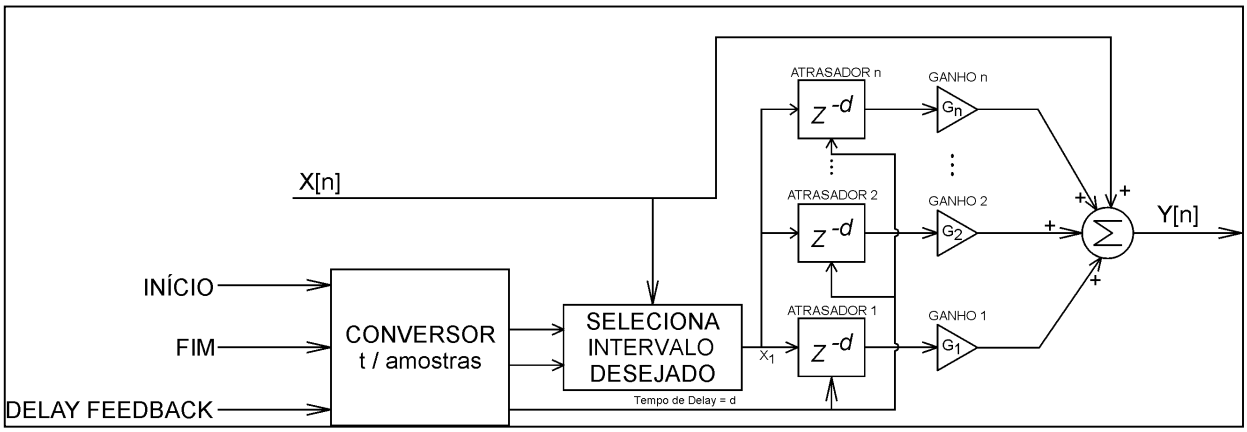


Figura 3 - Diagrama de Blocos do Efeito Delay.

ouvidos. Isto se explica pelo simples fato de que a audição humana é sensível a variações logarítmicas de amplitude sonora. Portanto para atingir um decaimento que soe "linear" para o ouvido humano, deveríamos aplicar um efeito que produzisse um decaimento exponencial de amplitude.

O efeito *delay* também pode ser facilmente percebido pelo ouvido humano caso o intervalo entre os atrasos seja razoável (maior que 0.1 segundos).

Assim como o eco, o *delay* também deve ser aplicado em instrumentos solo, pois pode-se obter um som não muito agradável aplicado-o em vários instrumentos de uma vez. Ao menos que objetivo seja obter uma sonoridade diferente, ou mesmo tornar o som confuso propositadamente.

Efeito Reverb

A implementação do efeito *reverb* se assemelha bastante com o *delay*. A diferença é que as amostras das repetições ímpares (q = 1, 3,5,7,...,n) são subtraídas do sinal original ao invés de somadas [3].

A Figura 4 apresenta o diagrama de blocos que descreve este efeito. O *Reverb* Decay *p* é a variável que irá definir o decaimento do *reverb* relacionando o número de repetições geradas. Os ganhos G_1, G_3, G_5, G_7 e todos os outros com índices ímpares assumirão valores negativos garantindo assim que as amostras das repetições ímpares serão subtraídas do sinal original ao invés de somadas.

O efeito *reverb* é utilizado para simular o que chamamos de "ambiência" de algumas salas ou ambientes com características de reverberação interessantes, como catedrais, cavernas e grandes salões. Ele pode ser usado em praticamente todos os instrumentos que compõem uma música, dando uma "encorpada" no som de cada um deles.

Neste trabalho, foi implementado um tipo de *reverb* bem simples. Profissionalmente, usam-se efeitos de *reverb* bem mais complexos, utilizando combinações do modelo simplificado para gerar as mais diversas reverberações encontradas em ambientes específicos.

EfeitoFlanging

O *flanging* produz um efeito parecido com o eco, porém o atraso $\beta(n)$ entre cada amostra varia em função dos valores inseridos para algumas variáveis de entrada [3].

O diagrama da Figura 5 representa as etapas referentes ao processamento deste efeito. O *Flange Delay d* é o tempo de

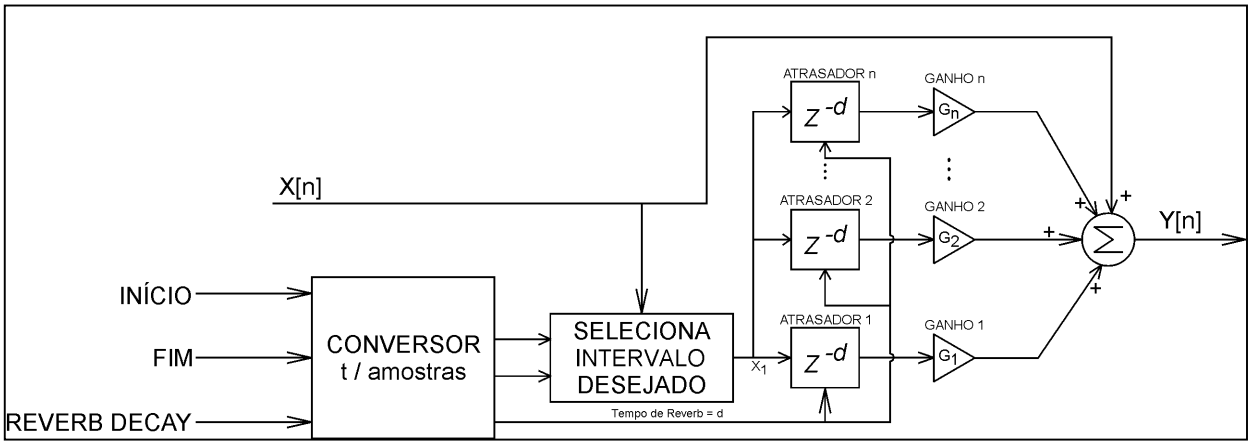


Figura 4 - Diagrama de Blocos do Efeito Reverb.

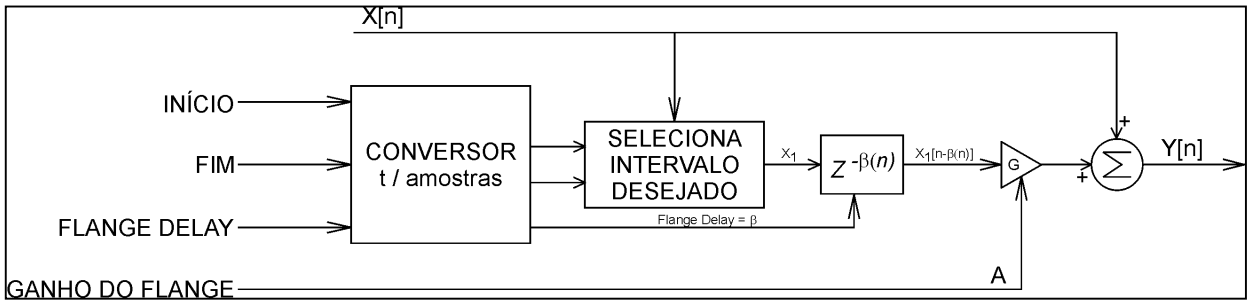


Figura 5 - Diagrama de Blocos do Efeito Flanging.

atraso máximo, em segundos, entre uma amostra do eco e sua amostra correspondente no sinal original.

O sinal x_1 sofrerá um atraso em relação ao sinal original, porém as amostras deste sinal selecionadas possuem um tempo de atraso entre elas variante. Esta variação é definida pela equação abaixo:

$$b = (d / 2) * (1 - \cos(w)) \quad (02)$$

Onde *b* é o atraso em amostras de uma amostra qualquer do sinal selecionado em relação a sua amostra correspondente no sinal original. Note que *b* assumirá no máximo valor igual a *d*. Temos ainda como variável de entrada a *Frequency Rate* (w), sendo esta uma frequência em radiano com que varia o atraso entre cada amostra do intervalo selecionado em relação ao sinal original.

Efeito Chorus

O *Chorus* produz um efeito parecido com o *delay*, porém o atraso entre cada intervalo varia em função dos valores inseridos para algumas variáveis de entrada. O diagrama da Figura 6 representa as etapas referentes ao processamento deste efeito. O *Chorus Delay d*, é o tempo de atraso máximo, em segundos, entre os intervalos [3]. Na prática, dependendo dos valores inseridos nos parâmetros de entrada, fica difícil diferenciar o som produzido por um efeito *Flanging* de um outro produzido por um efeito *Chorus*. Alguns instrumentos como violão, baixo e órgão, sobressaem-se melhor que outros instrumentos neste efeito.

Dependendo dos valores inseridos em alguns parâmetros, tais como, *Frequency Rate*, que determina a variação dos atrasos entre os intervalos, e o *Chorus Delay*, teremos resultados bastante estranhos que podem até descaracterizar o sinal com efeito incorporado.

EFEITOS NO DOMÍNIO DA FREQUÊNCIA

Falaremos agora um pouco sobre a implementação do efeito Equalizador. Como estamos trabalhando com um efeito que realiza modificações no espectro de frequência de um sinal de áudio digital, todas as características e alterações deste espectro só poderão ser observadas no domínio da frequência e não no domínio do tempo, como era percebido na implementação dos efeitos anteriores.

Efeito Equalizador

O objetivo é trabalhar com um equalizador de 32 bandas lineares, ou seja, igualmente espaçadas entre si. As bandas são como filtros passa-faixas que permitem a passagem de um som numa determinada faixa de frequência.

Para gerar as 32 bandas lineares utilizamos um filtro protótipo com 64 coeficientes. O número de coeficientes de um filtro determina a sua ordem, neste caso temos um filtro de ordem 63. Quanto maior a ordem de um filtro, melhor será a sua seleção de uma determinada faixa de frequência. A

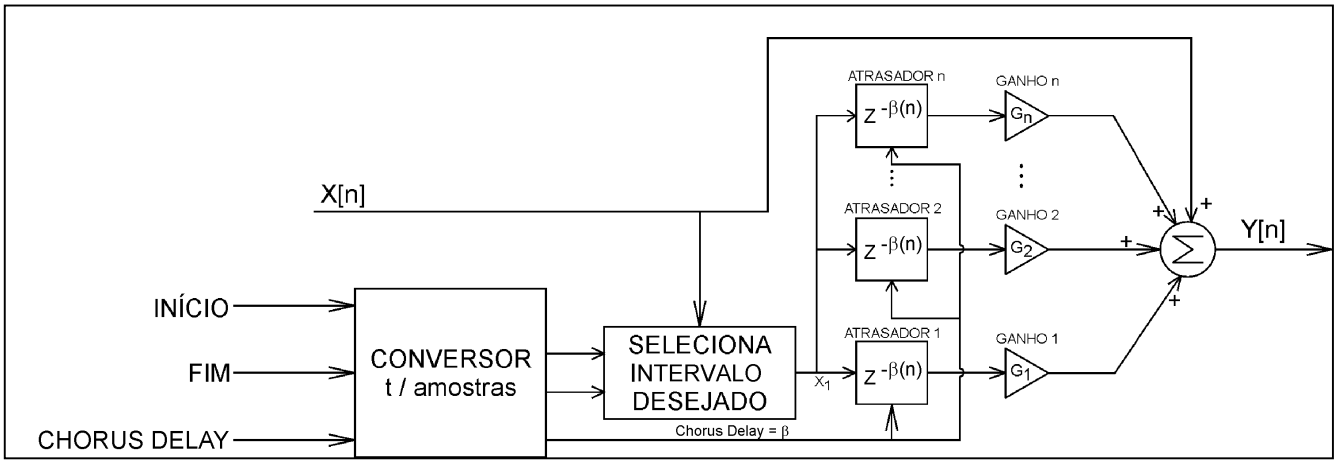


Figura 6 - Diagrama de Blocos do Efeito Chorus.

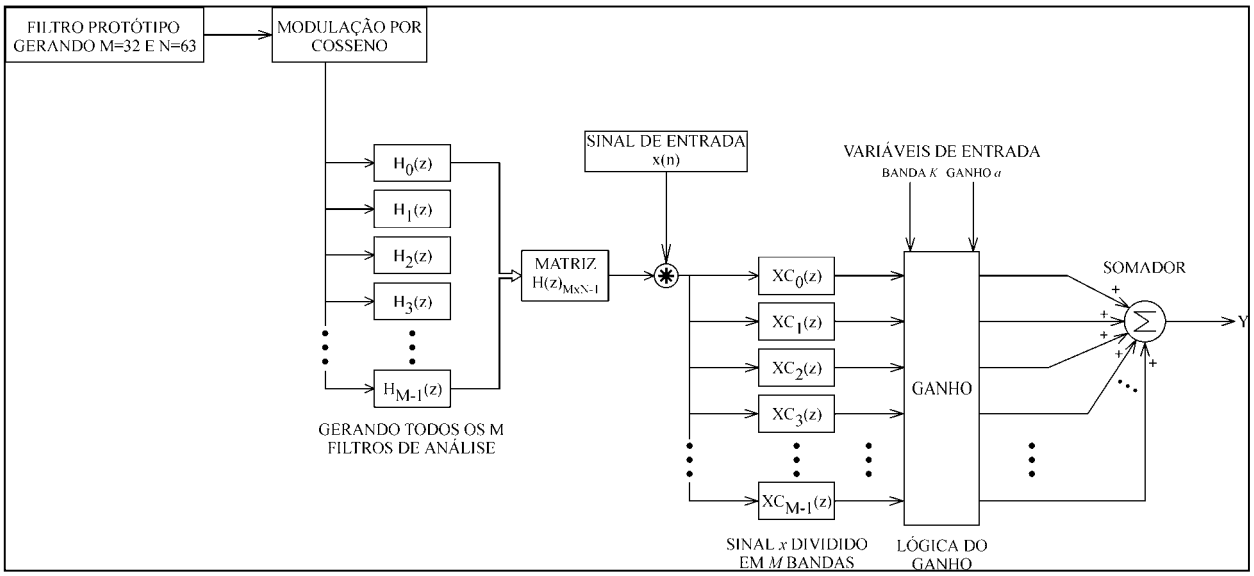


Figura 7 - Diagrama das Etapas da Lógica do Efeito Equalizador.

Figura 7 mostra um diagrama que descreve as etapas para a aplicação do efeito equalizador em um sinal *x* de entrada.

Para implementação do efeito Equalizador utilizamos a técnica do Banco de Filtros. Um Banco de filtros nada mais é do que uma sequência de filtros. Estes filtros, também chamados de Filtros de Análise, são convoluídos com o sinal de entrada *x* fazendo a divisão deste sinal em 32 bandas lineares de frequência [4]. O número de bandas do sinal resultante será sempre igual ao número de filtros de compõem o Banco de Filtros.

O Banco de Filtros é criado a partir da teoria da Modulação por Cosseno. Para simplificar este assunto, diremos que cada filtro de análise é criado aplicando a equação 3.

Onde *k* é o índice do filtro de análise, *n* é o número de canais para o sinal de áudio de entrada, *hp* é a resposta impulsiva do filtro protótipo utilizado, M é o número de bandas dos filtros e N a ordem dos filtros [5] [6].

A divisão do sinal de entrada em 32 bandas lineares, formam outros 32 sinais. Para cada sinal formado aplicamos um ganho independente criando assim um equalizador de 32 bandas de frequência.

Porém vale lembrar que o ouvido humano é sensível apenas às variações logarítmicas de frequência. Isto significa dizer que quando percebemos variações lineares de frequência, na verdade estamos variando a frequência logaritmicamente. Na prática, o intervalo entre cada banda é definido em oitavas. Transformaremos então as 32 bandas lineares em 6 bandas de oitavas.

Na implementação do efeito, esta transformação é feita ainda nos filtros de análise, pois assim necessitamos de menos filtros e consequentemente menos convoluções serão realizadas com o sinal de entrada, diminuindo o custo computacional no processamento. A transformação ocorre fazendo-se o somatório dos filtros lineares obtendo-se os filtros em oitavas. A equação 4 descreve de que forma este somatório é feito, com $0 \leq l \leq$ M-1 e $1 \leq o \leq$ C. Onde *l* é a banda linear e *o* sua banda correspondente em oitavas.

Temos também que $hh_o(n)$ é o Filtro em Oitavas correspondente ao Filtro Linear $h_l(n)$.

A Figura 8 mostra uma tabela que relaciona as bandas lineares, com as bandas em oitavas e suas respectivas larguras em frequência, que estaremos utilizando neste projeto.

Banda em Oitavas	Banda Linear	Faixa de Frequência
1	1	0-689 Hz
2	2	689-1378 Hz
3	3, 4	1,378 - 2,756 kHz
4	5,6,7,8	2,756 - 5,512 kHz
5	9,10, 11, 12, 13, 14, 15, 16	5,512 - 11,025 kHz
6	17, 18, 19, ... , 30, 31, 32	11,025 - 22050 kHz

Figura 8 –Tabela Relacionando Bandas, Oitavas e Frequências.

Sendo assim, sempre que o usuário deste programa quiser dar um ganho numa determinada banda (em oitavas), automaticamente o algoritmo encontrará qual ou quais serão as bandas correspondentes na escala linear.

Na Figura 9a é mostrada a resposta em frequência das bandas lineares, já a Figura 9b mostra a divisão destas bandas em oitavas.

Após a modificação do ganho das bandas selecionadas do sinal *x*, devemos agora recompor todas bandas somando os sinais XC_k e formar um único sinal de saída y.

$$H(k,n)=2hp(n)\cos\{[2(k-1)+1]*[\pi/(2M)][(n-1)-((N)/2)]+[(-1^{k-1})\pi/4]\} \quad (03)$$

$$hh_1(n) = h_0(n) \text{ e } hh_o(n) = \sum_l h_l(n),\ 2^{(o-2)} \leq l \leq 2^{(o-1)} -1 \quad (04)$$

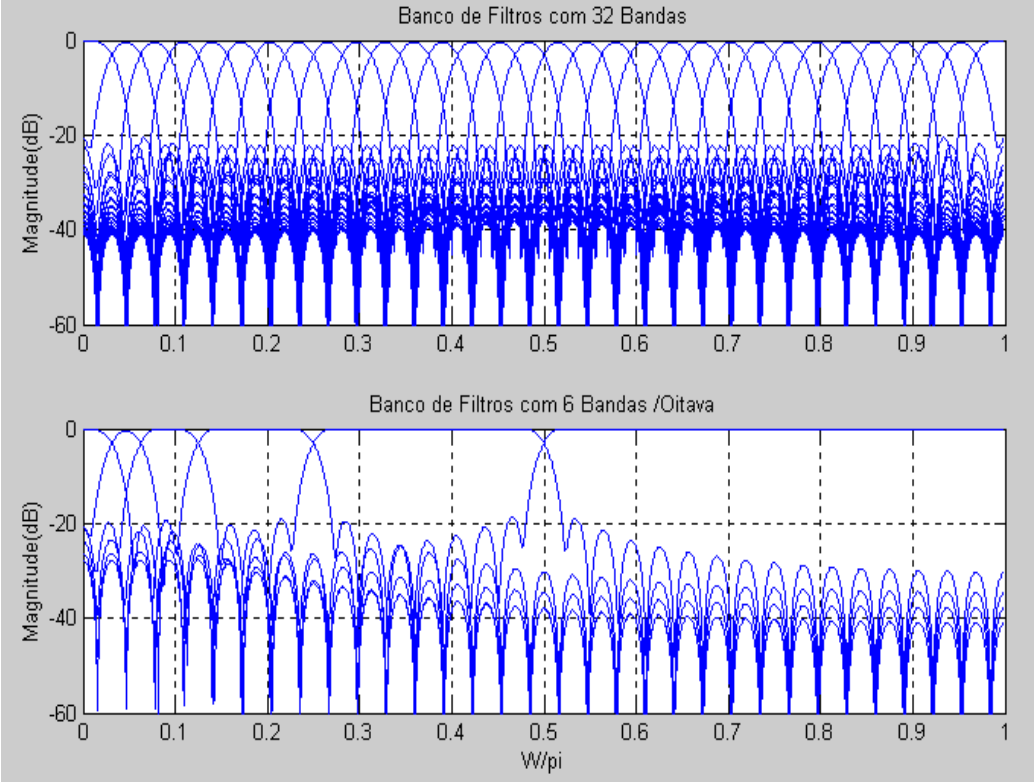


Figura 9 - ***(a) Resposta de Frequência das Bandas Lineares e (b) Separadas por Oitavas.***

CONCLUSÕES

Pudemos observar que todos os efeitos propostos neste trabalho obtiveram um bom resultado, se compararmos aos programas de processamento de áudio para PC existentes no mercado. Os processadores de efeitos profissionais de altíssima qualidade necessitam de melhor capacidade de processamento para suportar o custo computacional envolvido na implementação dos efeitos.

Vale mencionar ainda a necessidade de identificação de quando e onde cada efeito deve ser aplicado numa música, a fim de se evitar resultados indesejáveis no sinal processado. Normalmente os efeitos são aplicados de forma suave, a fim de não comprometer muito o som original.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] D. Christiansen, *Electronics Engineers' Handbook*, McGraw-Hill, Fourth Edition – IEEE Press, pp. 23.3-23.7; 23.13-23.15; 23.63-23.64; 23.80-23.87; 23.90-23.93.

[2] A. V. Oppenheim e A. S. Willsky, *Signals and Systems*, Prentice Hall International, Second Edition, p. 75; 112.

[3] S. K. Mitra, *Digital Signal Processing – A Computer-Based Approach*, McGraw-Hill, pp. 763-769, 1998.

[4] B. P. Lathi, *Modern Digital and Analog Communication Systems*, IEEE Press, Third Edition, pp. 97-98, 1997

[5] G. Strang e T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, pp. 1-2; 7-8; 15-17; 301-303; 325-331.

[6] T. Q. Nguyen, *Digital Filter Banks Desing – Quadratic Constrained Formulation*, IEEE Trans. Signal Processing, vol 43, pp. 2103-2108, 1994.

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na VII Convenção Nacional
26-28 de maio de 2003, São Paulo, 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.

Considerações Sobre o Uso de Funções de Transferência Clássicas no Projeto de Divisores de Freqüência

André Luís Dalcastagnê[1], Sidnei Noceti Filho[1] e Homero Sette Silva[2]
1 – LINSE – Laboratório de Circuitos e Processamento de Sinais
Departamento de Engenharia Elétrica, UFSC
Tel: (0xx48) 331-9504, Fax: (0xx48) 331-9091
Florianópolis, 88040-900, Santa Catarina, Brasil
andre@linse.ufsc.br, sidnei@linse.ufsc.br
2 - Eletrônica Selenium S.A.
Nova Santa Rita, 92001-970, Rio Grande do Sul, Brasil
homero@selenium.com.br

RESUMO

Neste trabalho, o uso de funções de aproximação clássicas no projeto de divisores de freqüência será abordado sob dois aspectos. Primeiramente, irá se mostrar que a utilização da função Bessel no projeto de um divisor de freqüência devido apenas à sua boa característica de fase não é coerente, uma vez que essa propriedade só ocorre em filtros passa-baixa. Em seguida, mostra-se que as características dos alto-falantes reais impedem que a resposta em freqüência de um divisor de freqüência possua as características das funções clássicas.

INTRODUÇÃO

A banda de áudio, normalmente considerada como sendo a faixa compreendida entre 20 e 20000 Hz, não pode ser reproduzida com qualidade por apenas um alto-falante devido à sua grande extensão. A solução para esse problema é separar o espectro de freqüência do sinal elétrico a ser convertido em pressão sonora em duas ou mais bandas e destinar um alto-falante específico para reproduzir cada uma dessas faixas de freqüência.

Realizar esta divisão é a principal função de um divisor de freqüência (DF), ou *crossover*, que é um conjunto de filtros elétricos responsável por determinar o espectro de freqüência do sinal elétrico enviado a cada transdutor do sistema. Neste trabalho, sistema é o conjunto formado pelo DF e pelos alto-falantes.

Normalmente, o projeto de um DF é baseado em uma função de aproximação clássica, sendo que as mais utilizadas são as funções Butterworth [1], Bessel [1] e Linkwitz-Riley [2], formada pela cascata de duas funções Butterworth idênticas. Essas funções possuem características particulares: a função Butterworth é monotônica em toda a faixa de freqüência e apresenta a máxima planura possível na banda de passagem; a função Bessel causa o menor desvio dos atrasos de fase

e grupo na banda passante; a função Linkwitz-Riley gera DFs de duas vias com resposta em freqüência plana, sem sobrepassamento em torno da freqüência de cruzamento. Entretanto, deve-se ter em mente que em um DF real essas características nunca são preservadas.

O primeiro objetivo deste trabalho é mostrar que não se deve utilizar como único critério de escolha da função Bessel para o projeto de um DF a sua boa característica de fase na banda passante, quando se compara com as demais funções clássicas, pois tal propriedade só existe em filtros do tipo passa-baixa. Na segunda parte, mostra-se que como o alto-falante não é uma simples resistência, os filtros de um DF passivo nunca são Butterworth, Bessel ou Linkwitz-Riley, pois a carga de um filtro passivo influencia na sua função de transferência (FT). Além disso, tanto nos DFs passivos quanto nos ativos, é interessante alterar os valores dos componentes calculados através de fórmulas clássicas, que consideram o alto-falante uma resistência pura, de modo a compensar os efeitos das não-idealidades dos alto-falantes na FT do sistema. Tal tarefa pode ser feita tanto por um método empírico quanto por otimização.

MODELAMENTO DE UM DIVISOR DE FREQÜÊNCIA

Grandezas Associadas a um Alto-Falante

Neste trabalho, duas grandezas associadas a um alto-falante serão mencionadas: a sua impedância equivalente e a sua FT. Considere a Fig. 1, onde "*vc*" significa *voice coil*, ou seja, a bobina de voz.

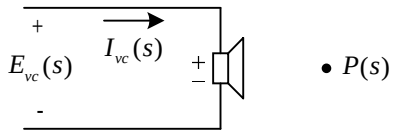


Fig. 1 - Grandezas associadas a um alto-falante.

A impedância equivalente de um alto-falante é definida como a relação entre a tensão aplicada nos seus terminais e a corrente que circula através deles:

$$Z_{vc}(s) = E_{vc}(s) / I_{vc}(s) \tag{1}$$

A FT de um alto-falante é a relação entre a pressão sonora produzida em um ponto do espaço e a tensão aplicada nos seus terminais:

$$T(s) = P(s) / E_{vc}(s) \tag{2}$$

Divisores de Freqüência Ativos e Passivos

O sistema formado pelo DF e pelos alto-falantes pode ser representado pelo diagrama de blocos da Fig. 2, onde na entrada se tem o sinal elétrico e na saída, a pressão sonora total produzida em um ponto do espaço. Idealmente, um DF deve fazer com que a pressão sonora na saída seja análoga ao sinal elétrico de entrada, independentemente da freqüência.

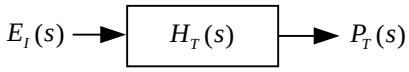


Fig. 2 – Sistema formado pelo divisor de freqüência e alto-falantes.

Este sistema pode ser construído com um DF passivo ou ativo. A Fig. 3 mostra a forma de ligação de um DF passivo e a Fig. 4, a de um DF ativo. Por simplicidade, representou-se DFs do tipo paralelo de duas vias. O índice "1" representa a via passa-baixa, o índice "2", a via passa-alta e o índice "*T*", o sistema completo.

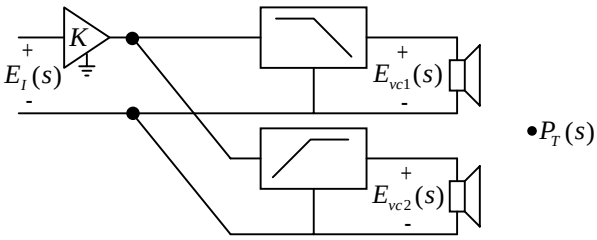


Fig. 3 – Divisor de freqüência passivo.

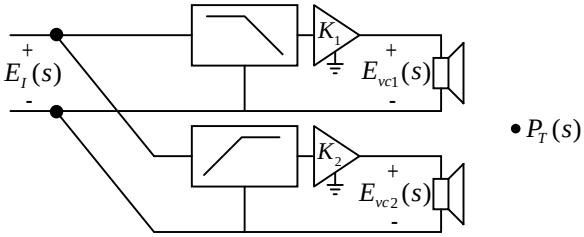


Fig. 4 – Divisor de freqüência ativo.

Idealmente, ambos os tipos podem realizar a mesma FT. Entretanto, há uma diferença básica na posição em que podem ser dispostos em relação ao amplificador de potência. Por simplicidade de projeto e disponibilidade no mercado, a grande maioria dos DFs ativos são construídos com amplificadores operacionais convencionais, que são incapazes de gerar a potência necessária para alimentar um alto-falante diretamente. Por esse motivo, os DFs ativos devem ser montados antes do amplificador de potência que alimenta o alto-falante (Fig. 4). Conseqüentemente, deve-se utilizar um amplificador de potência para cada via do sistema. No caso de DFs passivos, o problema de potência não existe e eles podem ser montados antes ou depois do amplificador. Se colocados antes (Fig. 4), há a vantagem de que a carga de cada um dos filtros do DF é a alta impedância de entrada do amplificador de potência. Se colocados entre o amplificador de potência e os transdutores (Fig. 3), a carga de cada filtro do DF é a impedância equivalente do transdutor que o filtro está alimentando. Esse fato complica bastante o projeto do DF, já que a carga de um circuito passivo altera a sua FT. Apesar dessa desvantagem, essa montagem é a mais utilizada e será considerada como padrão neste trabalho. Essa preferência se deve ao fato de que quando montado dessa forma, o DF pode ser incorporado dentro da caixa acústica,

juntamente com os seus alto-falantes, além de ser necessário apenas um amplificador de potência para todo o sistema.

Equacionamento de um Divisor de Freqüência

Nesta seção, será apresentado o equacionamento de um DF. A FT do amplificador de potência, ou dos amplificadores de potência no caso de DF ativo, será desconsiderada, pois na faixa de operação dos DFs pode-se assumir que o amplificador de potência está atuando na sua faixa plana.

A pressão sonora total produzida pelo sistema é a soma das pressões sonoras produzidas por cada uma de suas vias. Portanto:

$$P_T(s) = P_1(s) + P_2(s) \tag{3}$$

As FTs das vias *k* são independentes e dadas por:

$$H_k(s) = P_k(s) / E_I(s) \tag{4}$$

Combinando as equações (3) e (4), pode-se escrever a pressão sonora produzida pelo sistema como:

$$P_T(s) = H_1(s) \times E_I(s) + H_2(s) \times E_I(s) \tag{5}$$

de onde determina-se que:

$$H_T(s) = P_T(s) / E_I(s) = H_1(s) + H_2(s) \tag{6}$$

Portanto, fica evidenciado que o projeto de um DF paralelo pode ser feito via a via, pois a FT do sistema completo é a soma das FTs de cada uma de suas vias.

Conforme se observa na Fig. 3, a FT de uma via pode ser dividida em dois termos:

$$H_k(s) = \frac{P_k(s)}{E_I(s)} = \frac{E_{vck}(s)}{E_I(s)} \times \frac{P_k(s)}{E_{vck}(s)} \tag{7}$$

O termo:

$$F_k(s) = E_{vck}(s) / E_I(s) \tag{8}$$

é a FT do filtro da via. O termo:

$$T_k(s) = P_k(s) / E_{vck}(s) \tag{9}$$

é a FT do transdutor. Assim, cada via é formada por dois sistemas independentes em cascata, cuja FT é:

$$H_k(s) = F_k(s) \times T_k(s) \tag{10}$$

Portanto, iremos abordar neste trabalho três tipos de FT: a FT do sistema $H_T(s)$, a FT da via *k* $H_k(s)$ e a FT do filtro da via *k* $F_k(s)$.

EFEITO DAS TRANFORMAÇÕES EM FREQÜÊNCIA NA FASE DE FILTROS BESSEL

A função Bessel é uma aproximação de ordem *n* da função de fase linear, que é dada pela equação (11):

$$L(s) = K\, e^{-T_0 S} \tag{11}$$

Pode-se facilmente notar que esta função possui uma fase linear para $s = j\omega$, que é $-\omega T_0$.

Um parâmetro comumente utilizado para averiguar o comportamento da fase $\theta(\omega)$ de uma função é o atraso de fase [1], que é definido como:

$$\tau_p(\omega) = -\frac{\theta(\omega)}{\omega} \tag{12}$$

Portanto, o atraso de fase da função de fase linear é constante e igual a T_0.

O fato de a função Bessel ser uma aproximação da função de fase linear lhe confere algumas particularidades. A função Bessel é a que apresenta a menor dispersão do atraso de fase na banda passante dentre todos os filtros polinomiais clássicos, como Butterworth ou Chebyshev, o que significa que é a aproximação clássica cuja fase mais se aproxima de uma reta na banda passante. Além disso, ao contrário dos outros tipos de aproximação, quanto maior for a ordem *n* do filtro Bessel, melhor será a aproximação da equação (11) e, portanto, maior a faixa de freqüência em que a sua fase se mantém aproximadamente linear.

Entretanto, a boa característica de fase na banda passante da função Bessel passa-baixa não é preservada quando se aplica uma transformação em freqüência, para se projetar filtros Bessel passa-alta, por exemplo.

Para demonstrar este problema, iremos comparar um filtro Bessel com um Butterworth, ambos de quarta ordem. A variável *s* será normalizada pela freqüência de cruzamento do sistema, conforme a equação (13). As funções passa-baixa Bessel e Butterworth são dadas pelas equações (14) e (15), respectivamente. Os atrasos de fase destas duas funções estão mostrados na Fig. 5.

$$\bar{s} = \frac{s}{\omega_c} \tag{13}$$

$$H_{1BT}(\bar{s}) = \frac{1}{\bar{s}^4 + 2,613\bar{s}^3 + 3,41\bar{s}^2 + 2,613\bar{s} + 1} \tag{14}$$

$$H_{1BS}(\bar{s}) = \frac{1}{\bar{s}^4 + 3,123\bar{s}^3 + 4,39\bar{s}^2 + 3,201\bar{s} + 1} \tag{15}$$

Pode-se notar que a dispersão do atraso de fase na banda passante da função Bessel é bem menor que a da função Butterworth. Esse fato comprova que a fase da função Bessel se aproxima muito de uma reta na banda passante, diferentemente da função Butterworth.

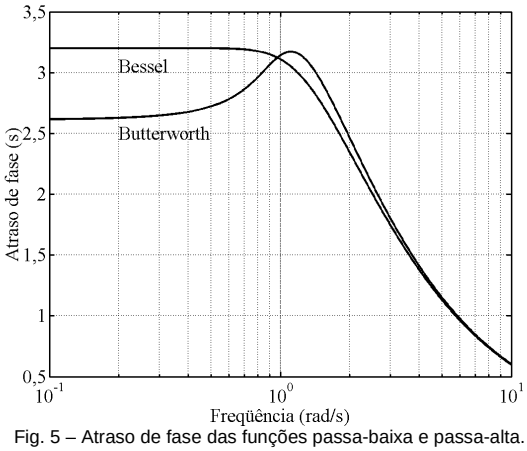


Fig. 5 – Atraso de fase das funções passa-baixa e passa-alta.

Aplicando a transformação passa-baixa passa-alta [1] nas equações (14) e (15), determinam-se as funções passa-alta de quarta ordem Butterworth e Bessel, dadas pelas equações (16) e (17), respectivamente.

$$H_{2BT}(\overline{s}) = \frac{\overline{s}^4}{\overline{s}^4 + 2,613\overline{s}^3 + 3,41\overline{s}^2 + 2,613\overline{s} + 1} \quad (16)$$

$$H_{2BS}(\overline{s}) = \frac{\overline{s}^4}{\overline{s}^4 + 3,123\overline{s}^3 + 4,39\overline{s}^2 + 3,201\overline{s} + 1} \quad (17)$$

As curvas de atraso de fase destas duas funções são iguais às das respectivas funções passa-baixa (Fig. 5). Em um filtro passa-alta, a banda passante inicia na freqüência de 1 rad/s. Nessa região, a função Bessel não possui mais um atraso de fase praticamente constante e, conseqüentemente, a sua fase não é mais próxima de uma reta. A vantagem da função Bessel sobre a Butterworth desapareceu, pois ambas possuem características de fase ruins. Sendo assim, fica provado que a vantagem de fase aproximadamente linear na banda passante do filtro Bessel sobre o Butterworth, ou sobre qualquer outra aproximação clássica, existe apenas para o caso de filtros passa-baixa.

Portanto, escolher a função Bessel para o projeto de um DF baseando-se apenas na sua boa característica de fase não é coerente, pois um DF é composto por, no mínimo, um filtro passa-baixa e um passa-alta. Deve-se verificar todas as características (magnitude, fase e tempo) antes de se optar pela função a ser utilizada.

PROJETO DE DIVISORES DE FREQÜÊNCIA

O uso de FTs clássicas no projeto de DFs seria perfeito se o alto-falante fosse um elemento ideal, ou seja, com impedância equivalente puramente resistiva e resposta em freqüência com magnitude constante e fase linear. Porém, sabe-se que a impedância equivalente e a resposta em freqüência de um alto-falante são funções complexas e que interferem na FT do sistema. Essa interferência é diferente em DFs passivos e ativos. No caso de um DF passivo, o alto-falante influencia de duas maneiras na FT de uma via: através de $T(s)$ e da sua impedância equivalente, que age sobre $F(s)$. Já no caso de um DF ativo, apenas a FT do transdutor age na FT da via, já que a FT do filtro é independente da impedância equivalente do transdutor.

Projeto de Divisores de Freqüência Passivos

Um DF passivo é formado unicamente por resistores, indutores e capacitores. Como exemplo de projeto, considere o DF passivo de segunda ordem mostrado na Fig. 6. O resistor em série com o transdutor da via passa-alta serve para compensar a maior sensibilidade, em geral, do alto-falante da via passa-alta em relação à do alto-falante da via passa-baixa.

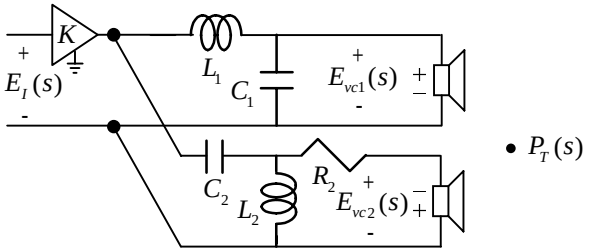


Fig. 6 – Divisor de freqüência passivo de segunda ordem.

A FT do filtro passa-baixa é:

$$F_1(s) = \frac{E_{vc1}(s)}{E_I(s)} = \frac{1/L_1C_1}{s^2 + \frac{1}{Z_{vc1}(s)C_1}s + \frac{1}{L_1C_1}} \quad (18)$$

A FT do filtro passa-alta é:

$$F_2(s) = \frac{E_{vc2}(s)}{E_I(s)} = \frac{G \cdot s^2}{s^2 + \frac{1}{(R_2 + Z_{vc2}(s))C_2}s + \frac{1}{L_2C_2}} \quad (19)$$

onde:

$$G = \frac{Z_{vc2}(s)}{R_2 + Z_{vc2}(s)} \quad (20)$$

O projeto do DF consiste em achar os valores das variáveis de (18) e (19) que satisfaçam a FT desejada para o circuito. Por exemplo, se quisermos projetar um DF do tipo Butterworth de segunda ordem, com $\omega_c = 2\pi 800$ rad/s , as FTs desejadas serão:

$$F_1(s) = \frac{2,5266 \times 10^7}{s^2 + 7108,6s + 2,5266 \times 10^7} \quad (21)$$

$$F_2(s) = G\frac{s^2}{s^2 + 7108,6s + 2,5266 \times 10^7} \quad (22)$$

Igualando as equações (18) e (21), e considerando a impedância equivalente do transdutor puramente resistiva e igual a 8 Ω, calculam-se os seguintes valores para os componentes da via passa-baixa: L_1 = 2,3 mH e C_1 = 17,584 μF. O procedimento deve ser repetido para a via passa-alta, sabendo-se qual o valor da atenuação *G* necessária para o circuito. As fórmulas para cálculo de DFs passivos, considerando transdutores com impedância equivalente puramente resistiva, podem ser facilmente encontradas na literatura, como em [3].

Entretanto, sabe-se que a impedância equivalente do transdutor não é puramente resistiva. Como esse parâmetro entra no cálculo das FTs dos filtros, conforme mostram as equações (18) e (19), os filtros calculados, na realidade, não são do tipo Butterworth.

Outro ponto importante é que as FTs das vias são dadas pelo produto entre as FTs dos seus filtros e FTs dos respectivos transdutores (equação (10)). Assim, mesmo se as FTs dos filtros fossem Butterworth, as FTs das vias não seriam Butterworth. Como a FT do sistema é a soma das FTs das vias (equação (6)), o sistema também não é Butterworth.

Projeto de Divisores de Freqüência Ativos

No caso de DFs ativos, a impedância equivalente do transdutor não interfere nas FTs dos filtros. Portanto, ao contrário do caso passivo, é possível projetar DFs ativos cujos filtros possuam uma $F(s)$ clássica.

O procedimento de projeto é o mesmo do caso passivo: basta deduzir as FTs dos filtros, compará-las com as FTs desejadas e calcular os valores dos componentes dos filtros.

Entretanto, as FTs dos transdutores continuam a interferir na FT do sistema completo, equação (10). Sendo assim, apesar de as $F(s)$ serem ideais, as FTs das vias $H(s)$ não serão, devido à ação das FTs dos transdutores $T(s)$. Como conseqüência, a FT do sistema $H_T(s)$ não será igual a função desejada.

Projeto de Divisores de Freqüência Considerando as Características Reais dos Alto-Falantes

No projeto de um DF, o que é realmente importante são as FTs das vias e não as FTs dos filtros, pois se está interessado em obter um bom resultado acústico e não propriamente que a tensão aplicada nos transdutores tenha o comportamento ideal. Como as FTs das vias são dadas pelo produto entre $F(s)$ e $T(s)$ (equação (10)), pode-se modificar $F(s)$ de modo que a resposta em freqüência da via $H(s)$ fique mais próxima da desejada.

Modificar $F(s)$ significa alterar os valores dos componentes do filtro. Pode-se partir dos valores calculados via fórmulas clássicas e modificá-los empiricamente até se obter um resultado satisfatório. Outra opção é utilizar métodos numéricos, que são a base dos programas de otimização de DFs.

Como exemplo deste segundo método, projetou-se um DF passivo de duas vias, segunda ordem, do tipo Butterworth, utilizando o programa DivCalc, apresentado em [4]. O DF projetado está mostrado na Fig. 7. Esse projeto foi feito com base nas curvas medidas de impedância equivalente e resposta em freqüência dos transdutores. Pode-se notar a diferença entre os valores calculados considerando alto-falantes com impedância equivalente de 8 Ω.

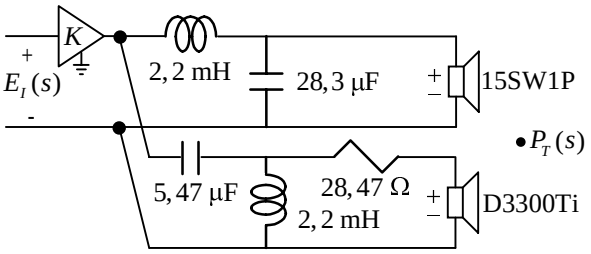


Fig. 7 – Divisor de freqüência passivo de segunda ordem otimizado.

A Fig. 8 mostra as curvas de *SPL* real e ideal. O *SPL* produzido pelas vias na banda passante é bem próximo do ideal, o que não ocorreu na banda de rejeição. Entretanto, é interessante que o *SPL* produzido pela via na banda de rejeição tenha uma inclinação maior do que a da função desejada, porque quanto menor a influência da via fora da sua faixa de operação, mais plano será o *SPL* produzido pelo sistema.

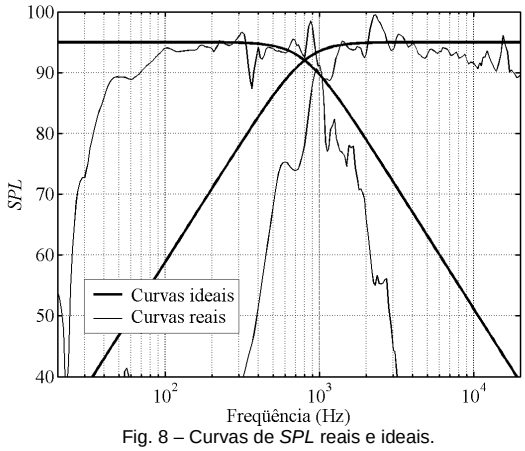


Fig. 8 – Curvas de *SPL* reais e ideais.

Na Fig. 9, mostra-se as curvas ideais e reais do *SPL* produzido pelo sistema. Um fato importante deve ser ressaltado. Idealmente, os transdutores de um DF de duas vias segunda ordem devem ser ligados fora de fase para evitar o ponto nulo da resposta na freqüência de cruzamento (Fig. 6) [3,4]. Porém, por simulação, foi verificado que com os transdutores conectados em fase, esse projeto apresenta um melhor resultado (*SPL* mais plano). Por esse motivo, os transdutores foram ligados em fase no DF protótipo (Fig. 7). Esse fato ocorreu porque tanto a resposta em freqüência quanto a impedância equivalente de um alto-falante são muito diferentes do considerado como ideal. Portanto, no projeto de um DF, todas as possibilidades de ligação devem ser testadas, para poder se optar pela que apresente o melhor resultado.

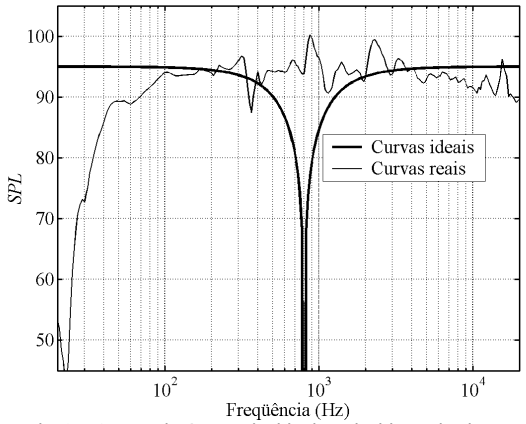


Fig. 9 – Curvas de *SPL* real e ideal produzidas pelo sistema.

Na Fig. 10, mostram-se as curvas de fase da pressão sonora real e ideal produzida pela via passa-baixa. Nota-se a grande diferença na fase da pressão sonora produzida em relação à da função Butterworth, que vai de zero a –180°. A fase iria decair abaixo de -4000°, o que só não ocorreu porque acima de 3 kHz, o *SPL* produzido já é muito baixo, menos de 30 dB, o que gerou erros de medida. O fato de a fase atingir tal valor indica que apesar de o filtro utilizado ser de segunda ordem, a FT da via possui uma ordem muito maior, devido ao alto-falante. Somente pela fase não se pode tirar conclusões a respeito da ordem da via, porque se a FT do transdutor possui zeros no semiplano lateral esquerdo, a fase decai mais, e se possui zeros no semiplano lateral direito, a fase decai menos. Mas as ondulações na curva de *SPL* (Fig. 8) evidenciam a alta ordem do sistema, muito maior do que a que seria obtida caso a ordem do alto-falante fosse aquela dada pelos modelos T-S [5]. Portanto, é impossível corrigir todas as imperfeições da resposta em freqüência de um alto-falante utilizando apenas um filtro de baixa ordem em série. Apesar disso, os DFs realizam a sua tarefa de permitir que alto-falantes que reproduzem faixas de freqüência diferentes operem no mesmo sistema.

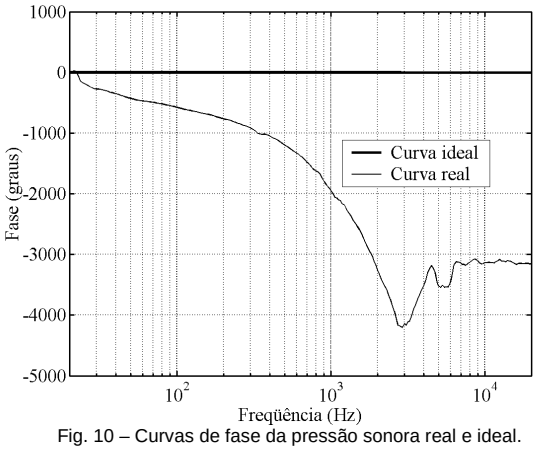


Fig. 10 – Curvas de fase da pressão sonora real e ideal.

CONCLUSÕES

Neste trabalho, o uso de FTs clássicas no projeto de DFs foi discutido. Mostrou-se que a boa característica de fase da função Bessel não deve ser o único aspecto a ser levado em consideração na sua escolha para o projeto de um DF, pois essa característica só ocorre em filtros passa-baixa. Em seguida, foi verificado que as características de impedância equivalente e resposta em freqüência dos transdutores reais impedem que um DF possua uma FT ideal, pois é impossível corrigir todas as imperfeições produzidas por um elemento de alta ordem como um alto-falante utilizando apenas um filtro de baixa ordem. Ainda assim, os DFs se prestam à função que desempenham: permitir que alto-falantes que reproduzem faixas de freqüência diferentes operem no mesmo sistema.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] NOCETI FILHO, S. **Filtros Seletores de Sinais**. 1. ed. Florianópolis: Editora da UFSC, 1998.

[2] LINKWITZ, S. H. Active Crossover Networks for Noncoincident Drivers. **Journal of the Audio Engineering Society**, v. 24, n. 1, p. 2-8, Jan./Feb. 1976.

[3] DICKASON, V. **Caixas Acústicas e Alto-falantes**. 5. ed. Rio de Janeiro: H. Sheldon Serviços de Marketing Ltda., 1997.

[4] DALCASTAGNÊ, A. L. **Desenvolvimento de um Programa para Projeto de Divisores de Freqüência Passivos Baseado nas Curvas de Resposta e Impedância dos Transdutores**. Florianópolis, 2002. 108 f. Dissertação (Mestrado em Engenharia Elétrica) – Centro Tecnológico, Universidade Federal de Santa Catarina.

[5] SILVA, H. S. **Análise e Síntese de Alto-falantes & Caixas Acústicas pelo Método T-S**. 1. ed. Rio de Janeiro: H. Sheldon Serviços de Marketing Ltda., 1996.

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na VII Convenção Nacional
26-28 de maio de 2003, São Paulo, 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.

Equalização de Áudio: Considerações Relevantes

Phabio Junckes Setubal, Sidnei Noceti Filho e Rui Seara
LINSE – Laboratório de Circuitos e Processamento de Sinais
Departamento de Engenharia Elétrica, UFSC
Tel: (0xx48) 331-9504, Fax: (0xx48) 331-9091
Florianópolis, 88040-900, Santa Catarina, Brasil
phabio@linse.ufsc.br, sidnei@linse.ufsc.br, seara@linse.ufsc.br

RESUMO

Equalizadores de áudio-freqüência são de grande importância na reprodução de sinais em sistemas de áudio e som. Assim, para um melhor entendimento do papel que os equalizadores desempenham em tais sistemas, este artigo discute algumas características básicas inerentes a todos os equalizadores, independente da tecnologia usada em seu projeto.

1. INTRODUÇÃO

Se existissem condições perfeitas para a reprodução dos sons, ou seja, se todo o sistema envolvido (fonte sonora + amplificadores + caixa acústica + ambiente + resposta do ouvido humano) tivesse uma resposta em freqüência perfeitamente plana, talvez os equalizadores de áudio (EQAs) não tivessem razão para existir. Entretanto, sabe-se que o sistema real está bem distante das condições ideais desejadas, justificando a necessidade de tais sistemas de equalização.

Considerando que a faixa de freqüência audível para os seres humanos localiza-se, aproximadamente, entre 20 Hz e 20 kHz, um EQA fornece a possibilidade de alterar as características de magnitude da resposta em freqüência do sistema na referida banda (Fig. 1), minimizando perdas auditivas dos ouvintes [1], ou eliminando realimentações acústicas do ambiente, tornando a resposta do sistema a mais plana possível. Além do mais, os EQAs são também utilizados para ajustar a distribuição de freqüências dos sinais audíveis ao nosso gosto pessoal, uma vez que a resposta em freqüência do nosso ouvido não é plana e não é a mesma para cada ouvinte em particular. Outra aplicação interessante dos EQAs é na produção de efeitos sonoros especiais.

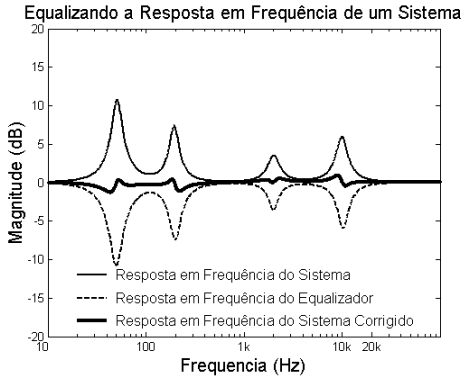


Fig. 1: Corrigindo a resposta em freqüência do ambiente.

Este artigo é organizado como segue. A Seção 2 apresenta um breve histórico sobre o surgimento e a evolução dos EQAs. A Seção 3 classifica os EQAs em relação aos seus parâmetros de controle. A Seção 4 apresenta a função de transferência (FT) do equalizador tipo *bump*, a qual é a base para o controle de amplitude dos EQAs gráficos,

paramétricos e paragráficos. A Seção 5 discute o efeito da combinação, ou interpolação, das amplitudes de freqüências vizinhas, comparando topologias com *Q*-constante e *Q*-não constante (proporcional), características inerentes aos EQAs gráficos. A Seção 6 estuda o efeito da largura de banda em EQAs gráficos de 1/3 oitava, considerando as configurações série e paralela. Uma breve discussão sobre o comportamento da fase, a qual apresenta característica de fase mínima, é discutida na Seção 7. Finalmente, na Seção 8 são apresentadas as conclusões deste artigo.

2. BREVE HISTÓRICO SOBRE EQUALIZAÇÃO DE ÁUDIO

Em um passado já distante, nos primórdios da telefonia, quando longos comprimentos de cabos eram empregados para transmitir a voz em grandes distâncias, verificou-se que uma importante porção do sinal era atenuado ao longo do caminho. Em princípio, imaginava-se que uma simples amplificação resolveria tal problema. Entretanto, observou-se que a atenuação do sinal dependia da freqüência, levando alguns componentes do espectro a sofrer maior degradação do que outros. Como solução, desenvolveram-se circuitos que amplificavam o sinal de voz de forma diferenciada, buscando preservar um mesmo nível de amplitude para todas as freqüências do sinal. Tais sistemas foram denominados equalizadores de sinais de voz.

A primeira vez que se utilizou um equalizador com características ajustáveis, objetivando melhorar a qualidade de sinais de áudio, foi nos anos 30 por John Volkman [2]. Tal aplicação visava equalizar o sistema de som de um cinema, que na época exibia os primeiros filmes que incorporavam áudio a imagens, mas cuja qualidade de reprodução era muito pobre. Ao contrário dos equalizadores empregados em redes telefônicas, que uma vez projetados não permitiam fáceis ajustes em suas características, ficando incorporado a uma rede telefônica específica, o equalizador desenvolvido por Volkman era um equipamento flexível, o qual permitia ser instalado a qualquer sistema de som já existente, possuindo ajustes relativamente fáceis. Neste mesmo período, a indústria cinematográfica de Hollywood começou a desenvolver outros sistemas de equalização ajustáveis, visando seu uso em pós-produção de efeitos sonoros, como também para realçar diálogos em alguns de seus filmes.

Durante os anos 40 e 50, devido à 2ª Guerra Mundial e suas conseqüências, não surgiu muita coisa nova nesta área. A maioria das aplicações dos EQAs restringia-se à pós-produção para a industria cinematográfica. Porém, em 1958, o professor W. Rudmose aplicou, com sucesso, na sonorização do Aeroporto de Dallas (Love Field) novas técnicas de equalização acústica. Em 1962, professor C. P. Boner da Universidade do Texas (Austin), muito conhecido por seus trabalhos em equalização acústica, desenvolveu uma simples e importante teoria. Essa teoria baseia-se na hipótese de que quando ocorre uma realimentação acústica esta acontece em uma determinada freqüência. Para eliminar tal realimentação, é suficiente atenuar esse componente de freqüência via um filtro rejeita-faixa sintonizado naquela freqüência. Assim, com o desenvolvimento desses filtros, o professor Boner fundamentou sua teoria para realimentação acústica. Desde estão, houve um grande desenvolvimento dos EQAs, podendo-se citar: os EQAs gráficos de 1/3 oitava, EQAs gráficos de *Q*-constante (largura de banda constante), EQAs paramétricos, dentre outras estruturas de equalização, usando tecnologias analógica ou digital.

3. CLASSIFICAÇÃO DOS PARÂMETROS DE CONTROLE DOS EQUALIZADORES DE ÁUDIO

Controle de Tonalidade

É um tipo de EQA muito simples e bastante utilizado. Geralmente opera em duas bandas: baixas freqüências (graves) e altas freqüências (agudos). É também conhecido como equalizador *shelving* [3], devido à sua curva característica, a qual apresenta resposta plana em baixas ou altas freqüências, ao contrário dos demais EQAs, que apresentam picos na freqüência de interesse (*bump*). Sua FT para baixas frequências é dada por

$$T(s) = \frac{s+A}{s+B}, \tag{1}$$

onde *A* e *B* são os parâmetros do sistema de equalização: para $A > B$, tem-se amplificação e para $A < B$, atenuação.

Para altas freqüências, sua FT é dada por (2). Assim,

$$T(s) = \frac{Cs+1}{Ds+1}, \tag{2}$$

sendo agora seus parâmetros representados por *C* e *D*. Para $C > D$, tem-se amplificação e para $C < D$, atenuação.

3.1. Equalizador Gráfico

O EQA gráfico apresenta uma FT tipo *bump*. Seu nome deve-se à disposição dos controles de amplitude, os quais representam a curva de resposta em freqüência do equalizador. Muitos EQAs gráficos utilizam o padrão ISO 266/1997 de freqüências centrais, dividindo o espectro em freqüências distanciados de um fator multiplicativo *k*. Por exemplo, para um EQA de 30 bandas (canais), também chamado de 1/3 de oitava, existem três freqüências centrais a cada oitava. Como uma oitava acima equivale ao dobro da freqüência e tem-se três freqüências separadas por um fator *k* a cada oitava, o valor de *k*, para uma dada freqüência inicial, é determinado por (3). Assim,

$$k^3 f_i = 2 f_i \;\therefore\; k = \sqrt[3]{2} \approx 1{,}26 \tag{3}$$

Portanto, para um EQA deste tipo, tem-se 30 freqüências centrais de controle, entre 25 Hz e 20 kHz, cada uma separada por $k \approx 1{,}26$. Há também EQAs de 1/3 oitava com 31 bandas, incluindo a freqüência inicial de 20 Hz. Encontram-se ainda EQAs de 1 oitava, 2/3 oitava, etc.

Pode-se também utilizar a divisão por décadas para classificar o número de bandas. Por exemplo, o EQA de 1/3 oitava equivale ao EQA de 1/10 de década. Isso porque uma década acima equivale a 10 vezes a freqüência e como, nesse caso, tem-se 10 freqüências centrais em uma década, obtém-se então o valor de *k* de (4). Portanto,

$$k^{10} f_i = 10 f_i \quad \therefore \quad k = \sqrt[10]{10} \approx 1{,}26 \tag{4}$$

de onde constata-se a mesma quantidade de bandas e os mesmos valores de freqüências centrais com respeito ao EQA de 1/3 de oitava. Esses EQAs dispõem apenas de controle de amplitude. Os valores das freqüências centrais e larguras de bandas são fixos.

3.2. Equalizador Paramétrico

Também apresentando uma FT tipo *bump*, o EQA paramétrico, além do usual controle de amplitude, permite controlar tanto os valores das freqüências centrais quanto os valores dos fatores de qualidade (*Q*), correspondentes aos filtros do equalizador. Esses EQAs possuem grande precisão e versatilidade, sobretudo para correção da resposta em freqüência de ambientes.

O controle independente de parâmetros, neste tipo de EQA, pode ser obtido através do projeto de filtros via variáveis de estado [2], [4]. É usual encontrarmos nos EQAs paramétricos comerciais cinco controles paramétricos, além de um controle *shelving* para baixas freqüências e um para altas freqüências.

3.3. Equalizador Paragráfico

Como o próprio nome sugere, o EQA paragráfico é uma combinação dos EQAs gráficos e paramétricos. Da mesma forma que os EQAs gráficos, os paragráficos empregam freqüências centrais definidas pelo padrão ISO (ou outras pré-determinadas) associadas à disposição dos controles de amplitude, que caracterizam sua curva de resposta em freqüência. Dos EQAs paramétricos, os paragráficos herdam o ajuste da freqüência central, podendo ou não possuir ajuste de largura de banda.

Em razão de seu maior custo, conseqüência de sua maior flexibilidade, não são disponíveis comercialmente EQAs paragráficos de 1/3 oitava (30 bandas). No entanto, são oferecidos comercialmente EQAs paragráficos de 1 oitava.

A Tabela 1 e a Fig. 2 resumem os principais parâmetros de um EQA e os associam aos EQAs citados.

Tabela 1 – Parâmetros de controle associados aos equalizadores

Equalizadores	Parâmetros de controle
Controle de tonalidade, paramétrico (primeira banda)	Amplitude (*shelving*) de baixas freqüências (graves) (Fig. 2(a))
Controle de tonalidade, paramétrico (última banda)	Amplitude (*shelving*) de altas freqüências (agudos) (Fig. 2(b))
Gráfico, paramétrico, paragráfico	Amplitude (*bump*) (Fig. 2(c)
Paramétrico, paragráfico	Freqüência central (Fig. 2(d))
Paramétrico, paragráfico (opcional)	Largura de banda (*Q*) (Fig. 2(e))

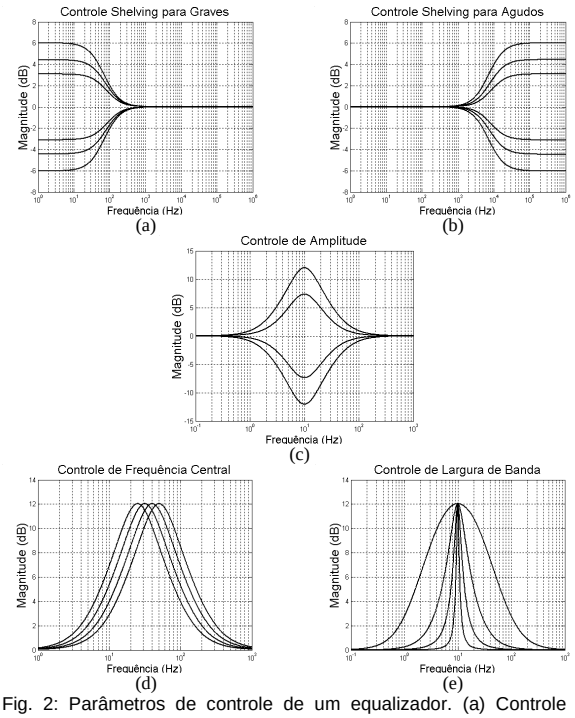


Fig. 2: Parâmetros de controle de um equalizador. (a) Controle de amplitude (*shelving*) para baixas freqüências; (b) controle de amplitude (*shelving*) para altas freqüências; (c) controle de amplitude (*bump*); (d) controle da freqüência central; (e) controle da largura de banda (valor de *Q*).

4. EQUALIZADORES *BUMP*

4.1. Função de Transferência

Como já visto, além do controle de amplitude do tipo *shelving*, cuja FT já foi apresentada, pode-se também destacar o controle de amplitude do tipo *bump*, bastante usado nos EQAs gráficos, paramétricos e paragráficos, sendo aqui apresentado com maior detalhe.

A FT de um equalizador *bump* de 2ª ordem (Fig. 2(c)) é dada por

$$T(s) = \frac{s^2 + (\omega_0 / Q_Z)s + \omega_0^2}{s^2 + (\omega_0 / Q_P)s + \omega_0^2} \,. \tag{5}$$

Decompondo $T(s)$ no produto de duas funções, representando o numerador e o denominador, tem-se:

$$T(s) = T_{\text{num}}(s)\,T_{\text{den}}(s) \tag{6}$$

onde

$$T_{\text{num}}(s) = s^2 + (\omega_0 / Q_Z)s + \omega_0^2 \tag{7}$$

e

$$T_{\text{den}}(s) = \frac{1}{s^2 + (\omega_0 / Q_P)s + \omega_0^2} \tag{8}$$

A Fig. 3 ilustra as curvas de magnitude das respostas em freqüências de $T(j\omega)$, $T_{\text{num}}(j\omega)$ e $T_{\text{den}}(j\omega)$, para freqüência de corte $f_0 = 20\,\text{Hz}$ e um fator amplificação de $12\,\text{dB}$.

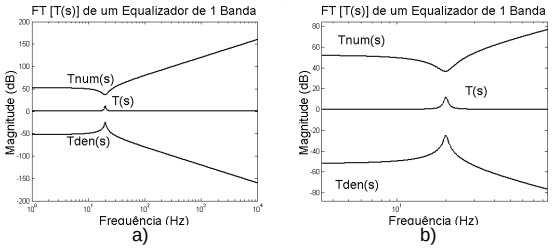


Fig. 3 – Curvas de magnitude de $T(s)$, $T_{\text{num}}(s)$ e $T_{\text{den}}(s)$.(a) Visão geral; (b) detalhe.

Da Fig. 3, constata-se que:

Se $Q_Z > Q_P$, tem-se uma atenuação próxima a f_0.

Se $Q_Z < Q_P$, tem-se uma amplificação próxima a f_0.

Existem diferentes maneiras para a obtenção da FT de um equalizador *bump* [4]-[7]. Uma das mais simples e muito utilizada é a baseada em um filtro passa-faixa de 2ª ordem (em [4] mostra-se que, em função da característica de fase, utilizar seções de 2ª ordem é a melhor opção), um amplificador inversor e uma estrutura de ganho *K* [4], representada por (9)

$$T(s) = 1 - P_F(s) + K\,P_F(s) \tag{9}$$

onde $P_F(s)$ é a FT do filtro passa-faixa, explicitada por (10):

$$P_F(s) = \frac{(\omega_0 / Q)s}{s^2 + (\omega_0 / Q)s + \omega_0^2} \,. \tag{10}$$

Assim, aplicando-se (10) em (9) obtém-se a FT do equalizador *bump*, dada por (11), apresentando um fator de ganho *K* na freqüência de interesse ω_0.

$$T(s) = \frac{s^2 + K(\omega_0 / Q)\,s + \omega_0^2}{s^2 + (\omega_0 / Q)\,s + \omega_0^2} \tag{11}$$

A Fig. 4 ilustra o diagrama em blocos de um equalizador *bump* para uma banda.

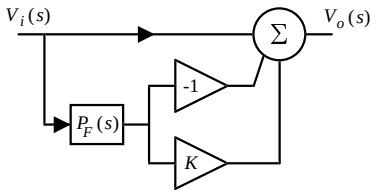


Fig. 4 – Diagrama em blocos de um equalizador *bump* de uma banda.

Pode-se também obter a FT do equalizador *bump* de outras formas, como a partir de um filtro passa-tudo [5]; passa-baixas [6]; passa-altas; da soma das saídas de um filtro passa-baixas, passa-altas e um passa-faixa, como o apresentado na estrutura KHN em [7].

4.2. Função de Transferência para *n* Bandas – Série/Paralelo

Circuitos equalizadores *bump* podem ser associados em série (Fig. 5(a)) ou em paralelo (Fig. 5(b)) para a implementação de EQAs com mais de uma banda [5]. Além disso, variações das topologias série e paralelo também podem ser empregadas. Em [4] e [6] são discutidos outros tipos de associações existentes na literatura.

Da mesma forma que existem diversas topologias, diferentes tecnologias podem ser empregadas para a implementação de um EQA gráfico de *n* bandas [2]. Assim, destaca-se: OTA-C [8], capacitor chaveado [9]-[10], MOSFET-C [11], filtros digitais [12]-[13], dentre outras.

Através da Fig. 5(a) pode ser verificada a FT total do EQA descrito por (12).

$$\frac{V_0(s)}{V_i(s)} = T(s) = \prod_{k=1}^{n} T_k(s) \tag{12}$$

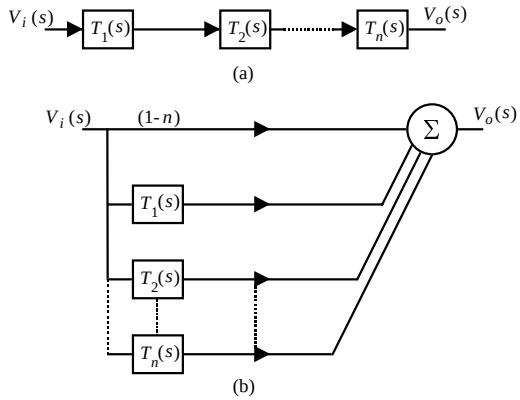


Fig. 5 - Equalizador de "*n*" bandas. (a) Associação série, (b) paralela.

Na configuração paralela não é tão simples visualizar a FT total do EQA de *n* bandas descrito por (13). Tal estrutura é ilustrada pelo diagrama em blocos da Fig. 5(b).

$$\frac{V_0(s)}{V_i(s)} = T(s) = 1 - n + \sum_{k=1}^{n} T_k(s) \tag{13}$$

Como $|T_k(j\omega)|$ é praticamente igual a 1 para todas as freqüências, exceto em torno de f_{0k} (ver Fig. 3), substitui-se $|T_k(j2\pi f_{0k})|$ por $1+\delta(f_{0k})$, onde $\delta(f_{0k})$ é o valor de amplificação ou atenuação em torno da correspondente freqüência f_{0k}. Assim, partindo de (13) obtém-se

$$|T(j\omega)| = 1 + \sum_{k=1}^{n} \delta(f_{ok}), \tag{14}$$

chegando ao resultado desejado para o controle de amplitude de um EQA de *n* bandas.

Para ilustrar, é mostrada na Fig. 6 a magnitude da resposta em freqüência de um equalizador *bump* com duas bandas. Neste exemplo é usada $f_{01} = 20\,\text{Hz}$, $f_{02} = 25\,\text{Hz}$, com $Q = 15$, sendo que essas são as freqüências centrais da primeira e da segunda banda de um EQA gráfico comercial de 31 bandas.

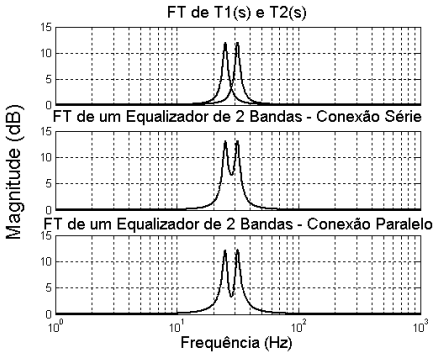


Fig. 6 – Equalizador de 2 bandas: Conexão Série e Paralelo.

A interação que ocorre entre bandas vizinhas, conforme observado pela Fig. 6, é conhecida como efeito de combinação ou interpolação, sendo uma característica bastante relevante em projeto de EQAs, a qual será tratada a seguir.

5. *Q*-CONSTANTE E *Q*-NÃO CONSTANTE (PROPORCIONAL)

Aqui *Q*-constante e *Q*-não constante (proporcional) estão relacionados ao comportamento da largura de banda dos filtros que compõem o EQA como uma função ou não dos níveis de amplificação/atenuação. Portanto, aqui esses termos não se referem à disponibilidade de controle de largura de banda dos EQAs paramétricos.

A abordagem *Q*-constante teve origem no projeto de um EQA gráfico de 1/3 oitava (que possui freqüências centrais relativamente próximas), no qual se verificou que havia uma considerável interação entre as amplitudes das freqüências vizinhas quando tais amplitudes não estavam ajustadas para seus valores máximos. Dessa forma, a resposta em freqüência do EQA não mais representava a curva esboçada pela posição dos controles de amplitude no painel do equipamento. Aqui é importante ressaltar que em seus valores máximos, as abordagens *Q*-constante e *Q*-não constante apresentam o mesmo efeito na combinação de amplitudes.

A abordagem *Q*-não constante significa que a largura de banda é uma função (inversamente proporcional) dos níveis de amplificação/atenuação. Para o valor máximo de amplificação/atenuação obtém-se a menor largura de banda. Quando estes níveis são diminuídos é obtido um aumento da largura de banda e, portanto, dessa forma tem-se a influência sobre um maior número de freqüências. Alguns autores preferem usar o termo *Q*-proporcional no lugar de *Q*-não constante, visto que a largura de banda nessa abordagem é proporcional aos níveis de amplificação/atenuação. Por outro lado, o termo *Q*-constante significa que a largura de banda não depende dos níveis de ajuste da amplitude, permanecendo esse fixo. A Fig. 7 ilustra a diferença entre as duas abordagens, evidenciando a vantagem do EQA gráfico de *Q*-constante.

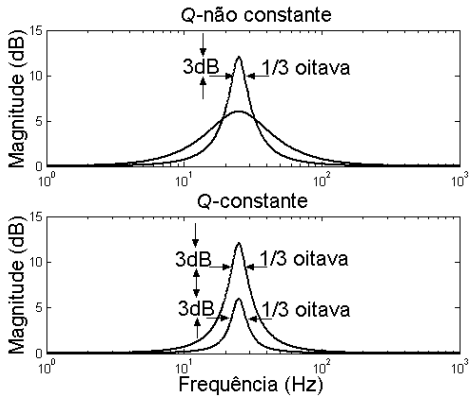


Fig. 7 – Comparação entre as abordagens *Q*-constante e *Q*-não constante.

Nota-se na Fig. 7 que no caso de *Q*-não constante há um aumento da largura de banda (diminuição de *Q*) quando se diminui a amplitude. Enquanto o EQA de *Q*-constante mantém a largura de banda de 1/3 oitava.

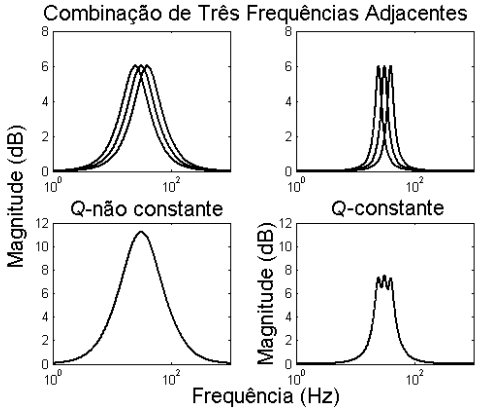


Fig. 8 –Combinação de 3 freqüências adjacentes ajustadas em +6 dB na amplitude.

A vantagem da abordagem que usa *Q*-constante fica mais evidenciada quando se compara o efeito da interação entre amplitudes de freqüências adjacentes, como mostrado pela Fig. 8. Nota-se que, quando usado *Q*-não constante, existe uma grande degradação na resposta em freqüência devido à interação entre as três freqüências adjacentes (distantes de 1/3 oitava), considerando um ajuste de amplitude de +6 dB para cada freqüência. Além de não se alcançar o efeito de amplificação desejado (+6 dB) em cada uma das freqüências, nota-se que existe um aumento da largura de banda. Por outro lado, na configuração de *Q*-constante percebe-se apenas o efeito de combinação das três freqüências, não havendo qualquer considerável degradação da resposta em freqüência do EQA (apenas uma pequena ondulação). No entanto, essa ondulação (*ripple*) pode ser reduzida através do uso de técnicas mistas (série-paralelo) como a apresentada em [4].

A condição necessária e suficiente para que se tenha um EQA gráfico de *Q*-constante é que a função que define a amplitude deve ser independente daquela do filtro passa-faixa que caracteriza o valor de *Q*.

Considera-se os EQAs gráficos de *Q*-constante (disponível comercialmente a partir de 1982) como um avanço importante em relação aos mais antigos EQAs que usam *Q*-não constante (proporcional).

6. LARGURA DE BANDA DO EQUALIZADOR 1/3 OITAVA (SÉRIE/PARALELO)

Como mostrado na Fig. 7 e discutido em [2] e [4], para os EQAs de 1/3 oitava, não apenas as freqüências de controle devem estar distantes de 1/3 oitava, mas também a largura de banda (em -3 dB) tem uma influência no intervalo de freqüências de 1/3 oitava (Fig. 9).

O valor de *Q* para EQAs gráficos, projetados a partir de filtros passa-faixa, é dado por

$$Q = \frac{\omega_0}{B} = \frac{\omega_0}{\omega_2 - \omega_1}\,, \tag{15}$$

onde ω_0 é a freqüência de controle (que está sendo equalizada), e ω_1 e ω_2 são as freqüências de amplitude –3 dB em relação ao valor da amplitude em ω_0. A relação entre ω_0, ω_1 e ω_2 é dada por

$$\omega_0 = \sqrt{\omega_1 \cdot \omega_2}\,. \tag{16}$$

Assim, *Q* pode ser determinado por (17). Portanto,

$$Q = \frac{\sqrt{\omega_1 \cdot \omega_2}}{\omega_2 - \omega_1}\,. \tag{17}$$

Como a largura de banda de um EQA gráfico de 1/3 oitava deve ser também de 1/3 oitava e sabendo que as freqüências devem estar separadas por um fator multiplicativo igual a 1,26 (Eq. (3)), obtém-se via Eq. (17) um valor de $Q = 4{,}137$. Entretanto, deve-se ressaltar que esse valor de *Q*, o qual produz os resultados obtidos na simulação da Fig. 8, é válido somente para a configuração em paralelo. Para um EQA de *n* bandas conectadas em série, o uso desse valor de *Q*, objetivando uma largura de banda de 1/3 oitava, não levará ao mesmo resultado obtido pela conexão em paralelo. Isso porque, como mostrado na Fig. 9, as freqüências de –3 dB referentes à freqüência superior f_{21} de uma banda f_{01}, e à freqüência inferior f_{12} de uma banda adjacente f_{02}, são coincidentes. Como na conexão em série as amplitudes simplesmente somam-se em dB, pode-se obter um pico muito alto na resposta de freqüências adjacentes.

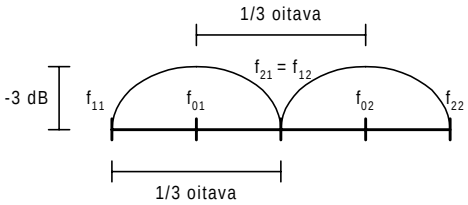


Fig. 9 – Equalizador gráfico de 1/3 oitava com largura de banda de 1/3 oitava.

Como exemplo, repetem-se na Fig. 10 os resultados da Fig. 6, agora não mais considerando um alto Q $(=15)$, mas o valor de $Q = 4{,}137$. Nota-se que a conexão série fornece um valor de amplitude cerca de 18 dB, referente à soma dos pontos de cruzamento de –3 dB. Caso se deseje obter resultados semelhantes aos da conexão paralelo para a conexão série, deve-se utilizar valores altos de *Q*. No entanto, o valor de $Q = 4{,}137$ para a conexão série apresenta uma característica que pode ser vantajosa, uma vez que o *ripple* resultante é menor do que o da conexão paralelo. Por esta razão, podem-se imaginar técnicas com configurações mistas visando diminuir o *ripple* da resposta em freqüência de EQAs gráficos [4].

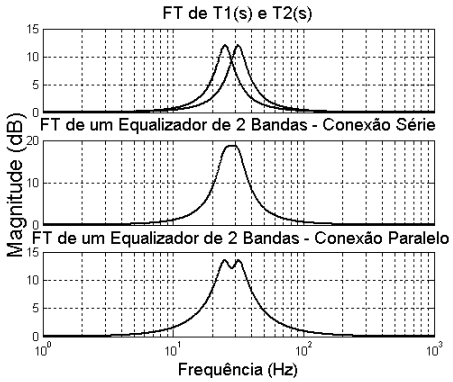


Fig. 10 – Duas bandas de um equalizador gráfico de 1/3 oitava: série-paralelo.

7. CONSIDERAÇÕES SOBRE A RESPOSTA DE FASE DE UM EQUALIZADOR DE ÁUDIO

Até este ponto, consideramos apenas as características de magnitude dos EQAs sem qualquer discussão sobre o comportamento de sua fase. Nesta seção é apresentada uma breve discussão da importância da fase para os EQAs.

Funções de transferência apresentando zeros situados no semiplano lateral esquerdo são funções de fase mínima. Nesse caso, o sistema apresenta uma mínima variação de fase, visto que os zeros situados no semiplano lateral esquerdo tendem a se "cancelar" com os pólos do sistema considerado. Quando uma rede de fase não-mínima é equalizada em magnitude, a tendência é aumentar o desvio de fase se compararmos com a rede não equalizada. Por outro lado, quando uma rede de fase mínima é equalizada em magnitude, o processo de equalização também equaliza a fase, podendo ser esse um resultado bastante interessante. É bom ainda lembrar que a característica de fase em aplicações de áudio não é tão importante, a menos que:

i) os tempos associados à variação de fase sejam maiores do que o tempo de resposta do ouvido humano;
ii) os sinais acústicos resultantes de mais de uma fonte possam interagir uns com os outros obtendo-se, assim, um processo de interferência (construtiva ou destrutiva) entre sinais de mesma freqüência;
iii) ocorram condições de osclilação. Tal situação, popularmente conhecida por microfonia, é decorrente de uma interação entre microfone, amplificador, alto-falante e ambiente.

Fase mínima é uma característica muito discutida e desejável em projetos de filtros em geral. Assim, para os primeiros EQAs projetados, sugeriu-se que tal característica também fosse contemplada por tais sistemas. De forma geral, hoje pode-se afirmar que exibir fase mínima não constitui nenhuma vantagem adicional para um EQA, uma vez que, quase na totalidade, todas as técnicas de projeto de EQAs conhecidas originam equalizadores com fase mínima. Em [6] é provado que todas as topologias básicas de EQAs são de fase mínima para qualquer combinação dos controles de ajuste. Apesar disso, alguns fabricantes se utilizam de tal característica para fazer *marketing* de seus produtos, tornando-se este assunto um dos mitos na discussão sobre EQAs [14].

8. CONCLUSÕES

Apresentou-se neste artigo diversas características comuns aos EQAs, iniciando-se por um breve histórico sobre o seu desenvolvimento; passando por uma classificação dos principais tipos; detalhando a FT clássica dos EQAs tipo *bump*, mostrando as conexões para formar um EQA de *n* bandas; e o efeito da combinação de freqüências vizinhas, comparando as topologias de *Q*-não constante e *Q*-constante, evidenciando-se a vantagem dessa última. Uma discussão sobre o valor de largura de banda dos EQAs gráficos de 1/3 oitava é apresentada, mostrando que o valor clássico de 4,317 (largura de banda de 1/3 oitava) é aplicável sem restrições somente para EQAs com bandas conectadas em paralelo. Finalmente, algumas considerações sobre o comportamento da fase, a qual apresenta característica de fase mínima para quase a totalidade de projetos envolvendo EQAs, é apresentada.

AGRADECIMENTOS

Os autores agradecem ao CNPq pelo suporte financeiro desta pesquisa.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] E. Rapoport, A. Petraglia, "Estudo e Correção de Perdas Auditivas Induzidas por Ruído usando Equalizadores Programáveis", XIV Congresso Brasileiro de Automática, Natal-RN, Setembro de 2002.
[2] B. Dennis, "Operator Adjustable Equalizers: An Overview," RaneNote 122 (1990).
[3] Yamaha – Sound Reinforcement Handbook. Section 14: 14.1 Equalizers, p. 244-255.
[4] B. Dennis, "Constant-Q Graphic Equalizers", 79th AES Convention, New York, USA, 1985; revised 1986.
[5] J. A. De Lima, A Petraglia, "On Designing OTA-C Graphic-Equalizers With MOSFET-TRIODE Transconductors", IEEE, p. I-212 – I-215, 2001.
[6] R.A Greiner and M. Schoessow, " Design Aspects of Graphic Equalizers ", presented at the 69th Convention of Audio Engineering Society, J. Audio Eng. Soc. (Abstracts), vol.29, p. 556 (1981 july/Ago), preprint 1767.
[7] S. Noceti, "Filtros Seletores de Sinais", 1ª ed., Florianópolis, Ed. da UFSC, 1998.
[8] A C. .M. de Queiroz, A. Petraglia and S. K. Mitra, "Tunable OTA-C Equalizers", IEEE, p. 2029-2032, 1992.
[9] J. F. Duque-Carrillo, J. Silva Martínez, E. Sánchez-Sinencio, "Programmable Switched-Capacitor Bump Equalizer Architecture", IEEE Journal Of Solid-State Circuits, vol. 25, n° 4, p. 1035-1039, 1990.
[10] J. Silva Martínez, "A Programmable Switched-Capacitor Filter", ISCAS 1994, p. 727-730.
[11] S. Sakurai, M. Ismail, J.-Y. Michel, E. Sánchez-Sinencio and R. Brannen, "A MOSFET-C Variable Equalizer Circuit with Simple On-Chip Automatic Tunning", IEEE Journal Of Solid-State Circuits, vol. 27, n° 6, p. 927-934, 1992.
[12] J. Dattorro, "The Implementations of Recursive Digital Filters for High-Fidelity Audio", Journal of the Audio Engineering Society, Vol. 36, n° 11, p. 851-878, 1988.
[13] A. J. S. Ferreira, J. M. N Vieira, "An Efficient 20 Band Digital Audio Equalizer", 98th AES Convention, Paris, France, 1995.
[14] B. Dennis, "Exposing Equalizer Mythology," RaneNote 115 (1986).

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na VII Convenção Nacional
26-28 de maio de 2003, São Paulo, 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 do Comportamento de Alto-falantes Excitados por Fonte de Corrente para Pequenos e Grandes Sinais

Rosalfonso Bortoni[1,2], Sidnei Noceti Filho[2], José Antônio Justino Ribeiro[3] e Rui Seara[2]

[1]Studio R Eletrônica
São Paulo, 04615-005, São Paulo, Brasil
rosalfonso@terra.com.br
[2]LINSE – Laboratório de Circuitos e Processamento de Sinais
Departamento de Engenharia Elétrica, UFSC
Florianópolis, 88040-900, Santa Catarina, Brasil
sidnei@linse.ufsc.br, seara@linse.ufsc.br
[3]INATEL
Santa Rita do Sapucaí, 37540-000, Minas Gerais, Brasil
justino@inatel.br

RESUMO

O projeto de caixas acústicas, que no passado era essencialmente heurístico, passou a ser, atualmente, uma técnica com maior rigor científico a partir dos trabalhos de Neville Thiele e Richard Small, na década de 70. Tais trabalhos proporcionaram a obtenção de um modelo linear para um alto-falante excitado por fonte de tensão, sendo que o modelo obtido é válido apenas para pequenos sinais de excitação. Adicionalmente, estudos têm sido feitos para introduzir características não-lineares aos modelos de alto-falantes, visto que esses últimos são na prática muitas vezes submetidos a sinais com grandes amplitudes. Este trabalho propõe uma nova abordagem para a análise do comportamento de alto-falantes excitados por fonte de corrente, considerando as mais importantes condições de operação envolvidas. Dessa forma, é mostrado ser possível diminuir distorções e compensar alguns de seus indesejáveis efeitos.

1. INTRODUÇÃO

A importância da interação entre amplificador e alto-falante, como único bloco de um sistema de áudio, tem sido estudada já há algum tempo, e diferentes técnicas de construção e excitação de alto-falantes [1,2,7-10,17,21] têm sido consideravelmente exploradas. Com raras exceções, alto-falantes e caixas acústicas são excitados por amplificadores de potência, cujo sinal de saída é modelado por uma fonte de tensão com baixa impedância [21]. O projeto de caixas acústicas, inicialmente realizado de forma heurística, passou a ter um maior rigor científico a partir dos trabalhos de Neville Thiele e Richard Small [2-5]. Tal procedimento, hoje denominado Método de Thiele-Small, baseia-se no fato de que um sistema alto-falante/caixa acústica pode ser adequadamente representado por um circuito equivalente eletromecânico-acústico, cuja função de transferência da pressão sonora/tensão de entrada é análoga à de um filtro passa-altas de segunda ordem (ou maior) [2-6]. Nesse método, os parâmetros são considerados lineares e o modelo é válido apenas para pequenos sinais.

O modelo caracterizado por um filtro de segunda ordem é mostrado na Figura 1. O amplificador é representado por um gerador de tensão senoidal (E_g) em série com uma impedância puramente resistiva (R_g). Os demais componentes (linha pontilhada) representam o equivalente elétrico do alto-falante. Thiele demonstra ser possível ajustar um sistema alto-falante/caixa acústica variando-se a resistência de saída do amplificador, fazendo-a positiva ou negativa conforme a necessidade [2], técnica proposta por Werner e Carrel [7]. Como os amplificadores atuais apresentam baixa impedância de saída, projetar uma caixa acústica significa trabalhar em suas características (volume

interno, sintonia, absorção, dentre outras) para adequá-la a um determinado alto-falante, ou adaptar um alto-falante à caixa acústica, ou adotar ambos os procedimentos. Assim, o processo concentra-se apenas no lado eletroacústico do sistema. Essa conduta não considera a resistência elétrica dos cabos que fazem a conexão entre o amplificador e a caixa acústica, o que pode levar a grandes erros [8].

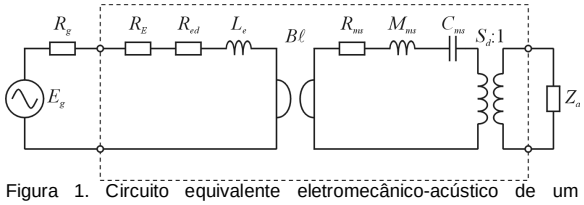


Figura 1. Circuito equivalente eletromecânico-acústico de um alto-falante.

Atuando em qualquer uma das subestruturas (amplificador, alto-falante e caixa acústica) modifica-se o comportamento final do sistema. Modificar as características mecânicas implica em alterar a forma de construir o alto-falante, o que não é trivial. O usual é ajustar o volume interno da caixa de som e/ou modificar a sua sintonia através da alteração nas dimensões do pórtico (duto). Com o método da resistência negativa [2,7], alteram-se as características elétricas do alto-falante, influenciando diretamente o fator de mérito, a eficiência do sistema e outros parâmetros de desempenho [2-8]. Um método para alterar as características mecânicas do alto-falante, com atuação no lado elétrico do sistema amplificador/caixa acústica, tem sido proposto por Stahl [16]. Ele consiste em fazer a resistência de saída do amplificador, em módulo, igual à resistência da bobina do alto-falante, porém de sinal contrário (negativo). Dessa forma, ambas se cancelam e o amplificador passa a ter como carga o circuito mecânico do alto-falante. Uma vez alterada as características de transferência do amplificador, modificam-se as características mecânicas do alto-falante. Esse método também foi explorado por Normandin [10]. O circuito proposto por Stahl era capaz de apenas gerar uma resistência negativa de saída proporcional à resistência elétrica da bobina do alto-falante e não à impedância elétrica. Esse fato restringiu o uso do circuito proposto a baixas freqüências. É importante lembrar que, em médias e altas freqüências, a reatância indutiva da bobina passa a ser relevante. Outra limitação é o aumento da resistência elétrica da bobina do alto-falante com a temperatura, o que obrigaria, quase sempre, o uso de um sistema termo-compensado [11,12].

Além das características lineares, existe a não-linearidade da impedância elétrica da bobina, causada pelas correntes induzidas no conjunto magnético do alto-falante [2,6]. Tal não-linearidade é representada pelos elementos R_{ed} e L_e da Figura 1, conforme o modelo proposto por Wright, onde:

$$R_{ed} = K_r \cdot \omega^{X_r} \quad \text{e} \quad L_e = K_i \cdot \omega^{(X_i - 1)}, \tag{1}$$

sendo K_r, X_r, K_i, X_i obtidos da curva de impedância do alto-falante [14].

A força aplicada ao conjunto móvel do alto-falante (bobina, cone e suspensão) é resultado do produto da corrente resultante do lado elétrico pelo fator de força $B\ell$ ($F = B\ell I_e(s)$), onde B é a indução magnética que atravessa o espaço de ar (*gap*), ℓ é o comprimento efetivo da bobina dentro do *gap* e $I_e(s)$ é a corrente resultante no lado elétrico (bobina móvel). Devido a não-uniformidade de distribuição do fluxo magnético nas bordas do *gap* e ao deslocamento da bobina móvel, o fator de força também contribui com não-linearidades. Em regime de grandes potências (grandes deslocamentos do cone) ocorre a variação da compliância (C_{ms}) do conjunto móvel, o que também representa uma não-linearidade do sistema [15,16]. Essas não-linearidades contribuem para as distorções que surgem no resultado sonoro final. Mills e Hawksford [17] mostraram ser possível reduzir tais distorções usando amplificadores de corrente no lugar dos tradicionais amplificadores de tensão. Para comprovar tal procedimento, foi construído um protótipo no qual o subsistema eletrônico e o eletroacústico eram subestruturas integrantes de um único módulo. Como resultado, as características sonoras superaram às dos sistemas convencionais. Mills e Hawksford lamentaram o fato dessa técnica não ter sido mais explorada. Posteriormente, Birt usou essa técnica, confirmando ser possível obter baixas distorções [18]. Ao se considerar uma excitação por fonte de corrente, são observados diferentes comportamentos, dentre eles: a correção de algumas não-linearidades (com conseqüente redução de distorções); e uma ampliação dos limites das respostas de baixas e altas freqüências.

Este trabalho apresenta uma nova abordagem de análise da pressão sonora de alto-falantes excitados por fontes de corrente para operação com grandes sinais. Mostra-se que as distorções e a compressão de potência podem ser reduzidas. Em nosso conhecimento, essa abordagem não foi, até então, formalmente apresentada na literatura.

2. EXCITAÇÃO POR FONTES DE TENSÃO E CORRENTE

A seguir, é apresentada uma análise comparativa entre os modos de excitação por fonte de tensão e por fonte de corrente, considerando um alto-falante instalado em um painel cuja menor distância entre a borda do alto-falante e qualquer de suas extremidades é muitas vezes maior do que o maior comprimento de onda acústica que se deseja reproduzir (aproximação de *Baffle* infinito) [6].

2.1 Excitação por fonte de tensão

Na Figura 2, mostra-se o circuito acústico equivalente de um alto-falante, instalado em um *baffle* infinito, excitado por fonte de tensão.

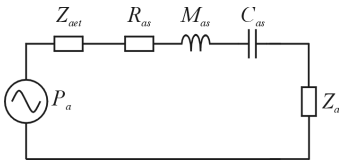


Figura 2. Circuito acústico equivalente de um alto-falante excitado por fonte de tensão instalado em um *baffle* infinito.

A pressão sonora a uma distância r do alto-falante é dada por $P_r(s) = (\rho/2\pi r)U(s)s$, onde $U(s)$ é o volume de ar deslocado [6]. O resultado final é dado em (2). O primeiro fator (pressão acústica) é função da impedância elétrica da bobina móvel, $Z_e(s)$, e do fator de força $B\ell$, que são duas grandezas não-lineares. O último fator é arranjado de forma a explicitar a influência que cada um dos subsistemas (elétrico, mecânico e acústico) tem sobre o sistema total; assim, tem-se para denominador desse fator: o primeiro termo é função das mesmas grandezas elétricas não-lineares citadas anteriormente, só que variando com

$(B\ell)^2$. O segundo termo é função de uma grandeza mecânica não-linear, C_{ms}; o terceiro termo caracteriza a impedância acústica de irradiação do ar, Z_a. Considerando $|Z_g(s)| << |Z_e(s)|, \forall s$, as não-linearidades ainda permanecem, como pode ser observado em (3).

$$P_r(s) = \frac{E_g(s)B\ell}{S_d[Z_g(s)+Z_e(s)]} \cdot \frac{\rho}{2\pi r} \times \frac{s}{\dfrac{(B\ell)^2}{S_d^2[Z_g(s)+Z_e(s)]} + \dfrac{1}{C_{as}} \cdot \dfrac{\dfrac{s^2}{\omega_s^2} + \dfrac{s}{\omega_s Q_{as}} + 1}{s} + Z_a} \tag{2}$$

$$P_r(s) = \frac{E_g(s)B\ell}{S_d Z_e(s)} \cdot \frac{\rho}{2\pi r} \cdot \frac{s}{\dfrac{(B\ell)^2}{S_d^2 Z_e(s)} + \dfrac{1}{C_{as}} \cdot \dfrac{\dfrac{s^2}{\omega_s^2} + \dfrac{s}{\omega_s Q_{as}} + 1}{s} + Z_a} \tag{3}$$

2.2 Excitação por fonte de corrente

Na Figura 3, tem-se o circuito acústico equivalente de um alto-falante instalado em um *baffle* infinito excitado por fonte de corrente.

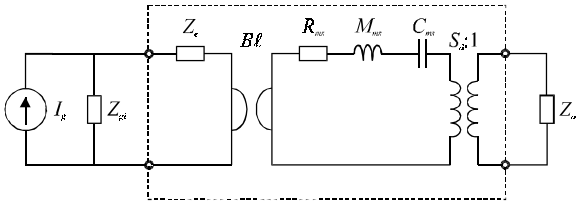


Figura 3. Circuito eletromecânico-acústico equivalente de um alto-falante excitado por fonte de corrente instalado em um *baffle* infinito.

Refazendo para a pressão sonora o mesmo que foi feito para o caso da excitação por fonte de tensão, obtém-se a equação (4). No primeiro fator, a pressão acústica $(I_g(s)B\ell/S_d)$ passa a ser função apenas de uma grandeza não-linear, o fator de força $B\ell$. O último fator, da mesma maneira que anteriormente, é arranjado de forma a explicitar a influência que cada um dos subsistemas (elétrico, mecânico e acústico) tem sobre o sistema total. Assumindo que $|Z_{gi}(s)| >> |Z_e(s)|$, $\forall s$, pois a excitação é agora por fonte de corrente, obtém-se (5). Esta equação mostra que se a influência de $|Z_e(s)|$ é eliminada, reduz-se as não-linearidades. Caso a impedância de saída do gerador de corrente seja muito elevada, o termo que contém $(B\ell)^2$ pode também ser desconsiderado, ficando apenas $B\ell$ e C_{as}, como mostrado na (6). Nesse caso, o fator de mérito do sistema passa a ser dependente dos subsistemas mecânico e acústico, o que pode resultar em um pico de ressonância na resposta, pois o fator de mérito mecânico do alto-falante, Q_{ms}, é comumente bem maior do que o fator de mérito elétrico, Q_{es}. No caso da excitação por fonte de tensão, o fator de mérito resultante é dado por $Q_{ts} = [(Q_{es})^{-1} + (Q_{ms})^{-1}]^{-1}$. Fazendo $Z_a = 0$ pode-se constatar que a característica de transferência desse sistema é igual a de um filtro passa-altas de segunda ordem (equação (7)), onde C_{as} é substituído por $S_d^2 C_{ms}$ para possibilitar uma mais fácil análise das duas grandezas não-lineares ainda restantes no sistema ($B\ell$ e C_{ms}). Alguns autores estudaram a variação desses dois parâmetros, para o alto-falante submetido a grandes potências (grandes deslocamentos do cone) [16-18], considerando os efeitos $B\ell$ e C_{ms} separadamente. Entretanto, a equação (7) mostra que o mais importante é o produto $B\ell C_{ms}$.

$$P_r(s) = \frac{I_g(s)B\ell Z_{gi}(s)}{S_d[Z_{gi}(s)+Z_e(s)]} \cdot \frac{\rho}{2\pi r} \times \frac{s}{\dfrac{(B\ell)^2}{S_d^2[Z_{gi}(s)+Z_e(s)]} + \dfrac{1}{C_{as}} \cdot \dfrac{\dfrac{s^2}{\omega_s^2} + \dfrac{s}{\omega_s Q_{as}} + 1}{s} + Z_a} \tag{4}$$

$$P_r(s) = \frac{I_g(s)B\ell}{S_d} \cdot \frac{\rho}{2\pi r} \cdot \frac{s}{\dfrac{(B\ell)^2}{S_d^2 Z_{gi}(s)} + \dfrac{1}{C_{as}} \cdot \dfrac{\dfrac{s^2}{\omega_s^2} + \dfrac{s}{\omega_s Q_{as}} + 1}{s} + Z_a} \tag{5}$$

$$P_r(s) = \frac{I_g(s)B\ell}{S_d} \cdot \frac{\rho}{2\pi r} \cdot \frac{s}{\dfrac{1}{C_{as}} \cdot \dfrac{\dfrac{s^2}{\omega_s^2} + \dfrac{s}{\omega_s Q_{as}} + 1}{s} + Z_a} \tag{6}$$

$$P_r(s) = I_g(s)B\ell \cdot \frac{\rho}{2\pi r} \cdot S_d C_{ms} \omega_s^2 \cdot \frac{s^2}{s^2 + s\dfrac{\omega_s}{Q_{as}} + \omega_s^2} \tag{7}$$

3. NÃO-LINEARIDADES DO ALTO-FALANTE

3.1 Partes do alto-falante

Estudos anteriores mostraram que as não-linearidades do alto-falante estão relacionados aos seus subsistemas elétrico e mecânico (conjunto móvel) [2,6,13-18]. A Figura 4 ilustra a estrutura básica de um alto-falante, com destaques para os seus elementos mais importantes. Reconhece-se o subsistema elétrico composto pela bobina, o conjunto magnético representado pelo ímã, pela peça polar e pelas arruelas superior e inferior e o conjunto móvel formado pela bobina, cone, aranha, calota e anel da suspensão. A carcaça é a estrutura de sustentação de todos os elementos citados.

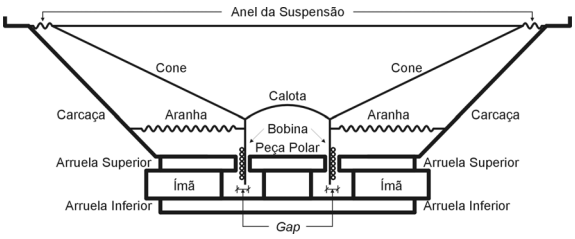


Figura 4. Estrutura básica do alto-falante.

3.2 Não-linearidades da bobina móvel

A impedância elétrica da bobina do alto-falante é representada pela equação (8):

$$Z_e(s) = R_E(T) + R_{ed}(\omega) + s\,L_e(\omega) \tag{8}$$

onde T é a temperatura em ^oC, ω é a freqüência angular $(2\pi f)$ e $s = j\omega$ [2,6,11-14]. Os termos $R_{ed}(\omega)$ e $L_e(\omega)$ estão definidos na equação (1).

A variação da resistência elétrica da bobina em função da temperatura é descrita em (9):

$$R_E(T) = R_E(T_o) \cdot \left[1 + \frac{T - T_o}{T_o - 25 + \dfrac{1}{\alpha_{25}}} \right], \qquad (9)$$

onde T_o é uma temperatura genérica de referência, $R_E(T_o)$ é a resistência elétrica da bobina na temperatura T_o, α_{25} é o coeficiente de temperatura do fio a 25 ^{o}C e T é a temperatura na qual se deseja determinar a nova resistência [11,12].

A partir de especificações técnicas de alto-falantes [19], obtêm-se as curvas de $R_E(T)$, $R_{ed}(\omega)$ e $L_e(\omega)$, mostradas nas Figuras 5, 6 e 7, respectivamente. Aplicando esses resultados à equação (8), determina-se a curva da impedância elétrica da bobina em função da freqüência e da temperatura (Figura 8).

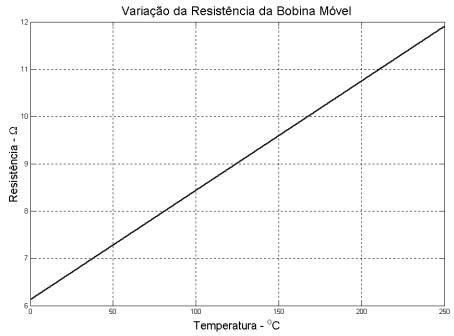


Figura 5. Variação da resistência elétrica da bobina em função da temperatura.

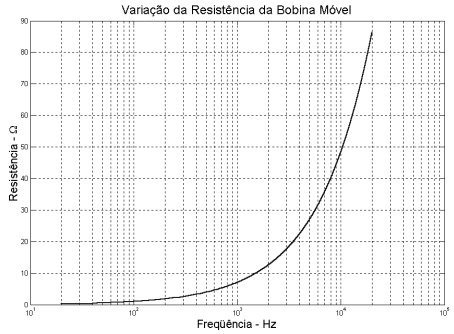


Figura 6. Variação da resistência elétrica da bobina devido às correntes induzidas no conjunto magnético do alto-falante em função da freqüência.

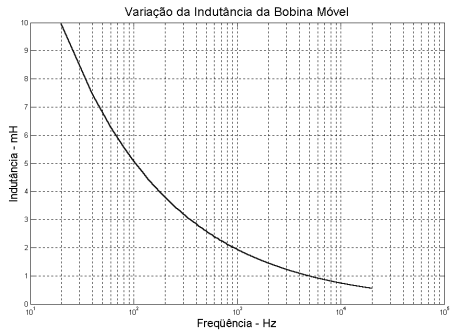
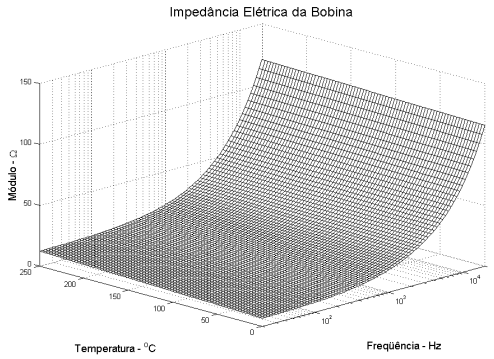
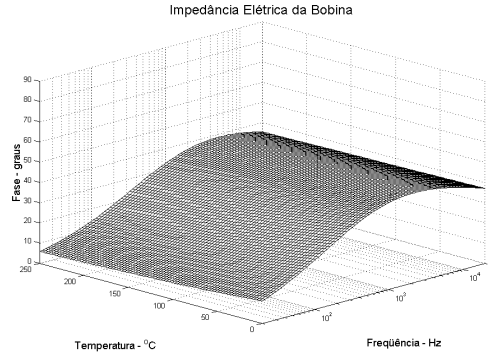


Figura 7. Variação da indutância da bobina devido às correntes induzidas no conjunto magnético do alto-falante em função da freqüência.



(a)



(b)

Figura 8. Impedância elétrica da bobina em função da freqüência e da temperatura: (a) módulo e (b) fase.

3.3 Não-linearidades do fator de força

O fator de força $(B\ell)$ é o produto da indução magnética (B) que atravessa o *gap* pelo comprimento efetivo (ℓ) da bobina dentro desse. Na Figura 4, pode-se observar que, com grandes sinais de excitação, o fator de força sofre uma variação devido ao deslocamento da bobina [15-18]. A partir dos dados reais [16], ilustra-se o comportamento deste parâmetro na Figura 9. O aumento do fator de força em deslocamentos negativos, nesse caso, justifica-se pela penetração da bobina no conjunto magnético (ver Figura 4) [16].

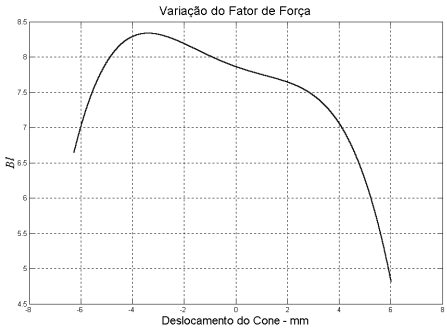


Figura 9. Exemplo do comportamento do fator de força em função do deslocamento do cone [16].

3.4 Não-linearidades da compliância

A compliância de um alto-falante (C_{ms}) é determinada pela maleabilidade das partes que fixam o conjunto móvel à carcaça, ou seja, o anel da suspensão e a aranha. Da mesma forma que para o fator de força $(B\ell)$, para grandes sinais estas estruturas sofrem um estiramento maior e, conseqüentemente, passam a ter uma maleabilidade menor, correspondente a uma diminuição da compliância

(Figura 10) [15-18]. Como a freqüência de ressonância do alto-falante ao ar livre é dada por:

$$F_s = \frac{1}{2\pi \cdot \sqrt{C_{ms} \cdot M_{ms}}}, \tag{10}$$

este parâmetro também sofrerá uma variação em função do deslocamento, como ilustrado na Figura 11.

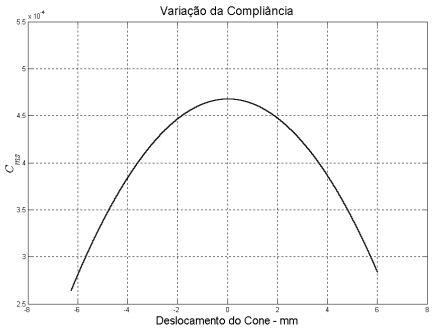


Figura 10. Exemplo do comportamento da compliância em função do deslocamento do cone [16].

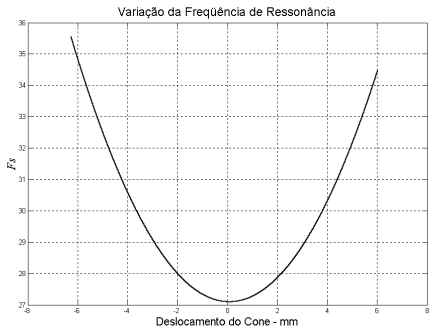


Figura 11. Exemplo do comportamento da freqüência de ressonância do alto-falante ao ar livre em função do deslocamento do cone [16].

4. COMPARAÇÃO ENTRE OS MODOS DE EXCITAÇÃO

4.1 *Baffle* infinito

Reescrevendo as equações (2) e (4), incluindo agora as não-linearidades aqui discutidas, têm-se as pressões sonoras, equações (11) e (12), para excitações por fontes de tensão e corrente, respectivamente (não é levado em conta a impedância acústica de irradiação do ar, Z_a). A distância r é feita igual a um metro, pois é a distância de referência em medidas de sensibilidade de alto-falantes e caixas acústicas. Os índices V e I são incluídos na pressão sonora (P_r) para distinção entre as duas condições. Então:

$$P_{rV}(s) = \frac{E_g(s) B\ell(x)}{S_d[Z_g(s) + Z_e(s)]} \cdot \frac{\rho}{2\pi r} \cdot A \tag{11}$$

$$P_{rI}(s) = \frac{I_g(s) \cdot B\ell(x) \cdot Z_{gi}(s)}{S_d \cdot \left(Z_{gi}(s) + Z_e(s)\right)} \cdot \frac{\rho}{2\pi r} \cdot B \tag{12}$$

$$A = \frac{s}{\dfrac{[B\ell(x)]^2}{S_d^2[Z_g(s) + Z_e(s)]} + \dfrac{1}{S_d^2 C_{ms}(x)} \cdot \dfrac{\dfrac{s^2}{[\omega_s(x)]^2} + \dfrac{s}{\omega_s(x) Q_{ms}(x)} + 1}{s}} \tag{13}$$

$$B = \frac{s}{\dfrac{[B\ell(x)]^2}{S_d^2[Z_{gi}(s) + Z_e(s)]} + \dfrac{1}{S_d^2 C_{ms}(x)} \cdot \dfrac{\dfrac{s^2}{[\omega_s(x)]^2} + \dfrac{s}{\omega_s(x) Q_{ms}(x)} + 1}{s}} \tag{14}$$

Com as especificações técnicas do alto-falante modelo WPU1807 e usando temperatura de 24 °C para a bobina [19], constroem-se os gráficos do nível de pressão sonora (*SPL*) em função da freqüência e do deslocamento (Figura 12), tomando-se como referência a pressão sonora de 20 μPa (considerada como o limiar de audição humana [20]). Desconsiderando-se o deslocamento e variando-se a temperatura (T) de 0 °C a 250 °C, pode-se observar a influência térmica no nível de pressão sonora (Figura 13). Os níveis de tensão e corrente de excitação são determinados de modo a se obter uma potência de 1 Watt *RMS* em uma carga resistiva de 8Ω, que é a impedância nominal do alto-falante, resultando em $E_g = 2{,}8284$ V e $I_g = 353{,}55$ mA. A impedância do gerador é feita igual a zero para o caso de excitação por fonte de tensão, e igual a infinito para o caso da excitação por fonte de corrente. Sob as mesmas condições, os dois modos de excitação são comparados diretamente através das Figuras 12 e 13.

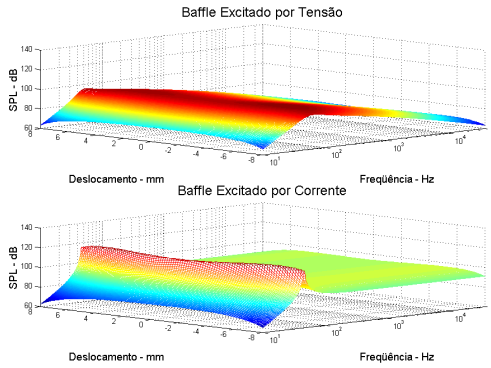


Figura 12. Nível de pressão sonora em função da freqüência e do deslocamento.

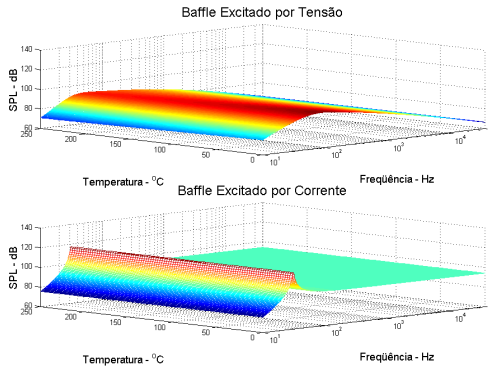


Figura 13. Nível de pressão sonora em função da freqüência e da temperatura.

Através da Figura 12, constata-se uma grande diferença entre os níveis de pressão sonora resultantes de ambos os modos de excitação. No caso da excitação por corrente, a freqüência de corte inferior situou-se em torno de 20 Hz, contra 40 Hz para o caso da excitação por tensão. Acima da região de ressonância, o nível de pressão sonora permaneceu constante para excitação por corrente, contra

uma queda gradual para excitação por tensão. Na região de ressonância, houve um aumento de aproximadamente 25 dB no nível de pressão sonora (excitação por corrente), o que é benéfico para a reprodução de sons em faixa estreita. No caso de se desejar reprodução em banda larga, com resposta maximamente plana, filtros equalizadores podem ser usados no sistema. Na Figura 13, tem-se, isoladamente, a influência da temperatura da bobina nos níveis de pressão sonora resultantes para os dois modos de excitação. Na excitação por fonte de corrente, não há influência da temperatura, como acontece na excitação por tensão. Isso se deve ao fato da corrente de excitação, no primeiro caso, não depender da impedância elétrica da bobina móvel.

5. CONCLUSÕES

Obter uma reprodução sonora fiel do sinal de áudio tem sido, há tempos, uma grande motivação para o aperfeiçoamento de sistemas de áudio. O alto-falante mantém os mesmos princípios de construção e funcionamento desde sua invenção. Novos métodos de projeto de caixas acústicas e de processamento dos sinais de excitação têm sido desenvolvidos. No entanto, um procedimento ainda não totalmente explorado é o modo de excitação por fonte de corrente. Nesse tipo de excitação, as distorções são reduzidas e abrem-se novas possibilidades de correções e compensações das não-linearidades, conforme discutido neste trabalho.

Uma nova abordagem à modelagem e ao estudo de alto-falantes e caixas acústicas tem sido apresentada, considerando a interrelacionada influência dos parâmetros não-lineares no sistema completo. Tal tratamento, em nosso entendimento, ainda não tinha sido apresentado na literatura. Essa unificação permitiu se ter uma visão mais ampla do comportamento de um *baffle* infinito excitado tanto por fonte de tensão quanto por fonte de corrente. O efeito térmico sobre a bobina é eliminado quando a excitação é através de fonte de corrente. A partir de tal análise, pode-se prever o comportamento não-linear do alto-falante e realizar uma prévia compensação do sistema através de processamento no sinal de excitação. Dessa forma, muitas das limitações discutidas na literatura especializada da área [2-6,9-18] poderiam ser superadas. Uma vez compensadas algumas dessas não-linearidades, poder-se-ia detectar outros efeitos não-lineares ainda não observados.

AGRADECIMENTOS

Os autores agradecem ao CNPq e à Studio R pelo financiamento desta pesquisa.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] BORTONI, Rosalfonso. *Análise, Dimensionamento e Avaliação de Estágios de Potência de Amplificadores de Áudio Classes A, B, AB, G e H*. Florianópolis, Abril de 1999. Dissertação (Mestrado em Engenharia Elétrica) – Centro Tecnológico, Universidade de Santa Catarina (UFSC).

[2] THIELE, A. Neville. "Loudspeakers in Vented Boxes: Parts I and II". *Loudspeakers Vol.1, An Anthology*, 2nd Edition, Audio Engineering Society, New York.

[3] SMALL, Richard H. "Direct-Radiator Loudspeaker System Analysis". *Loudspeakers Vol.1, An Anthology*, 2nd Edition, Audio Engineering Society, New York.

[4] SMALL, Richard H. "Closed-Box Loudspeaker Systems – Part I: Analysis, and Part II: Synthesis". *Loudspeakers Vol.1, An Anthology*, 2nd Edition, Audio Engineering Society, New York.

[5] SMALL, Richard H. "Vented-Box Loudspeaker Systems – Part I: Small-Signal Analysis, Part II: Large-Signal Analysis, Part III: Synthesis, and Part IV: Appendices". *Loudspeakers Vol.1, An Anthology*, 2nd Edition, Audio Engineering Society, New York.

[6] SILVA, Homero Sette. *Análise e Síntese de Alto-falantes & Caixas Acústicas pelo Método de Thiele-Small*. 1ª Impressão. Rio de Janeiro : H. Sheldon Serviços de Marketing Ltda., Inverno 1996.

[7] WERNER, R. E. and CARREL, R. M. "Application of Negative Impedance Amplifiers to Loudspeaker Systems". *Loudspeakers Vol.1, An Anthology*, 2nd Edition, Audio Engineering Society, New York.

[8] BORTONI, Rosalfonso. "Fator de Amortecimento. Uma Abordagem Introdutória". IV AES Brasil, São Paulo, Julho de 2000.

[9] STAHL, Karl Erik. "Synthesis of loudspeaker Mechanical Parameters by Electrical Means: A New Method for Controlling Low-Frequency Loudspeaker Behavior". *Loudspeakers Vol.2, An Anthology*, Audio Engineering Society, New York.

[10] NORMANDIN, R. "Extended Low-Frequency Performance of Existing Loudspeaker Systems". *Loudspeakers Vol.2, An Anthology*, Audio Engineering Society, New York.

[11] GANDER, Mark R. "Dynamic Lineariry and Power Compression in Moving-Coil Loudspeakers". *Loudspeakers Vol.2, An Anthology*, Audio Engineering Society, New York.

[12] SILVA, Homero Sette. "Variação da Resistência da Bobina Móvel em Função da Temperatura". IV AES Brasil, São Paulo, Junho de 2000.

[13] VANDERKOOY, John. "A Model of Loudspeaker Driver Impedance Incorporating Eddy Currents in the Pole Structure". *Loudspeakers Vol.2, An Anthology*, Audio Engineering Society, New York.

[14] WRIGHT, J. R. "An Empirical Model for Loudspeaker Motor Impedance". *Loudspeakers Vol.2, An Anthology*, Audio Engineering Society, New York.

[15] GANDER, Mark R. "Moving-Coil Loudspeaker Topology as an Indicator of Linear Excursion Capability". *Loudspeakers Vol.2, An Anthology*, Audio Engineering Society, New York.

[16] SILVA, Homero Sette. "O Alto-falante em Regime de Grandes Sinais". I AES Brasil, Rio de Janeiro, Outubro de 1996.

[17] MILLS, P. G. L. and HAWKSFORD, M. O. J. "Distortion Reduction in Moving-Coil Loudspeaker Systems Using Current-Drive Technology". *Loudspeakers Vol.3: Systems and Crossover Networks, An Anthology*, Audio Engineering Society, New York.

[18] BIRT, David R. "Nonlinearities in Moving-Coil Loudspeakers with Overhung Voice Coils". *Loudspeakers Vol.4: Transducers, Measurement and Evaluation, An Anthology*, Audio Engineering Society, New York.

[19] Woofer QCF® WPU1807. Especificações técnicas, Eletrônica Selenium S.A., Edição 00 – 04/01.

[20] GERGES, Samir N. Y. *Ruído: Fundamentos e Controle*. 1ª Edição, Imprensa Universitária, Universidade Federal de Santa Catarina, Florianópolis, 1992.

[21] BORTONI, Rosalfonso, FILHO, Sidnei Noceti and SEARA, Rui. *On The Design and Efficiency of Class A, B, AB, G and H Áudio Power Amplifier Output Stages*. Journal of The Audio Engineering Society, Vol. 50, No. 7/8, pp. 547-563, 2002 July/August.

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na VII Convenção Nacional
26-28 de maio de 2003, São Paulo, 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.

Transcrição Musical Automática com Bancos de Filtros

Cristiano N. dos Santos, Luiz W. P. Biscainho, Sergio L. Netto
Universidade Federal do Rio de Janeiro
Rio de Janeiro, 21945-970, RJ, Brasil
{csantos, wagner, sergioln}@lps.ufrj.br

RESUMO

Este trabalho investiga o uso de bancos de filtros FRM-CM como ferramenta auxiliar na transcrição automática de sinais musicais. Esses bancos possibilitam alta seletividade com elevado número de bandas, atingindo a alta resolução na freqüência requerida pela aplicação em questão. O artigo discute o problema da transcrição com suas dificuldades inerentes e compara o desempenho dos FRM-CMFBs com o da DFT na identificação de notas musicais.

I - INTRODUÇÃO

A transcrição de sinais de música [1] consiste em gerar, a partir de um sinal de áudio gravado, a representação da música executada numa forma que permita sua reprodução por um músico. A célula básica dessa representação é a nota musical. Assim, pode-se dizer que o coração de um sistema de transcrição é a identificação da altura, do tempo de início e da duração de cada nota emitida. A realização automática desse processo encontra aplicações que vão do auxílio ao ensino musical ao registro e estudo de interpretações de grande valor histórico-musical.

De modo geral, a identificação das notas musicais é feita a partir das informações espectrais do sinal em análise ao longo do tempo. Bancos de filtros [2] são um exemplo de ferramenta que permite a descrição dinâmica de sinais no domínio da freqüência de forma eficiente. Este trabalho analisa o uso de uma família específica de bancos de filtros no problema da transcrição de sinais musicais. Esses bancos se baseiam nas técnicas de mascaramento da resposta na freqüência (*frequency-response masking*, FRM) e de modulação por cossenos (*cosine-modulated filter bank*, CMFB), e possuem alta seletividade e grande número de bandas [3, 4].

A organização deste trabalho segue a seguinte estrutura: Na Seção II, apresenta-se o problema da transcrição musical, discutindo-se sua abrangência e revisando-se sua abordagem prática. Na Seção III, caracteriza-se o sistema de transcrição adotado neste trabalho. Na Seção IV, apresenta-se brevemente o banco de filtros FRM-CM usado neste trabalho. Na Seção V, compara-se o uso dos FRM-CMFBs com a DFT. Por fim, na Seção VI, apresenta-se um conjunto de experimentos demonstrando o uso dos FRM-CMFBs em diversas situações tópicas em transcrição.

II - O PROBLEMA DA TRANSCRIÇÃO

Uma das áreas mais abrangentes do processamento digital de sinais de áudio é a que se dedica à análise e à síntese de sinais musicais, a qual encontra aplicação na restauração e na remixagem de gravações, na síntese de instrumentos, na transcrição automática etc. De uma forma bem genérica, podemos dizer que os principais modelos utilizados em análise e síntese se enquadram em três categorias: puramente estocásticos (como o autorregressivo [5]), puramente

determinísticos (como o senoidal [6]) e simultaneamente estocásticos e determinísticos [6].

Neste trabalho, aborda-se particularmente o problema de transcrição musical automática [1]. Estritamente falando, a definição de sua meta poderia ser: “A partir de uma gravação musical, gerar uma partitura convencionalmente notada.” Nesses termos, a saída do sistema de transcrição deveria permitir a execução da peça musical por um músico sem nenhum treinamento adicional. Contudo, uma proposta genérica como essa pode atingir uma complexidade intratável, como veremos a seguir.

Deseja-se transcrever a peça musical com que grau de expressividade? Por exemplo, aspectos referentes à dinâmica (variações de intensidade) e à agógica (variações de velocidade) da execução musical podem estar associados à intenção do compositor (como ele as indicaria ao escrever a peça) ou à opção interpretativa do executante. Em alguns instrumentos, também o timbramento pode sofrer modificações essenciais (como no caso das diversas formas de se tanger uma corda de violão). Até que ponto é preciso notar os aspectos interpretativos contidos na gravação-fonte?

Outro aspecto de grande impacto na complexidade da tarefa de transcrição diz respeito aos tipos das fontes sonoras e à forma de encará-las. Por exemplo, pode-se considerar o caso de um instrumento solista; este pode ser monofônico (aqui no sentido de emitir uma única nota por vez) ou polifônico (capaz de emitir acordes). No caso de mais de um instrumento, além da classificação anterior, pode-se tratá-los como indivíduos solistas (como no caso de um quarteto de cordas) ou por famílias (como todos os segundos violinos de uma orquestra tocando em uníssono). A complexidade envolvida no reconhecimento de notas simultâneas emitidas por um mesmo (ou mesmo tipo de) instrumento é seguramente maior que no caso da emissão de uma nota por vez por tipo de instrumento. De qualquer modo, há que se determinar que notas são tocadas, e por quem.

Por último, tratar instrumentos temperados (no sentido de só serem capazes de emitir notas com afinação predeterminada, como um piano) permite simplificações que os demais instrumentos (como um violoncelo, em que se pode realizar um *glissando*) não admitem.

Claro que as soluções de melhor desempenho tendem a ser mais específicas; por outro lado, as soluções mais gerais têm o atrativo de serem mais automáticas. No final, a complexidade do sistema a projetar será determinada pela delimitação das metas a alcançar; vamos, então, dimensionar nossos objetivos e descrever as técnicas tipicamente associadas a eles.

III - METODOLOGIA BÁSICA

Primeiramente, a fim de não entrarmos no mérito da notação musical, assumiremos que se discute um sistema que tem como saída uma representação intermediária entre o sinal e a pauta, em princípio contendo em si todas as informações necessárias para gerá-la, se desejado. Como já foi dito, basta caracterizar as notas componentes do áudio sob análise. Aqui, cabe um comentário: a percepção musical pelo homem não se dá pela decomposição do áudio em suas notas individuais, mas a partir da cognição de entidades bem mais complexas e suas interrelações. Entretanto, se o objetivo final da transcrição é gerar uma representação segundo a notação musical convencional, a questão perceptiva perde bastante de sua importância na busca de soluções para este problema.

Outra simplificação que não prejudica tanto a generalidade da discussão é supor que os sinais são predominantemente "tonais" (mais uma vez, uma liberdade de terminologia): referimo-nos à presença exclusiva de fontes sonoras que emitem notas definidas, ao menos em regime permanente, excluindo, por exemplo, a maior parte dos instrumentos de percussão. Mais que isso, todos os instrumentos emitiriam sinais harmônicos, ou seja, compostos de uma componente fundamental numa freqüência f_0 (que definiria, afinal, a nota emitida) e componentes parciais em freqüências múltiplas inteiras de f_0.

Nesse contexto, um sistema típico de transcrição se preocuparia em reconhecer e descrever três aspectos: a entrada (*onset*) e duração de cada conjunto de notas, a individualização das notas com suas freqüências componentes e respectivas amplitudes e a identificação das fontes sonoras individuais, se for o caso.

A identificação do ataque de notas ou acordes geralmente é realizada a partir do exame da envoltória do sinal. A presença de picos na envoltória indica o início de novas emissões sonoras pelas fontes. Essa etapa permite extrair informações sobre a estrutura rítmica da música, que terá importância na sua notação final. A prévia separação do sinal em sub-bandas de freqüência permite, evidentemente, melhor desempenho, já que pode detectar melhor entradas com intensidades diferentes em faixas de freqüência distintas.

A parte mais importante do sistema de transcrição é a de descrição acurada do comportamento espectral do sinal no tempo. Isso envolve, basicamente, identificar cada linha espectral presente no sinal (incluindo todas as freqüências fundamentais e harmônicas que o compõem) e suas respectivas intensidades ao longo do tempo. As linhas precisam ser descritas de tal forma que se possa dizer quando nascem e morrem. O comportamento de uma fundamental e suas harmônicas associadas é coerente em regime permanente. Contudo, dependendo dos instrumentos tratados, as linhas podem variar continuamente em amplitude ou freqüência por efeitos como *tremolo* e *vibrato*. Num *glissando*, por exemplo, percorre-se uma série contínua de notas sem interromper a linha emitida. No caso de emissão simultânea de notas, é possível ocorrer o cruzamento e mesmo a superposição continuada de linhas de freqüência. Um dos casos mais complexos a enfrentar é a distinção entre outra fundamental presente na oitava superior de uma nota e sua segunda harmônica. Aqui, tudo se resume à busca de representações adequadas em tempo-freqüência, das quais as mais populares são a DFT e os bancos de filtros.

A DFT (*Discrete Fourier Transform*) fornece a descrição espectral de blocos de sinal, caracterizando módulo e fase de raias espectrais linearmente espaçadas. Associada ao modelo senoidal, que descreve cada linha freqüencial do sinal como uma soma de senóides harmônicas com amplitudes e fases lentamente variáveis, a DFT e suas variantes são a ferramenta preferida para implementar essa etapa dos sistemas de transcrição. Alternativamente, bancos de filtros permitem analisar continuamente no tempo a energia contida nas diversas regiões do espectro. Se os filtros individuais têm faixas de passagem suficientemente estreitas, permitem descrever acuradamente amplitude e freqüência ao longo do tempo para as componentes do sinal.

Dois aspectos ainda merecem ser mencionados: 1) A escala musical atualmente adotada no ocidente emprega temperamento igual, em que o menor intervalo (o semitom) corresponde a uma razão fixa de $2^{\frac{1}{12}}$ entre notas adjacentes. Então, mesmo que nos restrinjamos a instrumentos de afinação fixa, distinguir uma fundamental de outra exige acurácia melhor que 6%. 2) A importância da amplitude relativa das freqüências harmônicas está na identificação de instrumentos, quando for o caso: sua comparação contra um padrão pode fornecer uma pista muito importante.

Enquanto o problema de detecção de notas individuais já parece hoje satisfatoriamente solucionado, o polifônico ainda requer muito trabalho para ser resolvido. É nesse contexto que se insere este trabalho, onde propomos usar bancos de filtros de alta seletividade nessa etapa do processamento.

IV - O BANCO DE FILTROS FRM-CM

Os FRM-CMFBs são uma família de bancos de filtros que podem ser projetados com seletividade muito alta e elevado número de bandas. O detalhamento do projetos dos FRM-CMFBs pode ser encontrado, por exemplo, em [3] e [4].

Neste trabalho, utilizamos um banco de filtros de 1024 bandas com fator de *rolloff* ρ=0,1, atenuação máxima na banda passante A_p=0,2dB e atenuação mínima na banda de rejeição A_r=60dB . Com estas especificações, o projeto de um banco de filtros convencional seria impraticável. A resposta em magnitude deste FRM-CMFB é detalhada na Fig.1.

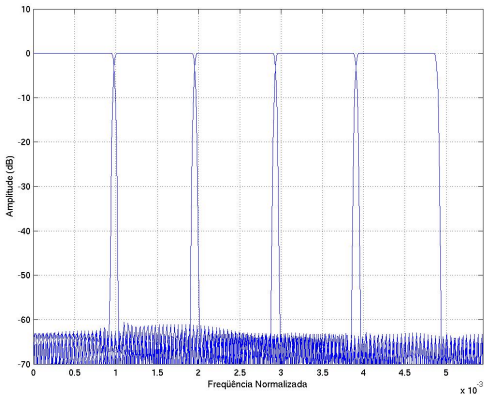


Figura 1: Resposta em magnitude do FRM-CMFB com 1024 em bandas (detalhe das 5 primeiras bandas).

V - FRM-CMFB X DFT

A separação de harmônicos é de fundamental importância na transcrição musical para garantir uma identificação segura das notas e das famílias de intrumentos. Esta separação deve ocorrer em dois níveis: devemos separar os harmônicos uns dos outros e também do ruído presente nas bandas vizinhas do sinal. O uso do FRM-CMFB procura resolver duas dificuldades na separação de harmônicos, comuns na tarefa de transcrição musical.

A primeira dificuldade é a da interferência entre bandas adjacentes. Analisando-se as transformadas mais usadas (DFT, CQT - *Constant-Q Transform* e BQT - *Bounded-Q Transform* [1]) como bancos de filtros, podem-se observar características comuns, como a baixa atenuação das bandas de rejeição de seus filtros. A conseqüência disso é a interferência de informação musical e/ou ruidosa entre bandas adjacentes, podendo tornar difícil a identificação de picos no domínio da freqüência. A Fig. 2 mostra a resposta em magnitude do banco de filtros correspondente à DFT, onde fica evidente a interferência entre bandas adjacentes devido à baixa atenuação (apenas 13 dB) na banda de rejeição dos subfiltros DFT. Comparando-se as Figs. 1 e 2, vê-se que a separação de harmônicos em bandas próximas deverá ser melhor realizada pela estrutura FRM-CMFB.

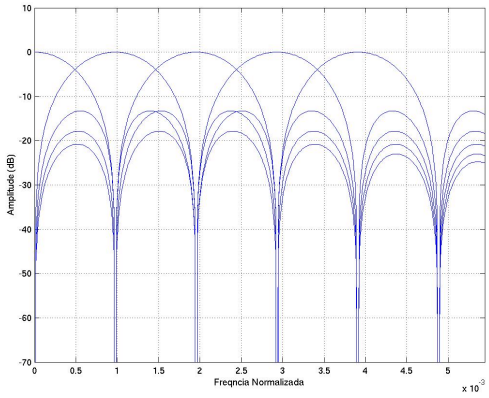


Figura 2: Resposta em magnitude da DFT com 1024 em bandas (detalhe das 5 primeiras bandas).

Em relação à DFT, cabe destacar ainda a presença de um *ripple* significativo na banda passante de seus filtros, o que pode gerar uma atenuação diferenciada das componentes de freqüência, dependendo das posições dos harmônicos em relação ao centro da banda passante. Este aspecto é crítico, por exemplo, na identificação de instrumentos musicais.

A segunda dificuldade diz respeito ao uso de transformadas logarítmicas, como a CQT e a BQT. Estas transformadas se destinam a separar faixas de freqüências a intervalos geométricos. Assim sendo, se é adotada uma resolução de um semitom, para cada nota da escala cromática há uma faixa entre um quarto de tom abaixo e um quarto de tom acima que é separada naquela banda. se a resolução é de quarto de tom, haverá duas faixas a cada nota, e assim por diante. Isso permite ter maior eficiência na distribuição de amostras da transformada, de acordo com a escala ocidental, logarítmica. Porém, essa distribuição tende a agrupar em uma mesma banda os harmônicos de notas diferentes que resultem muito próximos. Conseqüentemente, embora distintos, estes não são identificáveis individualmente.

Para exemplificar esse problema, vamos considerar o caso de um intervalo recorrente na música ocidental, a terça maior. Tomando-se a nota Dó como padrão, sua terça maior superior corresponde ao próximo Mi. Conforme a escala natural, a razão entre suas freqüências deveria ser de 4 para 5, o que levaria o quinto harmônico de Dó a coincidir exatamente com o quarto harmônico de Mi. Podem-se imaginar, em situações de mais notas simultâneas, as ambigüidades bastante complexas que essas superposições podem provocar. Entretanto, na escala de temperamento igual, a terça maior tem um erro de 0,8% em relação à razão de inteiros. Com Dó=262 Hz e Mi=330 Hz, esse desvio faria os dois

harmônicos citados distarem de 10,5 Hz. Uma acurácia suficientemente alta já resolveria a ambigüidade anterior, ao menos para instrumentos de afinação fixa e temperamento igual. As Figs. 3 e 4 mostram, respectivamente, as respostas da BQT, com resolução de um quarto de tom, e do FRM-CMFB, com resolução de 5,4 Hz, a um sinal com freqüências no intervalo musical descrito acima.

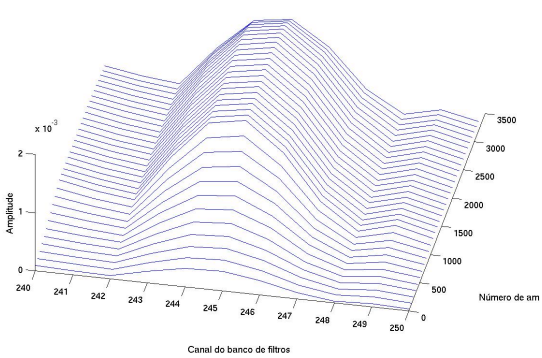


Figura 3: Envoltória da resposta da BQT no intervalo da terça maior.

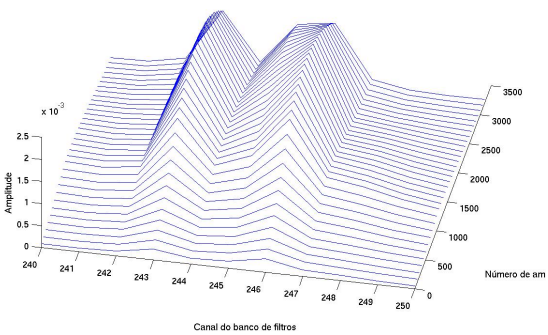


Figura 4: Envoltória da resposta do FRM-CMFB no intervalo da terça maior.

VI - EXEMPLOS DE TRANSCRIÇÃO

Exemplo 1 - Harmônicos Próximos:

O problema de interferência entre bandas pode ser visualizado a partir de um simples exemplo. Considere o caso em que há dois harmônicos situados um na banda *i* e o outro na banda *i*+2, deixando a banda do meio, *i*+1, sem informação relevante. Em caso de baixa atenuação nas bandas adjacentes (como na DFT), os harmônicos nas bandas *i* e *i*+2 podem se mesclar na banda *i*+1, confundindo o processo de transcrição.

Este fenômeno pode ser ilustrado usando-se uma senóide de 7 Hz a mais do centro da primeira banda e outra senóide de 7 Hz a menos do centro da terceira banda. A Fig. 5 mostra a envoltória das saídas de cinco bandas determinadas com a DFT. Note que neste caso não foi possível visualizar os dois harmônicos separadamente devido à baixa atenuação da DFT. O resultado do mesmo experimento usando o FRM-CMFB é visto na Fig. 6, onde podemos claramente perceber a existência dos dois harmônicos.

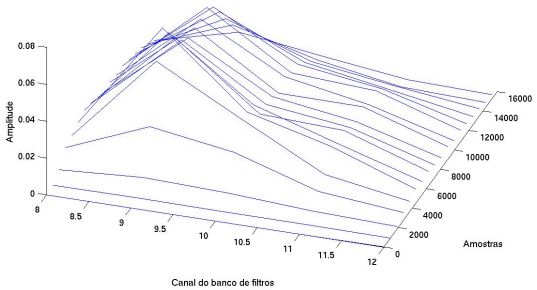


Figura 5: Exemplo 1- Envoltória dos sinais no banco de filtros DFT para o caso de dois harmônicos próximos.

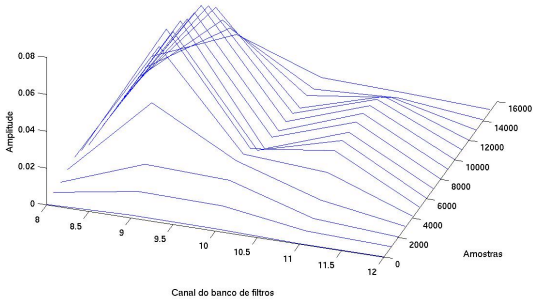


Figura 6: Exemplo 1- Envoltória dos sinais no FRM-CMFB para o caso de dois harmônicos próximos.

Exemplo 2 - Transitórios no Tempo:

Ao se filtrar um sinal do tipo musical, com componentes de freqüência bem definidas, deseja-se obter na saída a(s) componente(s) de freqüência correspondente(s) à banda deste filtro. Mas a filtragem altera a forma de onda, impondo transitórios. Os transitórios podem ser divididos em dois tipos: de subida e de estabilização.

Denominamos transitório de subida aquele em que a amostra central da resposta ao impulso do filtro ainda não chegou à amostra inicial do trecho estacionário. Denominamos de transitório de estabilização aquele em que a amostra central do filtro já ultrapassou a amostra inicial do trecho estacionário do sinal. Este transitório tem comprimento diretamente relacionado ao comprimento da resposta ao impulso do filtro.

De modo geral, os transitórios podem ser processados a ponto de não interferirem no problema de transcrição. Por exemplo, podemos cortar o transitório de subida através de um pré-processamento de detecção de inícios de notas, como observado nas Figs. 7 e 8, reduzindo-se, assim, a interferência entre notas subseqüentes numa mesma banda.

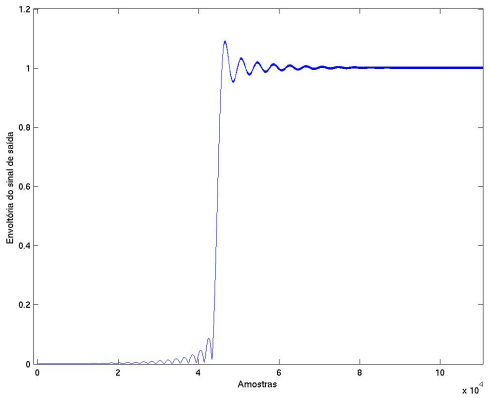


Figura 7: Exemplo 2 - Amplitude da resposta completa de uma banda do FRM-CMFB a uma senóide.

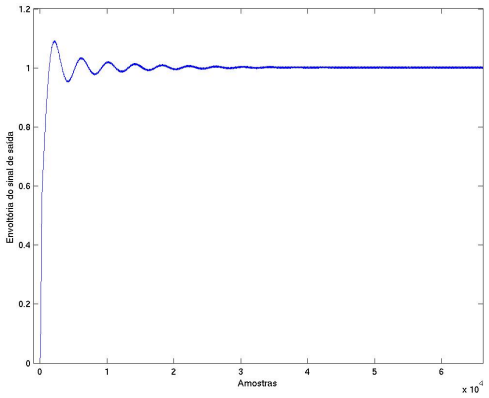


Figura 8: Exemplo 2 - Amplitude da resposta sem o transitório de subida de uma banda do FRM-CMFB a uma senóide.

Exemplo 3 - Freqüência Variável:

Uma forma simples de visualizar a acurácia de representação que se pode alcançar com os bancos de filtros é testá-los com uma senóide de freqüência variável. Neste exemplo, usamos uma senóide com sua freqüência variando linearmente de (f_c-20)Hz a (f_c+20)Hz, onde f_c é a freqüência central de uma banda do FRM-CMFB. Com isto, o sinal analisado consistia numa única freqüência variando no tempo ao longo de três bandas distintas do banco de filtros.

A parte superior da Fig. 9 mostra a variação de freqüência presente no sinal, atravessando as linhas horizontais que delimitam as bandas dos filtros do banco. A parte inferior da Fig. 9 mostra a envoltória das respostas a este sinal quando convoluído com os filtros de interesse neste exemplo. O resultado mostra como o banco de filtros foi capaz de perceber a banda correta da freqüência do sinal de entrada, inclusive determinando corretamente os momentos nos quais a freqüência mudava de banda dentro do banco de filtros.

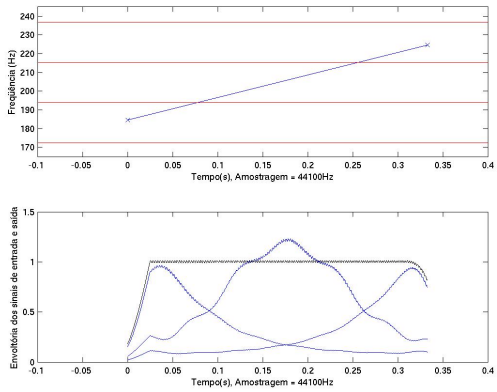


Figura 9: Exemplo 3 - (a) Variação da freqüência do sinal de entrada (as linhas horizontais representam limites de bandas no banco de filtros); (b) Envoltória de amplitude nas saídas das bandas correspondentes à parte (a).

VII - CONCLUSÃO

Neste trabalho, expusemos algumas premissas teóricas do problema da transcrição musical. Discutimos, ainda, o uso de bancos de filtros no reconhecimento de notas musicais, em particular a estrutura FRM-CMFB, ressaltando sua vantagem em relação à DFT quanto à seletividade no domínio da freqüência. Por fim, apresentamos exemplos práticos do problema da transcrição musical, enfatizando os aspectos da resolução no domínio da freqüência e dos transitórios no domínio do tempo.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] A. Klapuri, *Automatic Transcription of Music*, M.Sc. Thesis, Tampere University of Technology, Finland, Nov. 1997.

[2] S. R. Diniz, E. A. B. da Silva, S. L. Netto, *Digital Signal Processing: System Analysis and Design*, Cambridge, UK, 2002.

[3] P. S. R. Diniz, L. C. R. de Barcellos, S. L. Netto, "Design of cosine-modulated filter bank prototype filters using the frequency-response masking approach," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, UT, USA, May 2001.

[4] S. L. Netto, P. S. R. Diniz, L. C. R. de Barcellos, "Efficient implementation for cosine-modulated filter banks using the frequency-response masking approach," *Proc. IEEE International Symposium on Circuits and Systems*, Scottsdale, AZ, USA, vol. III, pp. 229-231, May, 2002.

[5] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, USA, 3.ed.., 1991.

[6] T. F. Quatieri, R. J. McAulay, "Audio signal processing based on sinusoidal analysis/synthesis," in *Applications of Digital Signal Processing to Audio and Acoustics*, eds. M. Kahrs, K. Brandenburg, Kluwer, 1998.

[7] S. W. Foo, W. T. Lee, "Application of fast filter bank on transcription of polyphonic signals," to appear in *Journal on Circuits, Systems and Computers*, vol. 12, no. 5, Oct. 2003 (expected).

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na VII Convenção Nacional
26-28 de maio de 2003, São Paulo, 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.

Sistema de Efeitos para Guitarra Aplicados em Tempo Real Utilizando DSP

Rodrigo Coura Torres e José Manoel de Seixas
Laboratório de Processamento de Sinais (LPS) COPPE / EE / UFRJ
Rio de Janeiro, CP 68504, RJ, Brasil
{torres, seixas}@lps.ufrj.br

RESUMO

A diversidade de sonoridades encontradas atualmente nos instrumentos deve-se principalmente ao processamento aplicado ao sinal oriudo do instrumento musical, antes de ser enviado ao sistema de alto-falantes. Neste aspecto, processamento digital de áudio utilizando DSPs surge como uma ferramenta poderosa. Assim sendo, um sistema de efeitos para guitarras foi implementado utilizando um DSP da *Analog Devices*, modelo ADSP-21065L. O sistema implementou dois efeitos muito utilizados por guitarristas: distorção e eco. Como resultados, são apresentadas as formas de ondas temporais e espectrais de cada efeito, bem como uma análise subjetiva dos mesmos.

1 Introdução

Como sabemos, processamento digital de sinais aparece em várias áreas de entretenimento atualmente [1]. Podemos observar sua valiosa contribuição em filmes, músicas, videogames, etc. Particularmente na música, esta ciência tem sido explorada com bastante sucesso, permitindo desde produções musicais mais elaboradas até a restauração de músicas armazenadas em meios deterioráveis com o tempo, como as velhas fitas magnéticas.

Este trabalho está focado em apresentar um sistema digital de processamento de áudio, mais particularmente destinado a tocadores de guitarra. Guitarristas são os músicos que mais exploram recursos que visam alterar a sonoridade de seu instrumento, visando dar mais vida a música. Uma guitarra difere-se sonoramente da outra em uma pequena parte pelo sistema de captação[1]. Entretanto, percebemos claramente as diferenças de sonoridade quando ouvimos as notas emitidas pelas guitarras de músicos como Gilberto Gil (MPB) e Jeff Hanneman (guitarrista do *Slayer*, uma banda de *Heavy Metal*). O que torna esta sonoridade tão diferente é justamente o processamento ao qual o sinal gerado pela guitarra é submetido antes de ser enviado aos alto-falantes. Assim sendo, guitarristas, ao longo de um *show*, por exemplo, utilizam-se de um conjunto de pedais, cada um responsável por alterar o sinal proveniente da guitarra de um modo específico, de forma a enriquecer a melodia.

O trabalho foi implementado utilizando-se um DSP, modelo ADSP-21065L, montado em uma placa (EZ-KIT Lite) que possibilita a comunicação das funções internas do mesmo com o usuário. A realização prática do trabalho consistiu em implementar nesta placa dois efeitos tradicionais de guitarra: distorção harmônica e eco, de forma a simular

[1] Sistema de captação é basicamente composto de uma ou mais bobinas posicionadas embaixo das cordas e que convertem a vibração das mesmas em sinal elétrico.

um sistema de pedais utilizado por guitarristas para aplicar estes efeitos[2].

Este trabalho será apresentado da seguinte forma: a sessão 2 abordará as principais características do processador utilizado neste trabalho, bem como descreverá alguns dos ambientes de desenvolvimento utilizados para a elaboração do mesmo. A sessão 3 apresenta a modelagem matemática de cada efeito, explicando os conceitos matemáticos envolvidos que realizam as alterações sonoras desejadas. Em seguida, a sessão 4 apresentará a forma de implementação deste trabalho no DSP, e descreverá alguns aspectos práticos de implementação. A seguir, a sessão 5 apresentará os resultados obtidos com o sistema de efeitos, e por fim, as conclusões serão apresentadas na sessão 6.

2 O Processador ADSP-21065L

O ADSP-21065L da família SHARC é um processador de sinais digitais de 32 bits de alta performance, utilizado em aplicações de telecomunicações, áudio digital, e instrumentação industrial, por exemplo [2].

Junto com um núcleo de processamento de alta performance (180 MFLOPS), este processador possui uma memória dual SRAM interna e periféricos de entrada/saída integrados em um processador de I/O dedicado. Com seu próprio *cache* de instruções interno, o processador pode executar cada instrução em um único ciclo. O ADSP-21065L possui código totalmente compatível com outros membros da família SHARC, o que possibilita *upgrades* de sistema sem maiores complicações. A família SHARC possui como principal característica quatro barramentos para acesso dual a dados, instruções e requisições de entrada/saída.

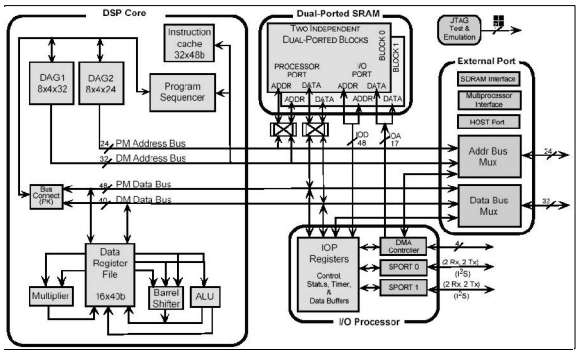


Fig. 1. Diagrama de blocos do ADSP-21065L.

A Fig. 1 apresenta os barramentos internos do ADSP-21065L. Temos o barramento PM (*Program Memory*), constituído pelos barramentos PMA (*Program Memory Address*) e PMD (*Program Memory Data*). Além deste, temos também o barramento DM (*Data Memory*), constituído pelos barramentos DMA (*Data Memory Address*) e DMD (*Data Memory Data*). Por fim, temos o barramento de entrada/saída (*I/O Bus*), constituído pelos barramentos IOA (*I/O Address*) e IOD (*I/O Data*).

O barramento PM pode acessar tanto dados como instruções. Durante um único ciclo, o processador pode acessar dois operandos de dados, um pelo barramento PM e outro pelo barramento DM, acessar uma instrução no *cache* interno e ainda executar uma transferência via DMA.

As portas externas do ADSP-21065L provêem a comunicação do processador com uma memória externa, o que é feito de forma direta, se for usada uma memória SDRAM, um processador hospedeiro ou outro processador ADSP-21065L. Esta porta externa realiza o controle interno do barramento e provê sinais de controle para a memória compartilhada, global, ou dispositivos de entrada/saída.

2.1 Características e Benefícios do ADSP-21065L

O ADSP-21065L possui os seis requerimentos básicos para pertencer a família ADSP-2106x de DSPs de 32 bits de ponto flutuante:

- **Unidades computacionais aritméticas rápidas e flexíveis:** O ADSP-21065L executa todas as instruções em um único ciclo. Junto com operações aritméticas tradicionais (multiplicação, adição, subtração e multiplicação acumulada), ele ainda possui um conjunto completo de operações aritméticas, incluindo *min*, *max*, *1/x*, *1/sqrt(x)*, etc. O ADSP-21065L é compatível com o padrão de ponto flutuante da IEEE e ainda permite tratamento de interrupções e exceções.
- **Fluxo de dados sem restrições de/para as unidades computacionais:** O ADSP-21065L possui a arquitetura da família SHARC, combinada com um arquivo de registro de dados de 10 portas, de tal maneira que, em um único ciclo, o processador pode:
  - Ler ou escrever dois operandos no arquivo de registro.
  - Fornecer dois operandos para a ALU.
  - Fornecer dois operandos para o multiplicador e,
  - Receber dois resultados da ALU e do multiplicador.

  A palavra de instrução ortogonal de 48 bits do processador permite total transmissão paralela de dados e operações aritméticas na mesma instrução.
- **Precisão estendida e larga faixa dinâmica nas unidades computacionais:** O ADSP-21065L opera no formato de ponto flutuante de 32 bits no padrão da IEEE, 32 bits em ponto fixo, complemento a dois e *unsigned*. Além disso, possui precisão estendida de 40 bits (formato IEEE) em ponto flutuante. O processador propaga esta precisão estendida através de suas unidades computacionais, limitando erros intermediários de truncamento. Quando trabalhando com dados internos, o processador pode transferir a mantissa de precisão estendida de 32 bits de e para todas as unidades computacionais. Os formatos em ponto fixo possuem um acumulador de 80 bits para cálculos de 32 bits em ponto fixo.
- **Geradores de endereços duais:** O ADSP-21065L possui dois geradores de endereços (DAGs) que

[2] Como guitarristas estão com ambas as mão ocupadas durante a música, os efeitos são acionados pelos pés. Daí o tradicional nome "pedaleira de efeitos". No nosso caso, a alteração de um efeito para outro será feita manualmente, dada a fragilidade da placa a ser utilizada e do produto ainda ser um protótipo.

permitem endereçamento imediato ou indireto. Eles permitem operações de módulo e bit-reverso, sem restrições na localização do *buffer* de dados.

- **Seqüenciamento de programa eficiente:** Em adição aos *loops* sem *overhead*, o ADSP-21065L permite configurações e saídas de *loops* em um único ciclo. Os *loops* podem ser cascateados (seis níveis em *hardware*) e interrompidos.
- **Busca Serial e Capacidade de Emulação:** O ADSP-21065L aceita o padrão IEEE P1149.1 *Joint Test Action Group* (JTAG) para testes de sistema. Este padrão define um método para buscar serialmente o estado de I/O de cada componente do sistema. O emulador EZ-ICE do ADSP-21065L também usa a porta serial JTAG para acessar as características de emulação interna do processador.

2.2 Placa de Desenvolvimento ADSP-21065L EZ-KIT Lite

Esta placa permite ao desenvolvedor explorar os recursos do ADSP-21065L, servindo de *interface* entre o processador e o desenvolvedor.

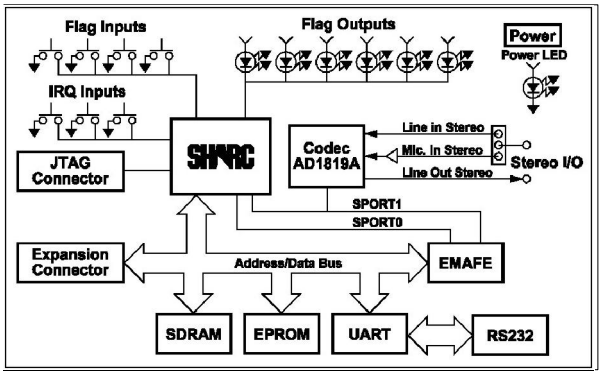


Fig. 2. Diagrama de blocos do sistema EZ-KIT Lite.

A Fig. 2 mostra bem a principal função desta placa, que é ser o meio de comunicação do DSP com o mundo exterior. O resto desta sessão visará descrever brevemente algumas das principais funcionalidades que esta placa possui. A Fig. 3 apresenta a localização de cada funcionalidade na placa.

2.2.1 Memória não volátil (EPROM) para *boot*

Esta funcionalidade permite que um total de 1M x 8 bits de programas armazenados possam ser carregados pelo DSP, quando configurado para *boot* via EPROM. A seleção do modo de *boot* é feita pelos *jumpers* BMS (*Boot Memory Select*) e BSEL. neste modo de *boot*, as primeiras 256 instruções (1536 bytes) são automaticamente carregadas pelo processador após o *reset*. O resto do código (se houver) precisa ser carregado pelo código contido nestas primeiras 256 instruções. Maiores detalhes podem ser encontrados em [3].

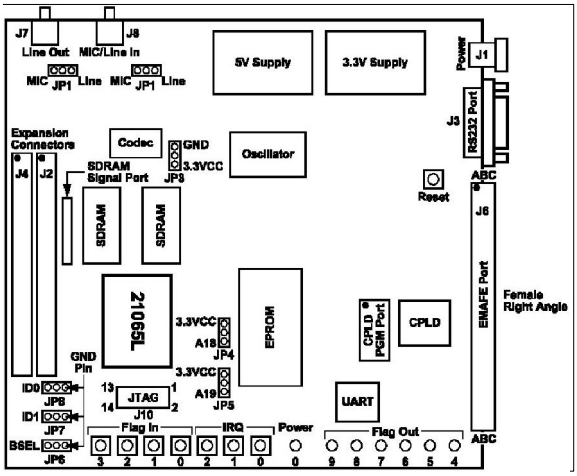


Fig. 3. Esquema da placa EZ-KIT Lite.

2.2.2 Botões de I/O

Para permitir ao usuário maior controle de operação do processador, existem 8 botões nesta placa: RESET, FLAG 0-3 e IRQ0-2.

- O botão de RESET permite reinicializar o DSP. Se o usuário perder contato entre o DSP e o computador hospedeiro, enquanto rodando um programa, este botão permite restabelecer a comunicação entre ambos.
- Os botões de FLAG 0-3 trocam o estado de quatro pinos de *flag* (FLAG 0-3) para o DSP.
- Os botões de IRQ 0-2 permitem enviar interrupções (IRQ 0-2) para o DSP. Esta interrupção manual permite a execução de uma rotina de interrupção pelo DSP durante a execução de um programa. A IRQ0 é dividida com o UART e a IRQ1 com o conector EMAFE.

2.2.3 LEDs de Usuário

Existem seis LEDs na placa para enviar informações para o usuário. Os LEDS (FLAG 4-9) são controlados pelas saídas de FLAG do DSP e são nomeados de acordo com a saída de *flag* que os controlam.

2.2.4 CODEC de Áudio

A placa possui um conversor A/D e D/A com freqüência máxima de amostragem de 48 kHz, tornando esta placa bastante útil para processamento de áudio em tempo real. Esta funcionalidade é implementada pelo circuito integrado modelo AD1819 da *Analog Devices*, e o mesmo comunica-se com o DSP pela porta serial (SERIAL 1).

2.2.5 *Interface* de Comunicação EMAFE

O conector EMAFE (*Enhanced Modular Analog Front End*) permite a *interface* padrão necessária para conexão com dispositivos externos. Este conector é composto por 96 pinos fêmeas organizados em 3 fileiras de 32 pinos cada. A *interface* aceita um *data path* de 16 bits em paralelo, duas portas seriais, uma saída de interrupção, e uma entrada de *flag*. Maiores detalhes podem ser vistos em [4].

3 Análise Matemática dos Efeitos

Nesta sessão, será apresentada a análise matemática envolvida na produção dos dois efeitos implementados neste trabalho. Para a realização desta análise, foi sempre considerado a utilização de uma senóide, por ser o tom mais puro conhecido. Esta senóide tem freqüência de 200 Hz.

3.1 Geração da Distorção Harmônica

Como o próprio nome indica, este efeito gera como resultado final uma forma de onda distorcida. O princípio básico deste efeito consiste em produzir harmônicos nos múltiplos da freqüência fundamental, de forma que estes harmônicos gerem distorção no sinal, mas sem alterar a nota musical que foi tocada. Com isso, devemos tomar cuidado para que a única distorção gerada seja harmônica, e não intermodular[3].

Assim, o principal método de distorção se resume em "truncar" a forma de onda, pois como sabemos pela teoria de séries de Fourier, esta deformação abrupta na forma de onda será representada na freqüência por harmônicos em freqüências mais altas. Como estão todos em múltiplos inteiros da freqüência fundamental, temos assim o efeito de distorção harmônica. A Fig. 4 apresenta o resultado deste processo.

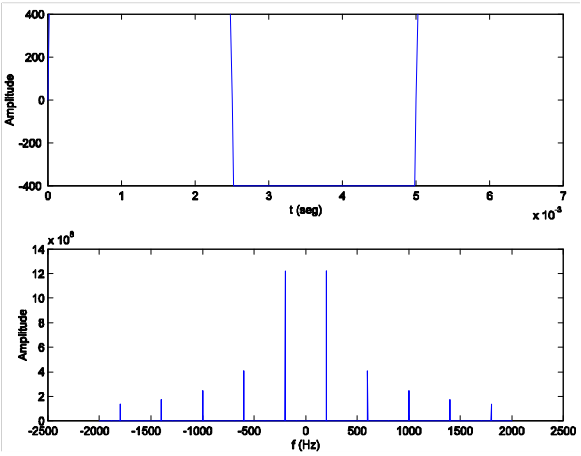


Fig. 4. Apresentação temporal e espectral da senóide após a distorção.

Outra técnica utilizada consiste em aplicar um ganho não linear as amostras do sinal. Assim sendo, cada amostra que chega é multiplicada pelo *log* da mesma, de forma a realçar as amostras de baixo valor, aumentando o tempo de decaimento da nota tocada.

3.2 Geração do Eco

Enquanto o efeito de distorção harmônica baseia-se em técnicas não lineares de implementação, o efeito de eco nada mais é do que um filtro linear de resposta infinita (IIR) [6] de função de transferência igual a apresentada na eq. 1.

$$H(Z) = \frac{1}{1 - g.Z^{-T_{eco}.f}} \qquad (1)$$

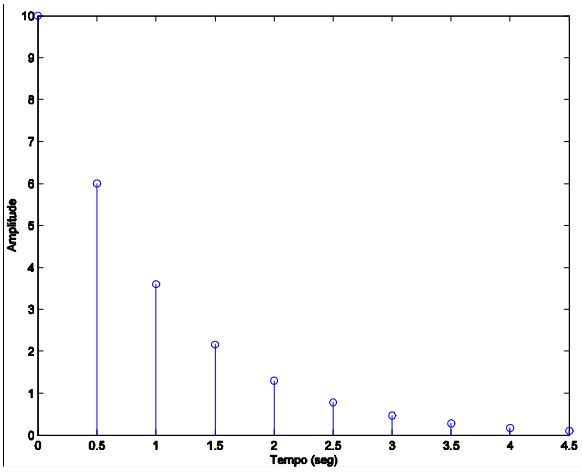


Fig. 5. Resposta ao impulso do sistema gerador de eco.

Onde g é o ganho dado ao eco ($|g| < 1$, para garantir estabilidade), T_{eco} é o período do eco e f é a freqüência de amostragem do conversor A/D. Assim, no tempo, teríamos a seguinte expressão:

$$y[n] = x[n] + g.y[n - T_{eco}.f] \qquad (2)$$

Neste caso, como estamos realimentando o sistema com amostras passada da saída, geramos eco, e ecos do eco. Caso o sistema fosse implementado por um filtro FIR, ou seja, se ao invés de amostras passadas da saída estivéssemos usando amostras da entrada, teríamos um *Delay*, que seria a reprodução única de uma amostra atrasada. Fig. 5 apresenta a resposta ao impulso para o caso onde: g = 0.6, T_{eco} = 0.5 seg e f = 48 kHz.

4 Implementação Prática

Nesta sessão, serão apresentados os procedimentos necessários para que o DSP pudesse implementar os efeitos discutidos na sessão 3. Para tal, selecionamos os seguintes parâmetros para os efeitos:

1. Distorção:
   - As amostras do sinal saturam em aproximadamente 40% do valor máximo (em módulo), o que gera uma distorção bastante forte.
2. Eco:
   - O ganho do eco (g) é igual a 0.6.
   - O Atraso do eco (T_{eco}) é de 0.5 seg.
   - A freqüência de amostragem (f) é de 48 kHz, o que gera um filtro IIR com T_{eco} x f = 24000 posições de memória (atraso).

[3] Distorção intermodular ocorre quando duas notas diferentes são tocadas ao mesmo tempo, gerando componentes nas freqüências iguais a soma e a diferença das freqüências das notas tocadas (fenômeno de batimento [5]).

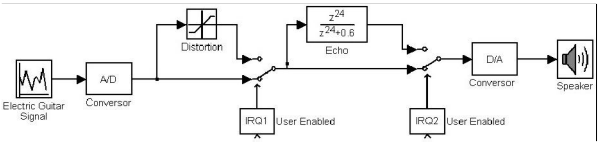


Fig. 6. Diagrama de blocos do trabalho desenvolvido.

A Fig. 6 apresenta o diagrama de blocos do trabalho. Como podemos observar, após a aquisição e digitalização do sinal pelo conversor A/D, as amostras podem seguir quatro caminhos distintos até a saída. O caminho a ser seguido dependerá da opção do usuário, que controla o sistema pelos botões de ativação das IRQs 1 e 2. Cada vez que uma dada IRQ é disparada, a chave correspondente muda de estado, mudando a trajetória, e conseqüentemente o efeito aplicado as amostras. Assim, podemos ter os dois efeitos combinados, apenas um, ou nenhum, dando maior controle ao sistema. Os *flags* 4 e 5 (representados por LEDs) são utilizados para indicar ao usuário se a distorção ou o eco, respectivamente, estão ativados (LED aceso significa que o efeito está ativo). Com isso, a qualquer momento, o usuário sabe exatamente o estado em que o sistema se encontra. Por fim, após o tratamento das amostras, as mesmas passam pelo conversor D/A, que por fim, enviará o sinal analógico resultante para o alto-falante.

4.1 Implementação do Sistema

O sistema foi implementado utilizando-se a linguagem C^{++} e *assembly*. O sistema possui uma estrutura seqüencial (de acordo com a Fig. 6), onde a saída do primeiro efeito é a entrada do segundo. Cada efeito pode ser desativado, gerando quatro estados distintos do sistema.

A chave da operação do sistema são duas variáveis booleanas. Cada uma responsável por indicar se um dado efeito está ou não ativo. A cada amostra que chega, o programa verifica o valor de cada variável, e se o mesmo for verdadeiro, a função correspondente é chamada para aplicar o efeito na amostra. Após as duas verificações, a amostra tratada (ou não, dependendo do valor das variáveis) é enviada ao conversor D/A, para que possa ser ouvida pelo usuário. Quando o usuário pressiona o botão de um efeito qualquer, a interrupção disparada por este botão chama uma função que altera o valor booleano da variável correspondente àquele efeito (e troca também o valor do respectivo LED), fazendo com que a cada vez que o botão seja pressionado, o efeito seja (ou não) ativado.

5 Resultados

Como formas de resultado, serão apresentadas as formas de onda e o espectro em freqüência do sinal após a aplicação de cada efeito separadamente. Primeiramente, na Fig. 7, podemos observar a forma de onda e o espectro em freqüência de uma nota tocada isoladamente e sem nenhum efeito aplicado.

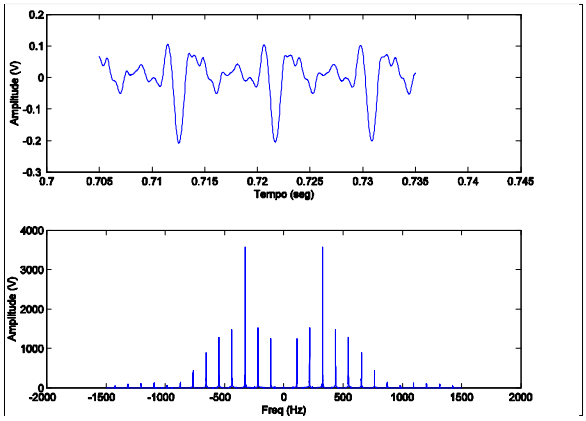


Fig. 7. Forma de onda e espectro de uma nota sem nenhum efeito.

Na Fig. 8 podemos observar as alterações surgidas quando aplicamos o efeito de distorção a nota. Percebemos o aumento de harmônicos, gerados pela "clipagem" da forma de onda (efeito este que pode ser observado na forma de onda), gerando a distorção harmônica desejada. Através do ganho não linear aplicado, percebe-se também que o decaimento da amplitude do sinal é mais lento, como esperado.

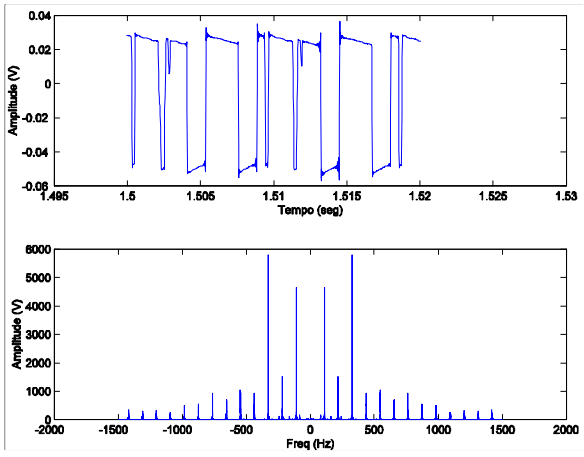


Fig. 8. Forma de onda e espectro de uma nota com distorção aplicada.

Na Fig. 9 observamos o resultado quando aplicamos o efeito de eco no sinal. Observa-se no espectro que não há geração de novos harmônicos, como esperado, havendo apenas uma pequena variação de amplitude dos harmônicos já existentes na nota sem efeito. No domínio do tempo, a variação da forma de onda também é muito pequena, contribuindo para manter a fidelidade do sinal.

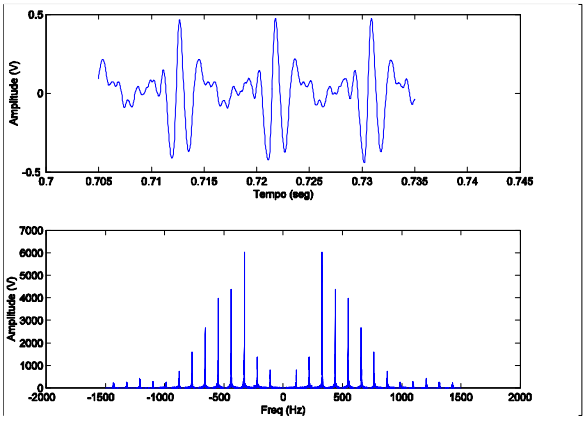


Fig. 9. Forma de onda e espectro de uma nota com eco aplicado.

Por fim, na Fig. 10, observamos as formas de onda de todos os três casos abordados durante todo o período de gravação. A maior importância nesta figura está na forma de onda do eco, onde percebemos as variações de amplitude que ocorrem a cada meio segundo, mostrando assim a presença do eco. A figura do meio apresenta o sinal distorcido, que como percebemos ficou bem diferente da forma de onda original.

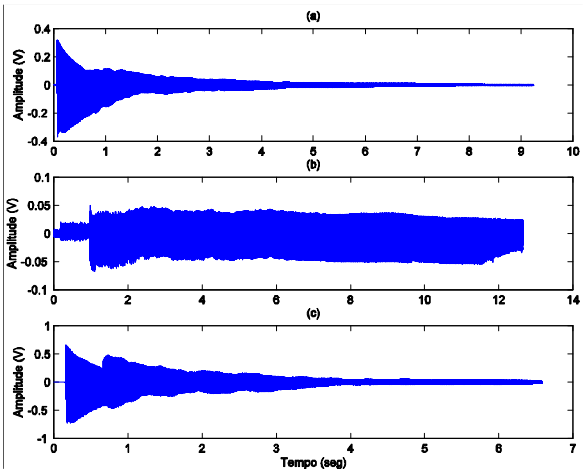


Fig. 10. Forma de onda do sinal sem efeito (a), com distorção (b) e com eco (c).

Em *www.lps.ufrj.br/~torres* estão os arquivos sonoros com os resultados deste trabalho para avaliação do sistema proposto. Como forma de análise subjetiva, o sistema foi apresentado a um grupo de 10 pessoas, que continha tanto músicos experientes como leigos. Todos os avaliadores fizeram sugestões que foram levadas em consideração no desenvolvimento do projeto, até que a qualidade dos resultados estivesse do agrado geral, fazendo assim, com que este produto possa, com a adição de outros recursos, atingir o estágio de comercialização.

6 Conclusão

Foi desenvolvido um sistema de efeitos para guitarra atuando em tempo real numa placa contendo um processador DSP da *Analog Devices*. O trabalho apresentou os conceitos matemáticos envolvidos na elaboração dos dois efeitos implementados, além de abordar aspectos práticos de implementação destes efeitos em DSP.

Visando facilitar novos trabalhos na área de processamento de áudio utilizando o CODEC da placa, uma classe em C^{++} foi desenvolvida, permitindo uma utilização mais simples deste recurso.

6.1 Trabalhos Futuros

Como propostas de trabalhos futuros, poderia ser implementado um controle de volume para o sinal de saída do CODEC. Outra proposta poderia ser utilizar os *flags* de entrada para controle de parâmetros dos efeitos (atraso do eco, nível de distorção, etc), caminhando assim, rumo a aplicação comercial para este trabalho.

Agradecimentos

Gostaríamos de agradecer à CAPES, FAPERJ e CNPq por todo apoio oferecido para a realização deste trabalho.

Referências

[1] TOHYAMA, M., KOIKE, T., *Fundamentals Of Acoustic Signal Processing*. Academic Press, 1998.

[2] VASSALI, M. R., SEIXAS, J. M., ESPAIN, C., *Real-Time Speech Recognition System for Portuguese Language Based on DSP Technology*, IEEE South-American Workshop on Circuits and Systems, 2000.

[3] ANALOG DEVICES, *ADSP-21065L SHARC User´s Manual*, September 1998.

[4] ANALOG DEVICES, *ADSP-21065L EZ-KIT Lite Evaluation System Manual*, December 2000.

[5] HALLIDAY, D., RESNICK, R., WALKER, J., *Fundamentals Of Physics*. 5 ed. John Wiley, 1996.

[6] OPPENHEIM, A. V., SCHAFER, R. W., *Discrete-Time Signal Processing*. Prentice Hall, 1989.

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na VII Convenção Nacional
26-28 de maio de 2003, São Paulo, 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.

Conversão de arquivos WAVE em MIDI

F. Paiva, G.C.R. Abrahão, R.J.R. Cirigliano, R.S. Maia, F.G.V. Resende Jr.
Universidade Federal do Rio de Janeiro, Dept. de Engenharia Eletrônica e Computação
Rio de Janeiro, Po Box 68504 , RJ, Brasil
Email:{paiva, gabrahao, rjcirig, maia, gil}@lps.ufrj.br

RESUMO

Neste artigo está sendo apresentado um sistema de conversão WAVE-MIDI que consiste de três etapas: a extração de pitch, o tratamento desta informação e a geração do arquivo MIDI. Testes realizados mostram um desempenho similar ao de outros softwares comerciais na extração de pitch e um melhor desempenho do algoritmo proposto na geração do arquivo MIDI.

1. INTRODUÇÃO

Atualmente, para se obter a partitura de uma música recorre-se ao trabalho de um profissional com a audição treinada e a exaustivas repetições da mesma melodia. Algoritmos para automatizar este processo vêm sendo desenvolvidos desde o final dos anos 70. Neste artigo está sendo proposto um algoritmo de conversão WAVE-MIDI[1], uma vez que os softwares disponíveis no mercado apresentam muitos erros nessa conversão.

O trabalho de conversão se divide em três etapas: a extração do pitch[2], o seu tratamento e a geração do arquivo MIDI. Para a extração do pitch foi usado o método da autocorrelação, uma vez que ele se mostrou consistente mesmo em situações desfavoráveis e por ter uma baixa complexidade computacional. O tratamento da informação extraída é feito para que sejam eliminados erros nos trechos de silêncio e nas transições entre notas.

Foram realizados testes comparativos com softwares comerciais[3-5] e os resultados mostram um melhor desempenho do algoritmo proposto.

O artigo está organizado da seguinte maneira. Na Seção 2 é apresentado o algoritmo de extração de pitch. Na Seção 3 são feitas considerações relativas ao tratamento do pitch e geração das notas. Os testes realizados são apresentados da Seção 4 e finalmente são mostradas as conclusões na Seção 5.

2. ANÁLISE DE PITCH

A análise de pitch é feita com base no método da autocorrelação[2]. Primeiramente é calculada a autocorrelação do sinal de entrada. Em seguida esta autocorrelação é dividida em três faixas, sendo seus tamanhos respectivamente 14%, 21% e 65% do tamanho original do bloco. Esta divisão é feita para uma melhor estimativa do pitch, uma vez que evita que valores múltiplos do pitch fundamental sejam escolhidos. O valor máximo da autocorrelação dentro de cada uma dessas faixas é calculado e normalizado pela energia do bloco. É então feita uma análise seguindo a ordem dos blocos. Se o pitch da faixa $n+1$ é maior que o pitch normalizado da faixa n ajustado por uma constante ?=0.92, então este será o pitch do bloco. Caso isto não ocorra, o pitch do bloco será o pitch da primeira faixa.

3. TRATAMENTO DO PITCH

O processo de conversão de pitch para MIDI apresenta duas dificuldades: a primeira são as falhas no processo de extração de pitch; a segunda é o fato de que muitas vezes o próprio arquivo de áudio não reproduz fielmente o que tem

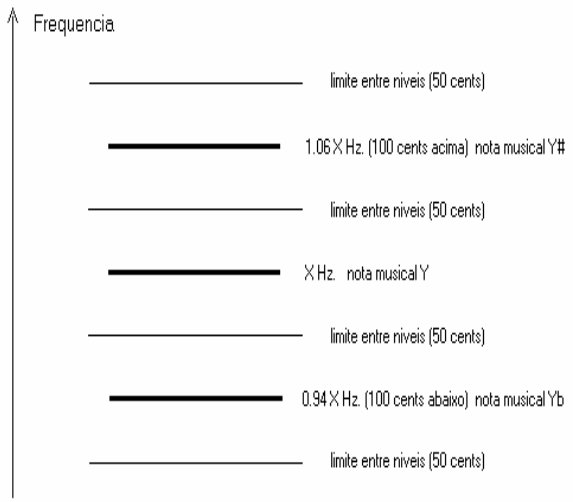


Figura 1 - Níveis para definição das notas.

que ser convertido em partitura. Foram utilizados quatro algoritmos para tentar amenizar esses problemas. O primeiro elimina os trechos de silêncio, enquanto os outros três tratam de suavizar o pitch. Cada um deles será abordado nas seções seguintes.

3.1. Remoção de silêncio

Para remover os trechos de silêncio, utilizamos a informação de intensidade. O algoritmo utilizado calcula a energia de cada frame normalizando os valores das amostras, somando-os e dividindo o valor final pelo número de amostras no frame. Calculamos então a média da intensidade em todos os frames e estabelecemos que, para nossos testes, valores abaixo de 0.05 do valor da média fossem considerados silêncio.

Uma vez removidos os trechos de silêncio, é iniciada a conversão do pitch para os dados MIDI. A conversão é feita pela equação

$$f = f_o \, 2^{c/1200} \qquad (1)$$

onde $f_0 = 440\ Hz$ (A4) , c é um valor dado em cents e f é a freqüência do frame. Cada 100 cents representam uma distância de um semi-tom entre a nota que representa f e A4.

Com base nessa equação, estipulamos níveis que relacionam os valores de pitch e suas respectivas notas, como pode ser observado pela Figura 1.

3.2. Suavização do pitch

A simples divisão do pitch em níveis e sua posteiror conversão não é satisfatória sendo necessário o desenvolvimento de algoritmos que alterem esses níveis de quantização conforme os valores que se encontram ao redor de um determinado dado. Foram criados três tipos de análise: ruído, flutuações e transições. Todos esses algoritmos tratam de variações em pequenos instantes, assim eles se mostram válidos e atuantes de forma útil na maioria dos casos.

3.2.1. Eliminação do ruído

Observando-se a curva de pitch, nota-se que alguns pontos destoam dos demais ao seu redor. Esses pontos são o que chamamos nesse trabalho de ruído e costumam causar erros na geração das notas. Para reduzi-los, definimos que qualquer ponto que tenha seu pitch maior que o dobro, ou menor que a metade, do pitch dos frames anterior e posterior seja descartado na conversão para MIDI. O limiar de dobro (ou metade) é usado porque se trata de uma variação de exatamente uma oitava.

Instrumentos musicais podem passar desse limite porém nosso compromisso final é o de converter a voz humana, que em nenhum caso estudado chegou a alcançar essa variação entre frames vizinhos.

3.2.2. Eliminação da flutuação

Vários instrumentos, assim como a voz, podem flutuar entre certos valores de freqüência quando reproduzem uma nota qualquer e isso não é imperceptível para o extrator de pitch. A princípio, a separação de níveis feita no início do processo deveria contornar esse problema. No entanto, testes mostraram que em alguns casos a separação não era suficiente, e o sistema acabava por errar a nota. Para tentar solucionar o problema foi criado um algoritmo que analisa uma seqüência de três frames. Caso o primeiro e o terceiro frame estejam com o pitch no mesmo nível, e o frame intermediário esteja um pouco acima do limite imposto pela quantização dos níveis, desconsideramos essa pequena flutuação para a criação dos dados MIDI. A Figura 2 ilustra este processo.

Para todos os casos testados o algoritmo removeu as flutuações sem eliminar as nota corretas.

3.2.3. Tratamento de transições

Os instrumentos musicais e a voz humana, quando estão transitando de uma nota para outra, assumem em pequenos instantes valores de pitch intermediários, que são apontados pelo extrator de pitch e podem atrapalhar a conversão. Além disto, a janela utilizada na análise de pitch, apesar de pequena, pode coincidir com um trecho de transição de notas e gerar um valor de pitch totalmente errado. Esses dois fatores tornam necessária a utilização de um algoritmo que remova esses pontos. O algoritmo utilizado toma o pitch de cinco frames consecutivos $[p_1 ... p_5]$. Então definimos que se $p_1 = p_2$, $p_4 = p_5$ e $p_2 \neq p_3 \neq p_4$, então o valor de p_3 é descartado. Esse processo se mostrou útil em vários testes sem que tenha removido nenhuma nota correta.

3.3. Geração do arquivo MIDI

Para a geração do arquivo MIDI, é importante a ordem com que estes algoritmos são aplicados. Para que se obtenha o resultado esperado é preciso que eles sigam a ordem mostrada na Figura 3.

4. EXPERIMENTOS

O primeiro teste realizado compara o desempenho do algoritmo proposto com o de alguns softwares comerciais. Este teste foi realizado utilizando-se um arquivo de áudio gerado a partir de um arquivo MIDI. Assim, podemos analisar com precisão os pontos positivos e negativos de cada sistema. O segundo teste foi feito utilizando-se um arquivo de áudio gravado por uma pessoa sem experiência em canto usando um microfone comum na gravação visando a comparação entre a análise de pitch do nosso sistema e a de um sistema comercial.

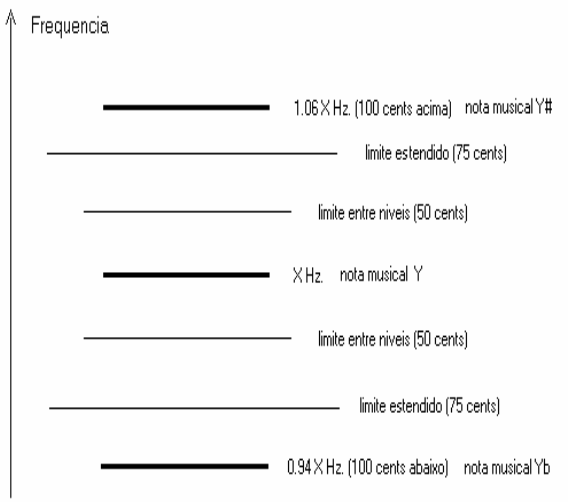


Figura 2 - Nova definição de níveis para eliminação de flutuações.

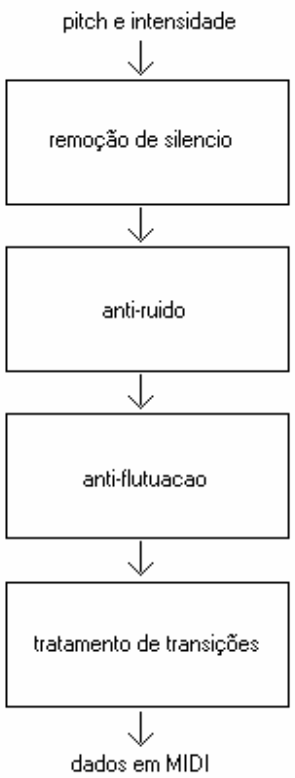


Figura 3 - Diagrama de blocos do sistema de conversão.

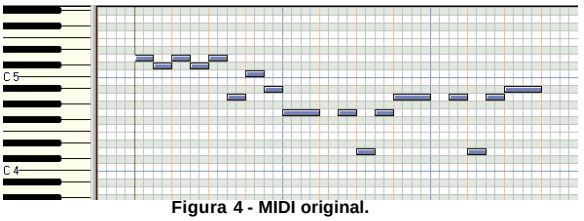


Figura 4 - MIDI original.

4.1. Primeiro teste

Nesse primeiro teste temos o arquivo MIDI da Figura 4 no qual escrevemos um trecho da música Fur Elise (de Beethoven).

Neste teste várias análises podem ser feitas: precisão na geração de notas, análise de silêncio entre duas notas iguais e notas com diferentes durações.

Para realizar o teste, convertemos o arquivo original de MIDI para WAVE utilizando um piano do banco de MIDI disponível. Depois aplicamos o resultado nos diferentes conversores WAVE-MIDI. Os resultados gerados pelos softwares testados são apresentados as Figuras 5, 6, 7 e 8.

Podemos observar pela Figura 5 [3] que as notas geradas não estão todas corretas, pois o software não conseguiu distinguir notas muito próximas como vemos no início do arquivo. Outro erro está no fato de que não há nenhum instante de silêncio no arquivo gerado, enquanto no original temos vários. Já na análise da duração das notas podemos também observar um erro na duração da última nota do arquivo. O software da Figura 6 [4] se mostrou totalmente ineficiente como pode ser notado. A Figura 7 [5] tem notas se sobrepondo além de algumas totalmente fora do local original. Porém, o algoritmo proposto se mostra mais eficiente que os demais, apesar de alguns pequenos erros, apresentando os pontos de silêncio corretamente, além de precisão total nas notas geradas.

4.2. Segundo teste

Para esse teste utilizamos um arquivo de voz gravado. Nele temos uma frase de "Atirei o pau no gato" (música popular). A Figura 9 compara a análise de pitch do sistema proposto neste artigo com a análise de pitch do software Solo Explorer (que se mostrou o melhor dentre os comerciais no primeiro teste). Podemos observar que a análise de pitch está muito semelhante, o que nos permite inferir que o algoritmo proposto apresenta ganho na conversão do pitch para MIDI.

5. CONCLUSÕES

Neste trabalho foi proposto um sistema de conversão WAVE-MIDI divido em três partes: análise de pitch usando o método da autocorrelação, tratamento do pitch e geração do arquivo MIDI. Testes realizados mostram que o sistema proposto consegue converter arquivos de áudio em arquivos MIDI com mais precisão que softwares comerciais.

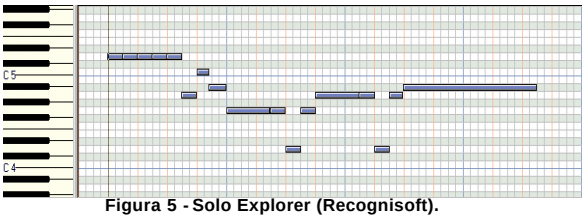


Figura 5 - Solo Explorer (Recognisoft).

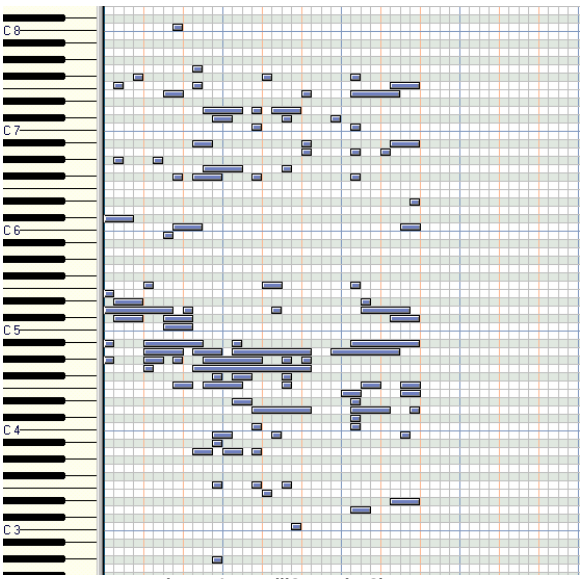


Figura 6 - IntelliScore (IMS)

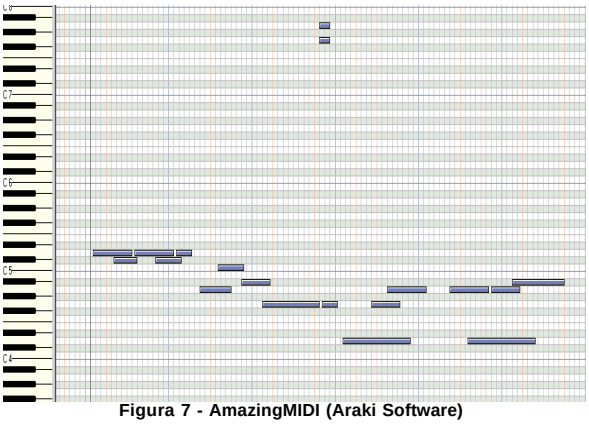


Figura 7 - AmazingMIDI (Araki Software)

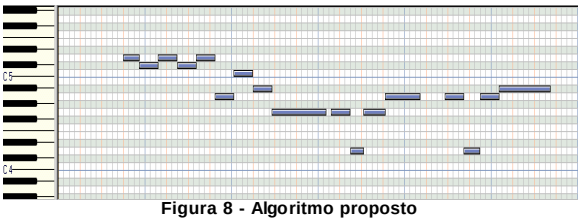


Figura 8 - Algoritmo proposto

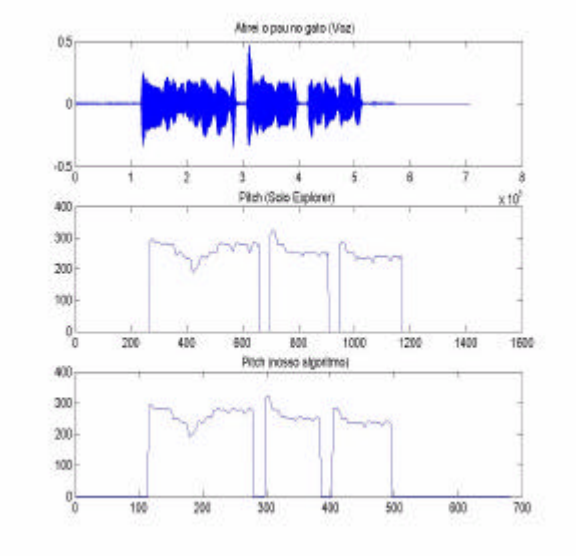


Figura 9 - Comparação entre a análise de pitch do sistema proposto e a do Solo Explorer.

6. REFERÊNCIAS BIBLIOGRÁFICAS

[1] Bloom M., *Music through MIDI*, Editora Microsoft Press, 1987.

[2] Deller J., Proakis J., Hansen J., *Discrete-Time Processing of Speech Signal*, Editora Macmillan, 1993.

[3] Solo Explorer, Recognisoft, www.recognisoft.com

[4] AmazingMIDI, Araki, *www. pluto.dti.ne.jp/~araki/amazingmidi/index.html*

[5] Intelliscore, Intelliscore, www.intelliscore.com

Sociedade de Engenharia de Áudio

Artigo de Convenção

Apresentado na VII Convenção Nacional
26-28 de maio de 2003, São Paulo, 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.

Modulação Sigma Delta em Áudio

Christian Gonçalves Herrera
Departamento de Engenharia Eletrônica, Universidade Federal de Minas Gerais
Av. Antônio Carlos, 6627, Campus - Pampulha
CEP 31270-901 Belo Horizonte, MG - Brasil
herrera@cpdee.ufmg.br

RESUMO

A modulação sigma delta resulta numa codificação do sinal que tem se mostrado extremamente útil em aplicações de engenharia de áudio. Estas aplicações vão desde a conversão A/D e D/A até a amplificação em potência de sinais de áudio. Neste trabalho são abordados as principais características e aplicações relacionadas a este tipo de modulação.

INTRODUÇÃO

Uma tendência relativamente recente, porém já consolidada, é a de lidar com sinais no domínio digital, embora no mundo real estes se apresentem, na maioria das vezes, na forma analógica. As vantagens compreendem, entre outras, as facilidades em se transmitir, armazenar e processar diversos tipos de sinais. Entre eles, podemos destacar os sinais de áudio e vídeo e os sinais biomédicos, ambos por demandarem alta precisão na representação e, em muitas aplicações, sofisticado processamento realizado por computadores ou processadores digitais de sinais dedicados (DSPs).

A conversão de um sinal do domínio analógico para o digital, e vice-versa, é realizada pelos conversores A/D e D/A, que são circuitos eletrônicos que se apresentam sob diversas topologias, cada qual com suas vantagens e limitações. As características que diferenciam estes conversores entre si são a resolução e a faixa de freqüências com que são capazes de trabalhar.

Este artigo trata de uma solução específica para a conversão A/D e D/A: os conversores sigma delta. Sua principal característica é a utilização de circuitos cuja precisão pode ser muito menor do que a precisão do conversor como um todo. Como será apresentado no decorrer do texto, é possível obter uma resolução equivalente a 16 bits na conversão A/D e D/A utilizando um quantizador de apenas 1 bit (comparador), uma solução que encontra larga aplicação hoje em dia em sistemas de gravação e reprodução de áudio (e.g.: CD players).

Outra aplicação é a amplificação em potência de sinais de áudio. No caso da modulação sigma delta, a seqüência de bits resultante pode ser diretamente aplicada a um amplificador classe D e entregue com potência suficiente a um alto falante.

Serão apresentados e discutidos neste artigo os principais aspectos da conversão A/D, como a amostragem e a quantização. A modulação sigma delta será caracterizada e comparada com a modulação por código de pulsos (PCM), que é a mais tradicional e utiliza quantizadores de alta resolução.

Uma análise sobre as não linearidades do processo de quantização será desenvolvida, o que levará a uma razoável compreensão sobre o desempenho dos conversores A/D no que diz respeito à resolução. Serão listadas as principais áreas nas quais a modulação sigma delta encontra aplicabilidade,

assim como alguns circuitos integrados que realizam esta modulação.

CONVERSÃO ANALÓGICO DIGITAL

Amostragem e quantização

Na conversão analógico digital, ou digitalização, de um sinal elétrico, duas operações básicas são efetuadas sobre o sinal: a amostragem no tempo e a quantização na amplitude [1].

O processo de amostragem para um sinal analógico *x(t)* é descrito pela Eq.1 e representado graficamente na Fig.1.

$$x[n] = x(nT_s), \tag{1}$$

onde T_s é o período de amostragem e *x[n]* é o sinal amostrado no instante *n*.

No domínio da freqüência, o efeito da amostragem é de criar versões periodicamente repetidas do espectro do sinal amostrado, centradas em freqüências múltiplas da freqüência de amostragem $f_s = 1/T_s$. Este comportamento é descrito pela Eq.2 e mostrado na Fig.2.

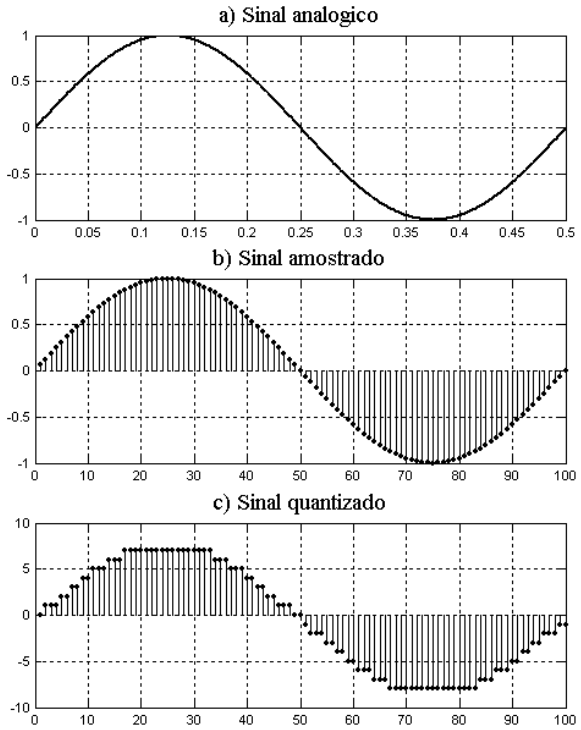


Fig.1: a) Sinal senoidal analógico (2 Hz). b) amostragem (200 pontos por segundo) e c) quantização (16 níveis ou 4 bits).

$$X_s(f) = \frac{1}{T_s}\sum_{k=-\infty}^{\infty} X(f - kf_s), \tag{2}$$

onde $X_s(f)$ representa o espectro do sinal amostrado. Como é possível observar na Fig.2, se a largura da faixa espectral f_B do sinal amostrado for maior que a metade da freqüência de amostragem f_s, ocorrerá uma "invasão" do conteúdo espectral entre duas repetições adjacentes. Este fenômeno é chamado *aliasing*, e é prevenido aplicando o sinal a ser amostrado num filtro passa baixas cuja freqüência de corte é a metade da freqüência de amostragem. Na Fig.2 a resposta deste filtro *antialiasing* aparece em linhas tracejadas. É importante mencionar, embora seja óbvio, que trata-se de um filtro exclusivamente analógico, uma vez que precede o processo de digitalização do sinal. Assim sendo, sua implementação sofre das limitações práticas intrínsecas aos filtros analógicos, como por exemplo, a inviabilidade de se conseguir um filtro de alta ordem (6ª ou maior), o que é necessário quando deseja-se uma faixa de corte relativamente estreita.

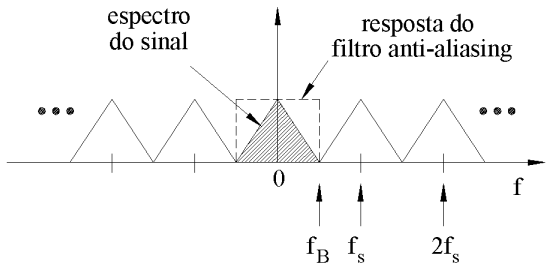


Fig.2: Representação no domínio da freqüência do sinal de largura de banda f_B amostrado com freqüência de amostragem f_s.

O processo de amostragem é uma operação reversível, uma vez que nenhuma informação é perdida e o sinal pode ser perfeitamente reconstruído.

A quantização na amplitude de um sinal é um processo não reversível, uma vez que mapea um número infinito de valores de entrada num número finito de valores de saída. Estas saídas são representadas normalmente como uma palavra digital com um número finito de bits. Um quantizador com *Q* níveis de saída é usualmente caracterizado como tendo *N* bits de resolução, onde $N=\log_2(Q)$. Assim, *Q* deve ser escolhido como sendo uma potência de 2, para que *N* seja um número inteiro. Se *V* é a amplitude máxima do sinal de saída, somente valores de entrada separados de no mínimo $\Delta=2V/(Q-1)$ podem ser distinguidos entre dois níveis diferentes (Fig.3).

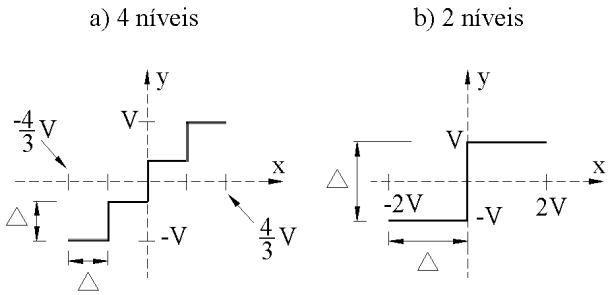


Fig.3: Função de transferência: a) quantizador de 4 níveis e b) quantizador de 2 níveis (comparador).

Na Fig.1 é mostrado um exemplo de um sinal senoidal quantizado com 16 (2^4) níveis. Uma palavra digital de 4 bits é suficiente para representar todos os possíveis níveis de

amplitude deste sinal. Na Fig.3 são mostradas as funções de transferência de dois tipos de quantizadores uniformes.

O processo de quantização é inerentemente não linear, o que dificulta a sua análise. Além disso, o erro introduzido no processo depende da diferença entre a amplitude da entrada e o nível referido na saída. Este erro é responsável pela introdução de um ruído no sinal, e o desempenho de um conversor A/D é mensurado justamente pela quantidade de ruído introduzido na quantização. É claro que quanto maior for o número de níveis de quantização, melhor será o desempenho do conversor.

Entretanto, a implementação eletrônica de um quantizador de alta resolução é extremamente complexa, principalmente quando os circuitos são integrados em chips VLSI (*Very Large Scale Integration*). Uma das limitações refere-se à precisão necessária no casamento dos capacitores usados para realizar repetidas divisões por 2 de uma tensão de referência. Numa conversão A/D de N bits, a precisão requerida é de, no mínimo, uma parte em 2^N. Por exemplo, um quantizador de 10 bits demanda uma precisão da ordem de 0,1% no casamento dos capacitores, o que é extremamente difícil em VLSI.

Análise de desempenho

O interessante na análise do desempenho de um conversor A/D é o comportamento do erro de quantização $e[n]$, que é a diferença entre o sinal de saída $y[n]$ e o sinal de entrada $x[n]$:

$$e[n] = y[n] - x[n], \tag{3}$$

Como $e[n]$ é resultante de um processo genuinamente não linear, para facilitar a análise algumas aproximações são adotadas a respeito do seu comportamento estatístico e da sua dependência com relação ao sinal de entrada [2].

1. $e[n]$ é não correlacionado com o sinal de entrada $x[n]$.
2. $e[n]$ é uma seqüência amostrada de um processo estocástico estacionário.
3. $e[n]$ é uma seqüência independente e identicamente distribuída (i.i.d.), i.e., $e[n]$ é um ruído branco com média zero.
4. $e[n]$ é uniformemente distribuído no intervalo $[-\Delta/2, \Delta/2]$.
5. $x[n]$ é também um processo estocástico i.i.d., uniformemente distribuído no intervalo [-V, V] e de média zero.

Estas aproximações simplificam muito a análise do sistema, uma vez que substituem uma não linearidade determinística por um sistema linear estocástico, permitindo o uso de técnicas de análise de sistemas lineares.

Embora sejam largamente adotadas na literatura científica sobre conversão A/D, estas suposições sofrem sérias limitações. A principal delas refere-se ao fato do erro $e[n]$ ser de fato uma função determinística da entrada, não podendo, pois, ser estatisticamente independente desta (propriedade 1). Em [2] são discutidas com detalhes as limitações das suposições consideradas acima, bem como métodos matemáticos alternativos de análise para o processo aqui abordado.

No caso de valerem todas as aproximações, a variância ou potência de $e[n]$ é definida pela Eq.4, para o caso do quantizador de $Q=2^N$ níveis:

$$\sigma_e^2 = \frac{\Delta^2}{12} = \frac{(2V/Q-1)^2}{12} = \frac{\left(2V/2^N - 1\right)^2}{12} \cong \frac{\left(2V/2^N\right)^2}{12} \tag{4}$$

Um importante parâmetro de análise do desempenho é a relação sinal ruído SQNR (*Signal to Quantization Noise Ratio*). Sendo σ_x^2 a variância do sinal de entrada, a SQNR é expressa pela Eq.5:

$$SQNR = 10\log\left(\frac{\sigma_x^2}{\sigma_e^2}\right) = 10\log\left(\frac{\sigma_x^2}{V^2}\right) + 4{,}77 + 6{,}02N \ \text{(dB)} \tag{5}$$

Assim, para cada bit a mais na resolução do conversor (i.e., para cada incremento em N, há uma melhora de aproximadamente 6 dB na SQNR. Ou seja, há uma relação direta entre a resolução do conversor e a sua SQNR.

Outro parâmetro a ser definido é a faixa dinâmica R, que é uma medida do intervalo de valores de entrada para os quais o conversor produz uma SQNR positiva. Para entradas senoidais, a faixa dinâmica é definida como a razão entre a potência máxima do sinal senoidal ($V^2/2$)e a potência mínima do sinal senoidal que resulta numa SQNR igual a 0 dB ($\sigma_x^2 = \sigma_e^2 = \Delta^2/12$):

$$R = \frac{V^2/2}{\Delta^2/12} \cong \frac{V^2/2}{\left(2V/2^N\right)^2/12}$$

$$R = 6{,}02N + 1{,}76 \ \text{(dB)} \tag{6}$$

A mesma análise de desempenho pode ser feita no domínio da freqüência. Aplicando a transformada Z na Eq. 3, obtemos:

$$Y[z] = X[z]H_x[z] + E[z]H_e[z], \tag{7}$$

onde foram introduzidas as funções de transferência $H_x[z]$ e $H_e[z]$ relativas à modulação do sinal de entrada e do erro de quantização, respectivamente. A densidade de potência espectral na saída do conversor, $P_{ey}(f)$, fornece a informação necessária para se determinar a potência do ruído gerado por $e[n]$:

$$P_{ey}(f) = P_e(f)\left|H_e(f)\right|^2, \tag{8}$$

onde $P_e(f)$ é a densidade de potência espectral do erro de quantização. No caso de $|H_e(f)| = 1$, e de $e[n]$ obedecer às condições citadas, pode-se afirmar que:

$$P_{ey}(f) = P_e(f) = \sigma_e^2 / f_s, \tag{9}$$

onde f_s é a freqüência de amostragem. A integral da densidade de potência espectral $P_{ey}(f)$ somente na faixa de freqüências de interesse f_B, tem como resultado a covariância cruzada ou potência média σ_{ey}^2 gerada pelo ruído de quantização na saída:

$$\sigma_{ey}^2 = \int_{-f_B}^{f_B} P_{ey}(f)df = 2\int_0^{f_B} P_{ey}(f)df$$

$$\sigma_{ey}^2 = \int_{-f_B}^{f_B} P_{ey}(f)df = 2\int_0^{f_B} P_{ey}(f)df$$

$$= \int_0^{f_B} \frac{2\sigma_e^2}{f_s}df = \sigma_e^2\left(\frac{2f_B}{f_s}\right) \qquad (10)$$

Este é um resultado muito interessante, uma vez que relaciona a potência do ruído na saída do conversor com a razão entre a largura da faixa espectral do sinal amostrado f_B e a freqüência de amostragem f_s. Se o conversor utiliza a taxa de Nyquist, $f_s = 2 f_B$ o resultado é o mesmo da Eq. 4. No entanto, se o sinal for amostrado a uma taxa superior à taxa de Nyquist, a Eq. 10 informa que a potência do ruído gerado será menor que no primeiro caso. Esta técnica é chamada sobreamostragem, e seu efeito na prática é o de trocar velocidade por resolução. Em outras palavras, aumentando a freqüência de amostragem de um sinal, consegue-se uma mesma relação sinal ruído utilizando-se um quantizador de menor resolução que gera um ruído de maior potência.

Uma explicação razoável para este fenômeno parte da suposição que o erro de quantização é um ruído branco distribuído uniformemente entre $-f_s/2$ e $f_s/2$. Se $f_s = 2f_B$, então toda a potência do ruído encontra-se distribuído na mesma banda espectral do sinal amostrado. Entretanto, no caso da sobreamostragem, a mesma potência será distribuída numa extensão maior que f_B, restando apenas uma fração desta potência no interior da banda do sinal. Um filtro passa baixas digital é capaz de facilmente remover o ruído que está fora da banda de interesse. O resultado é o aumento na SQNR, e conseqüentemente no desempenho do conversor A/D.

A expressão da SQNR pode ser rescrita na forma mais generalizada:

$$SQNR = 10\log\left(\frac{\sigma_x^2}{\sigma_{ey}^2}\right)$$

$$= 10\log\left(\sigma_x^2\right) - 10\log\left(\sigma_e^2\right) + 10\log\left(\frac{f_s}{2f_B}\right) \ (dB) \qquad (11)$$

Se a razão fs/2fB = 2r, a expressão assume a forma:

$$SQNR = 10\log\left(\sigma_x^2\right) - 10\log\left(\sigma_e^2\right) + 3{,}01r \ \ (dB) \qquad (12)$$

Assim, cada vez que a freqüência de amostragem tiver seu valor dobrado, ocorrerá uma melhora de aproximadamente 3 dB na relação sinal ruído do conversor.

Um dos benefícios da sobreamostragem é a possibilidade de se utilizar um filtro *anti aliasing* com a faixa de corte menos radical do que no caso da amostragem na taxa de Nyquist.

MODULAÇÃO SIGMA DELTA

As funções de transferência $H_x[z]$ e $H_e[z]$ introduzidas na Eq. 7 podem ser escolhidas de tal forma que os sinais $x[n]$ e $e[n]$ sofram modulações que venham a melhorar o desempenho da conversão A/D. É o caso da modulação sigma delta, onde $H_e[z]$ é projetado de forma a concentrar o ruído de quantização fora da banda do sinal a ser amostrado. Contudo, $H_x[z]$ deve ser tal a manter $x[n]$ inalterado. Conclui-se, então, que $H_e[z]$ deve ser a função de transferência de um filtro passa altas, uma vez que o sinal a ser amostrado ocupa a banda espectral entre DC e f_B.

Outra possível implementação acontece no caso do sinal apresentar uma largura de banda igual a $[f_c - f_B/2, f_c + f_B/2]$, onde $f_c >> f_B$ é a freqüência central da banda do sinal. Neste caso, uma função de transferência realizando um filtro passa banda pode ser utilizado.

A implementação básica de um modulador sigma delta é representado pela Eq. 13:

$$Y[z] = X[z]z^{-1} + E[z]\left(1 - z^{-1}\right)^L, \qquad (13)$$

onde L é a ordem do filtro passa altas, e conseqüentemente, a ordem do modulador. É fácil perceber que quanto maior for L, uma maior energia do ruído $e[n]$ será colocado para fora da banda de interesse e melhor será o desempenho do conversor. O sinal de entrada aparece na saída intacto, apenas com um atraso imposto por z^{-1}.

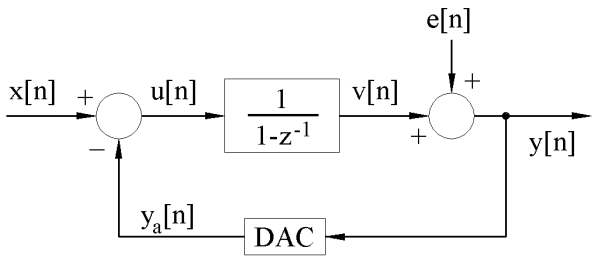


Figura 4: Diagrama de blocos do modulador sigma delta de primeira ordem.

Diversas variações para $H_x[z]$ e $H_e[z]$ já foram propostas. Entre elas, destacam-se as topologias em cascata ou multiestágio [3-6], paralelo [7,8] e multibanda [9]. Quantizadores de diversos níveis também são utilizados. Neste artigo, contudo, serão abordados apenas os moduladores sigma delta de primeira e segunda ordens com quantizadores de 2 níveis (1 bit).

Modulação Sigma Delta de Primeira Ordem

O diagrama de blocos de um modulador sigma delta de primeira ordem é mostrado na Fig. 4. A saída do modulador *Y(z)* é expressa pela equação:

$$Y[z] = X[z]z^{-1} + E[z]\left(1 - z^{-1}\right). \qquad (14)$$

No domínio do tempo, a Eq. 14 pode ser rescrita:

$$y[n] = x[n-1] + e[n] - e[n-1]. \qquad (15)$$

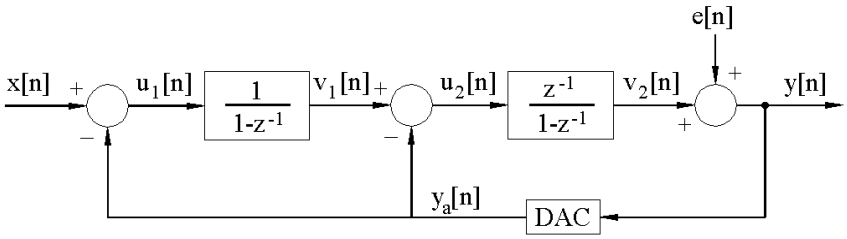


Figura 5: Diagrama de blocos do modulador sigma delta de segunda ordem.

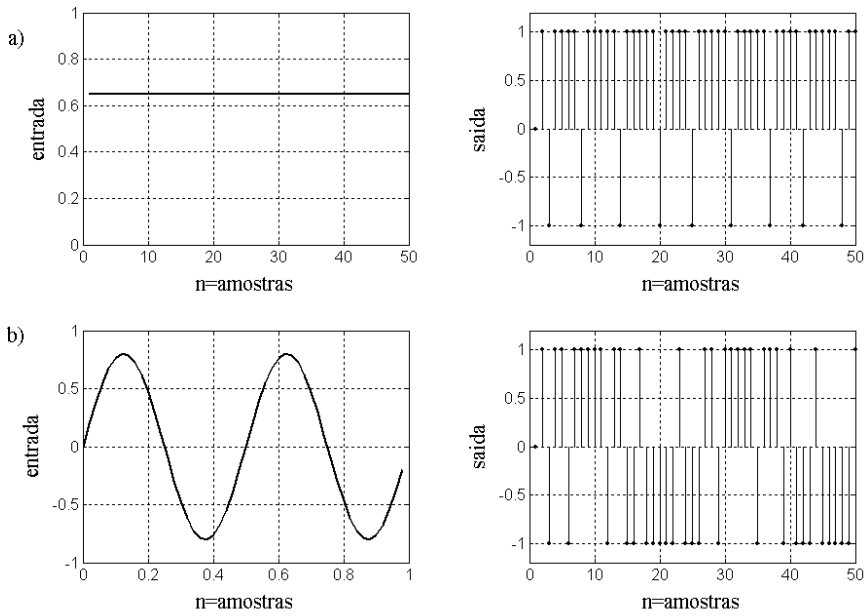


Figura 6: Resposta de um modulador sigma delta de primeira ordem: a) sinal DC e b) sinal senoidal.

Utilizando $H_z = z^{-1}$ e $H_e = (1-z^{-1})$ nas Eqs. 8, 9 e 10, a potência do erro de quantização na saída do conversor é:

$$\sigma_{ey}{}^2 = \sigma_e{}^2 \frac{\pi^2}{3}\left(\frac{2f_B}{f_s}\right)^3 \tag{16}$$

A relação sinal quantizado ruído é, então:

$$SQNR == 10\log\left(\sigma_x^2\right) - 10\log\left(\sigma_e^2\right) - 10\log\left(\frac{\pi^2}{3}\right)$$

$$+30\log\left(\frac{f_s}{2f_B}\right)\ (dB) \tag{17}$$

Se a razão $f_s/2f_B = 2r$:

$$SQNR = 10\log\left(\sigma_x^2\right) - 10\log\left(\sigma_e^2\right)$$

$$-10\log\left(\frac{\pi^2}{3}\right) + 9{,}01r\ \ (dB) \tag{18}$$

Cada vez que a freqüência de amostragem tiver seu valor dobrado, ocorrerá uma melhora de 9 dB na relação sinal ruído do conversor, ou equivalentemente, um aumento de 1.5 bit na resolução.

Modulação Sigma Delta de Segunda Ordem

O diagrama de blocos de uma modulador sigma delta de segunda ordem é mostrado na Fig. 5. A estrutura agora contém dois integradores. As equações para a saída do conversor e para a SQNR são:

$$Y[z] = X[z]z^{-1} + E[z]\left(1 - z^{-1}\right)^2 \tag{19}$$

$$y[n] = x[n-1] + e[n] - 2e[n-1] + e[n-2]. \tag{20}$$

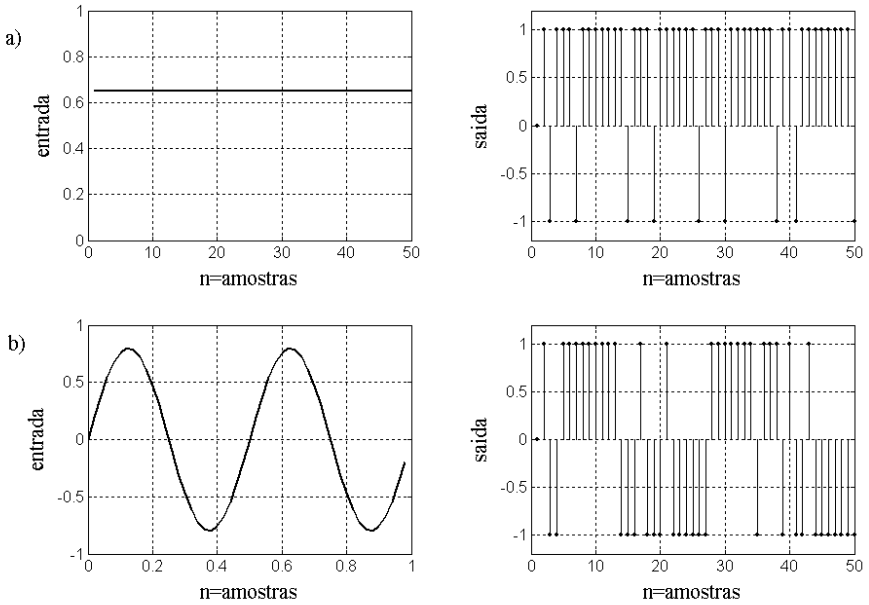


Figura 7: Resposta de um modulador sigma delta de segunda ordem: a) sinal DC e b) sinal senoidal.

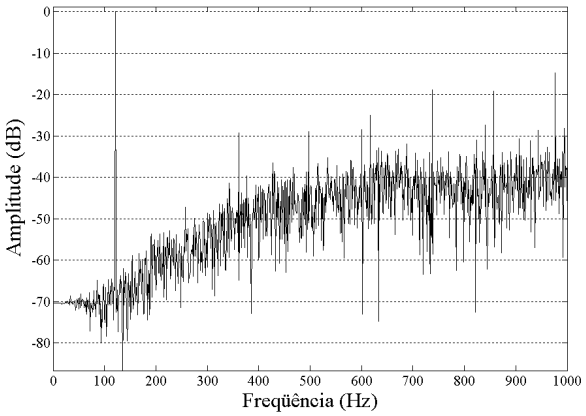


Figura 8: Resposta em freqüência de um modulador sigma delta de segunda ordem com f_s = 8129 Hz e um sinal senoidal (pico mais alto) de 120 Hz e amplitude igual a 0.8 na entrada.

$$\sigma_{ey}{}^2 = \sigma_e{}^2 \frac{\pi^4}{5}\left(\frac{2f_B}{f_s}\right)^5 \tag{21}$$

$$SQNR = 10\log\left(\sigma_x^2\right) - 10\log\left(\sigma_e^2\right) - 10\log\left(\frac{\pi^4}{5}\right) + 15{,}05r \ \text{(dB)} \tag{22}$$

Neste caso, um incremento em r correspondente a uma multiplicação por 2 na freqüência de amostragem, proporciona um aumento de 2.5 bits na resolução do conversor, ou 15 dB de melhora na SQNR.

O modulador de segunda ordem incorpora um integrador a mais em sua função de transferência. Teoricamente, este integrador contribui para atenuar o ruído de quantização na banda de interesse e reforça-lo em freqüências altas.

Comportamento Qualitativo

A modulação sigma delta pode ser vista como um conversor PCM com realimentação, que tenta forçar a saída $y[n]$ a ser igual à entrada $x[n]$. Considerando o caso onde um quantizador de 1 bit é utilizado, sendo V =1, a saída do comparador assume os valores +1 ou –1. Sendo assim, sempre haverá um erro $u[n] \neq 0$, exceto quando a entrada assumir exatamente um dos valores citados acima.

Considerando uma entrada DC entre 0 e 1 no modulador de primeira ordem da Fig. 4, quando $y[n]$ = 1 o erro $u[n]$ acumulado pelo integrador é negativo, já que $u[n] = x[n] - y[n]$. Após um determinado número de ciclos estes valores negativos acumulados serão suficientes para mudar o estado do comparador para $y[n]$ = -1. O sinal do erro muda para um valor positivo e é acumulado novamente no integrador por alguns ciclos até mudar o estado do comparador de volta para +1. A densidade de +1's e –1's num período de tempo é relacionado ao valor DC da entrada, de maneira que quanto maior o valor da entrada, maior a densidade de +1's na saída, e vice-versa. Por essa razão, o modulador sigma delta utilizando um quantizador de 1 bit é comumente denominado modulador por densidade de pulsos (PDM).

No modulador de segunda ordem o comportamento é parecido. Contudo, a entrada do quantizador $v_2[n]$ é uma versão mais refinada, ou precisa, do erro entre a entrada e a saída, uma vez que este erro $u_1[n]$ é aplicado num integrador antes de ser subtraído de $y[n]$e gerar $u_2[n]$. O resultado é uma representação mais precisa do sinal de entrada.

Nas Figs. 6 e 7 são mostradas as respostas dos moduladores de primeira e segunda ordens aos sinais DC e senoidal. É possível perceber as diferentes distribuições de pulsos implementadas por cada modulador.

A resposta em freqüência da saída de um modulador sigma delta de segunda ordem é mostrada na Fig. 8. O pico mais alto refere-se ao sinal senoidal de freqüência igual a 120 Hz e amplitude igual a 0.8 aplicado à entrada. A freqüência de amostragem neste caso é de 8192 Hz, o que corresponde a uma taxa de sobreamostragem de aproximadamente 68. É interessante observar a atenuação do ruído de quantização nas baixas freqüências, provida pelo filtro passa altas $H_e = (1-z^{-1})^2$, enquanto o sinal de entrada permanece inalterado.

Algumas imperfeições nos circuitos eletrônicos responsáveis pela implementação do modulador podem degradar o seu desempenho [2], como é o caso do integrador. Se o ganho não for unitário e houver um fator de fuga considerável, a função de transferência do erro de quantização sofre mudanças em seus parâmetros que se traduzem numa pior atenuação do ruído dentro da faixa de freqüências de interesse.

Imperfeições no conversor D/A e no quantizador também são susceptíveis de ocorrer, no entanto não chegam a comprometer o desempenho do conversor como um todo.

Contudo, um aspecto que merece especial atenção refere-se à natureza não linear do conversor, à realimentação, e ao fato de o ruído de quantização não ser branco como assumido. Tudo isso leva ao aparecimento de componentes periódicas (*limit cycle oscillations*) na saída do conversor [10]. Estas componentes podem se situar dentro da banda do sinal de entrada, principalmente em conversores de primeira ordem, o que é extremamente indesejável em aplicações de áudio e voz.

Comportamento Quantitativo

Nesta seção o objetivo é traçar um paralelo entre as diversas implementações de conversores A/D discutidas neste artigo com um exemplo numérico baseado nas equações propostas ao longo do texto.

O exemplo é um conversor A/D para aplicação em áudio de alta fidelidade, onde a largura de banda é igual a 20 kHz e a relação sinal ruído requerida é de 98 dB, ou 16 bits como é o caso dos conversores PCM utilizados nos CDs *players* de áudio.

A Tabela 1 reúne os resultados das freqüências de amostragem necessárias para atingir o desempenho esperado.

Implementação Prática

Um modulador sigma delta de segunda ordem foi implementado num DSP (*Digital Signal Processor*) da família TMS320C2407 do fabricante Texas Instruments.

O algoritmo escrito na linguagem de programação C++ executa uma leitura nos conversores A/D de 10 bits internos ao DSP e calcula a saída $y[n]$ baseado nas equações desenvolvidas anteriormente neste artigo. A freqüência de amostragem escolhida foi a maior possível, respeitando o *clock* do DSP de 30 MHz. Como foi medido que o algoritmo gasta cerca de 90 ciclos de *clock* para realizar todas as operações do cálculo de $y[n]$, a freqüência de amostragem ficou estabelecida em 300 kHz.

Tabela 1: Freqüência de Amostragem para Conversão A/D em Áudio de Alta Fidelidade
f_B = 20 kHz, SQNR = 98 dB

Tecnologia	Freqüência de Amostragem
PCM 16 bits	44.1 kHz
PCM Sobreamostrado 12 bits	10 MHz
PCM Sobreamostrado 8 bits	2.64 GHz
Sigma Delta Primeira Ordem 1 bit	96.78 MHz
Sigma Delta Segunda Ordem 1 bit	6.12 MHz
Sigma Delta Segunda Ordem 5 bits	1.53 MHz
Sigma Delta Terceira Ordem 1 bit	1.92 MHz

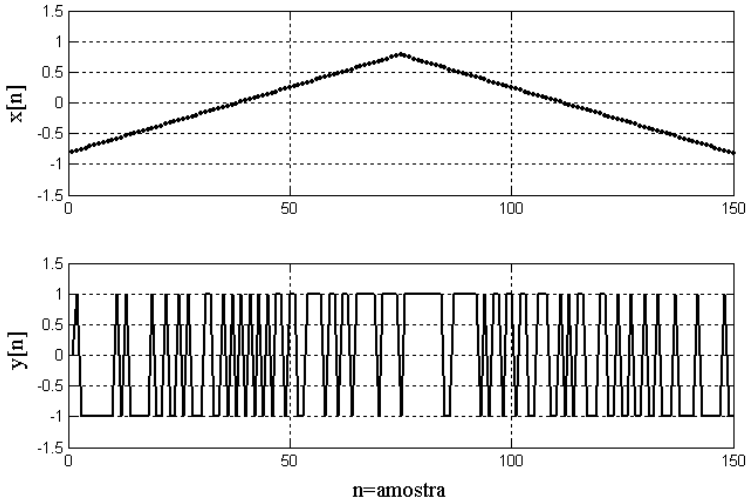


Fig. 9: Resultado da simulação de um conversor sigma delta de segunda ordem com taxa de sobreamostragem igual a 150. O sinal de entrada é uma onda triangular com amplitude variando entre –0.8 e +0.8.

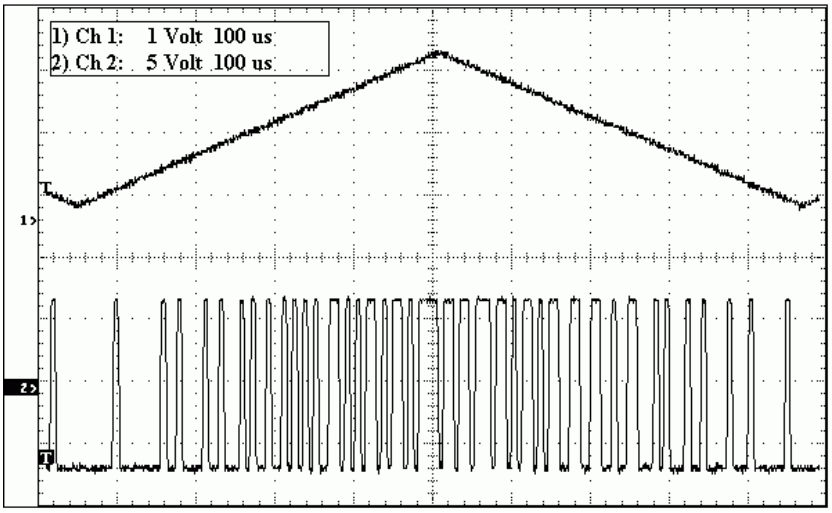


Fig. 10: Formas de onda na entrada (triangular, 1 kHz) e na saída do DSP TMS320C2407 com taxa de sobreamostragem igual a 150.

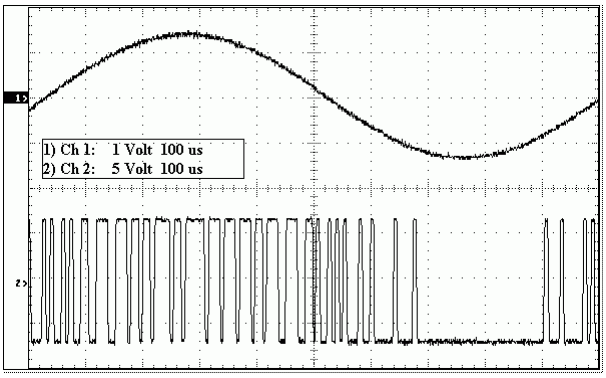


Fig. 11: Saída modulada do DSP para um sinal senoidal de entrada com freqüência igual a 1 kHz.

Na Fig. 9 é mostrado o resultado de uma simulação na qual é aplicada à entrada do modulador um sinal triangular, onde $f_s/2f_B$ = 150. Esta figura é comparada à forma de onda adquirida da saída do DSP, mostrada na Fig. 10.

A Fig. 11 mostra os sinais de entrada e saída do modulador, onde a entrada é um sinal senoidal de 1000 Hz.

Infelizmente não foi viável realizar uma análise no domínio da freqüência da saída do modulador sigma delta implementado devido à impossibilidade em se adquirir com o equipamento disponível um número de pontos que permitisse uma resolução em freqüência razoável.

APLICAÇÕES DE CONVERSORES SIGMA DELTA

Nesta seção são apresentadas algumas das principais situações onde os conversores sigma delta encontram grande aplicabilidade. Os parâmetros resolução e largura de banda são os mais importantes na avaliação dos resultados conseguidos pelas diversas implementações. Na Tabela 2 estão agrupados vários exemplos encontrados na literatura.

A conversão de dados em instrumentação demandam altos níveis de resolução (19-20 bits), embora envolvam freqüências até aproximadamente 500 Hz, na maioria dos casos. Taxas de sobreamostragem (*tsa*) iguais a 128 (f_s = 128 kHz) são comumente encontradas e a SQNR pode chegar a 120 dB.

Em codificação de voz para aplicações de telefonia, a largura de banda requerida é de 4 kHz com uma resolução de 13 ou 14 bits. Aqui, para atingir 80 dB de SQNR, pode-se encontrar conversores de 1 bit e primeira ordem com *tsa* = 500, ou de segunda ordem com *tsa* = 128.

Sistemas de áudio digital de alta fidelidade tem especificações mais rigorosas, com largura de banda igual à 20 kHz e resolução de, no mínimo, 16 bits. Várias arquiteturas têm sido propostas, utilizando quantizadores de 1 ou 4 bits, topologias de primeira, segunda ou quarta ordens, ou mesmo circuitos em cascata.

Em sistemas de telecomunicações como ISDN ou telefonia celular, a modulação sigma delta também encontra aplicabilidade. A largura de banda nestes casos é relativamente grande, e a resolução varia entre 12 e 14 bits na maioria dos casos. A topologia mais utilizada é a cascata com filtragem passa banda.

Na Tabela 3 estão listados alguns dos principais fabricantes de circuitos integrados para aplicações em conversão A/D que mantêm modelos utilizando a modulação sigma delta.

CONCLUSÕES

Este trabalho apresentou a tecnologia de codificação de sinais utilizando a modulação sigma delta. Para isso, foi feita uma revisão dos conceitos envolvidos em conversão analógico digital. Foram definidos os principais parâmetros para a avaliação do desempenho destes conversores, e apontados os principais pontos fracos que podem levar a um comprometimento na qualidade do conversor.

A modulação sigma delta se apresenta como uma excelente solução para a conversão A/D e D/A de sinais de áudio. Sua principal vantagem sobre outras topologias é o fato de sua precisão não estar limitada pela precisão dos circuitos elementares que conformam o conversor. Além disto, a seqüência de bits resultante da conversão pode ser diretamente aplicada num amplificador classe D para amplificação em potência. A demodulação consiste simplesmente numa filtragem das freqüências fora da banda espectral de interesse.

O aumento crescente na velocidade de funcionamento dos circuitos integrados tem tornado cada vez mais viável a utilização da modulação sigma delta em aplicações de áudio.

Tabela 2: Aplicações Típicas de Conversores Sigma Delta*

Largura de Banda do Sinal (f_B)	Freqüência de Amostragem (f_s)	TSA ($f_s/2f_B$)	Resolução (bits)	Estrutura do Modulador	Quantizador	Aplicação	Referência
492 Hz	128 kHz	128	20	4ª ordem	1 bit	instrumentação	[11]
500 Hz	128 kHz	128	20	4ª ordem	1 bit	abalos sísmicos	[12]
4 kHz	4 MHz	500	13	1ª ordem	1 bit	voz	[13]
4 kHz	1.024 MHz	128	13	2ª ordem	1 bit	voz	[14]
20.5 kHz	5.25 MHz	128	16	2ª ordem	4 bits	áudio	[15]
24 kHz	6.144 MHz	128	18	4ª ordem	4 bits	áudio	[16]
24 kHz	3.072 MHz	64	16	2ª ordem	1 bit	áudio	[17]
25 kHz	6.4 MHz	128	17	"2-1" cascata	1 bit	áudio	[6]
40 kHz	10.24 MHz	128	14	2ª ordem	1 bit	ISDN	[18]
40 kHz	2.56 MHz	32	13	"2-1" cascata	1 bit	ISDN	[4]
100 kHz	3.25 MHz	16	15	"2-2-2" cascata	3 níveis	celular digital	[19]
160 kHz	20.48 MHz	24	16	"2-1" cascata	1 bit	-	[5]
250 kHz	32 MHz	32	14	4ª ordem	1 bit	-	[20]
1 MHz	50 MHz	20	12	"2-1" cascata	1 bit e 3 bits	ultra-som	[21]

* Retirado de [23].

Tabela 3: Conversores Sigma Delta em Circuitos Integrados

Fabricante	Modelo	Largura de Banda do Sinal (f_B)	Freqüência de Amostragem (f_s)	TSA ($f_s/2f_B$)	Resolução (bits)	Quantizador	Estrutura do Modulador
Analog Devices	AD7720	98 kHz	25 MHz	128	16	1 bit	7ª ordem
	AD1878	21.7 kHz	3.5 MHz	81	16	1 bit	5ª ordem
Texas Instruments	PCM1760	24 kHz	3.072 MHz	64	20	4 bits	4ª ordem
	TLC320AD57	24 kHz	3.072 MHz	64	18	1 bit	4ª ordem
National Semiconductor	ADC16071	21.6 kHz	6.144 MHz	142	16	1 bit	2ª ordem
Motorola	DSP56ADC16	45.5 kHz	6.144 MHz	71	16	1 bit	3ª ordem

REFERÊNCIAS BIBLIOGRÁFICAS

[1] A. Oppennheim e R. Schafer, *Discrete Time Signal Processing*, (Prentice-Hall, 1989).

[2] S. R. Norsworthy, R. Schreier, G. C. Temes, *Delta Sigma Data Converters*, (IEEE Press, 1996).

[3] K. Uchimura, T. Hayashi, T.Kimura e A. Iwata, "Oversampling A-to-D e D-to-A converters with multistage noise shaping modulators," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, pp. 1899-1905, December, 1988.

[4] L. Longo e M. Copeland, "A 13 bit ISDN-band oversampled ADC using two-stage third order noise shaping," *Proceedings, IEEE Custom Integrated Circuits Conference*, pp. 21.2.1-21.2.4, 1988.

[5] G. Yin, F. Stubbe, W. Sansen, "A 16-b 320 kHz CMOS A/D converter using two-stage third order noise shaping," *IEEE Journal of Solid State Circuits*, pp.640-647, June, 1993.

[6] L. Williams e B. Wooley, "Third order sigma-delta modulator with extended dinamic range," *IEEE Journal of Solid State Circuits*, pp.193-202, March, 1994.

[7] W. Black e D. Hodges, "Time interleaved converter arrays," *IEEE Journal of Solid State Circuits*, pp.1022-1029, December, 1980.

[8] A. Petraglia e S. Mitra, "High speed A/D converters using QMF banks," *IEEE International Symposium on Circuits and Systems*, pp. 2797-2800, 1990.

[9] P. Aziz, H. Sorensen, J. Van der Spiegel, "Multiband sigma delta modulation," *Electronics Letters*, pp. 760-762, April 29, 1993.

[10] V.Friedman, "The structure of limit cycles in sigma delta modulation," *IEEE Transactions on Communications*, pp. 972-979, August, 1988.

[11] C. Thompson, S. Bernadas, "A digitally corrected 20b delta-sigma modulator," Digest of Technical Papers, *International Solid State Circuits Conference*, pp. 194-195, 1994.

[12] D. Derth, D. Kasha, et al, " A 120 dB linear switched-capacitor delta-sigma modulator," Digest of Technical Papers, *International Solid State Circuits Conference*, pp. 194-195, 1994.

[13] B. Leung, R. Neff, P. Gray, R. Broderson, "Area-efficient multichannel oversampled PCM voice-band coder," *IEEE Journal of Solid State Circuits*, pp.1351-1357, December, 1988.

[14] V. Friedman, D. Brinthaupt, et al, "A dual-channel voice-band PCM codec using sigma delta modulation technique," *IEEE Journal of Solid State Circuits*, pp.274-280, April, 1989.

[15] M. Sarhang-Nejad e G. Temes, "A high-resolution multibit sigma-delta ADC with digital correction and relaxed amplifier requirements," *IEEE Journal of Solid State Circuits*, pp.648-660, June, 1993.

[16] R. Adams, "Design and implementation of an audio 18-bit analog-to-digital converter using oversampling techniques," *Journal of the Audio Engineering Society*, pp. 153-166, March, 1986.

[17] D. Welland, B. Del Signore, et al, "A stereo 16-bit sigma-delta A/D converter for digital audio," *Journal of the Audio Engineering Society*, pp. 476-486, June, 1989.

[18] S. Norsworthy, I. Post, H. Fetterman, "A 14-bit 80kHz sigma-delta A/D converter: modeling, design and

performance evaluation," *IEEE Journal of Solid State Circuits*, pp.256-266, April, 1989.

[19] I. Dedic, "A sixth-order triple-loop sigma-delta CMOS ADC with 90 dB SNR and 100 kHz bandwidth," Digest of Technical Papers, *International Solid State Circuits Conference*, pp. 188-189, 1994.

[20] F. Op't Eynde, G. Yin, W. Sansen, "A CMOS fourth-order 14-bit 500k-sample/s sigma-delta converter," Digest of Technical Papers, *International Solid State Circuits Conference*, pp. 62-63, 1991.

[21] B. Brandt and B. Wooley, "A CMOS oversampling A/D converter with 12-bit resolution at conversion rates above 1 MHz," Digest of Technical Papers, *International Solid State Circuits Conference*, pp.64-65, 1991.

[22] A. Leon-Garcia, "*Probability and Random Processes for Electrical Engineering*," (Addison-Wesley Publishing Company, 1989).

[23] P. M. Aziz, H. V. Sorensen, J. Van der Spiegel, "An Overview of Delta-Sigma Converters," *IEEE Signal Processing Magazine*, pp.61-84, January 1996.

Sociedade de Engenharia de Áudio
Artigo de Convenção
Apresentado na VII Convenção Nacional
26-28 de maio de 2003, São Paulo, 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.

Difusores Acústicos III

Ing. Alejandro Bidondo
AES Member
Ciudad de Buenos Aires, Buenos Aires, Argentina,

abidondo@ingneieriadesonido.com

RESUMO
En esta tercera etapa del estudio de la difusión acústica se trabajó específicamente sobre la medición y análisis del SFD en distintos recintos, la sala AB del Centro Cultural San Martín, el teatro del Colegio Euskal Echea, el salón Dorado de Teatro Colón y la sala principal del Teatro Colón. De estas experiencias se extrajeron los parámetros a los que es sensible el SFD y los datos para un nuevo punto de partida para analizar la acústica de recintos.

Teoría:

Toda investigación nace de ciertos cuestionamientos y en este caso algunos fueron:

- ¿Puede un solo número, el "coeficiente de difusión" obtenido del desparramo espacial de la energía, caracterizar completamente el funcionamiento de una superficie, siendo éste tan complejo?.
- ¿Cuánta superficie difusora es necesaria para conformar un campo difuso?.
- ¿Cuántos tipos de campos difusos se pueden conformar?.
- ¿Es medible la difusión de un campo difuso como para poder repetir valores obtenidos en proyectos ya realizados?.
- ¿Qué superficie difusora es necesaria en un control y en una sala de un estudio de grabación?.
- ¿Es perfectible un campo difuso?.
- ¿Es estudiable un campo difuso?, ¿Cómo?.
- ¿Qué diferencia hay entre conformar un campo difuso con superficies aleatorias respecto de superficies devenidas de teorías numéricas y softwares de optimización?

Definiciones:

Difusión:
Propiedad de las superficies por la que desparraman la energía acústica en el espacio en forma no especular y en el tiempo.
Campo Difuso:
Espacio físico donde existe similar *decorrelación binaural* de la energía acústica.

En las anteriores definiciones se hacen tres afirmaciones ya demostradas en trabajos anteriores (Difusores Acústicos II):

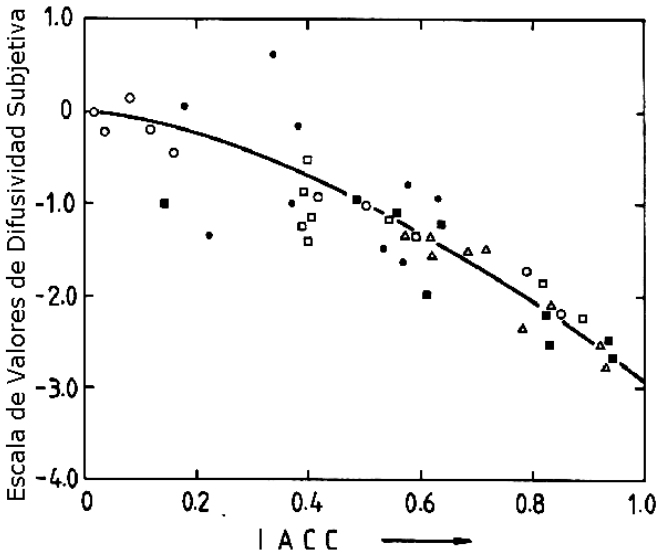
1. Los difusores desparraman la energía acústica *en el tiempo*.
2. El método de medición de los campos difusos es intrínsecamente *binaural*.
3. Los campos difusos poseen un *valor medio* de *difusividad* y un *desvío estándard* que mide su homogeneidad.

SFD:
Es un *índice de Espacialidad Subjetiva* que indica el *grado de decorrelación de reflexiones entre el oído izquierdo y derecho* (binaural).

Post procesado de las mediciones binaurales:
Sabiendo que existe una correlación entre el IACC y la *sensación subjetiva* de difusión, a partir de los valores del primero se llegó a los valores del segundo por medio de una *ecuación* fruto de una regresión matemática entre resultados obtenidos en anteriores estudios.

IACC:

$$IACC = \max \quad |\phi_{lr}(\tau)| \quad para \quad |\tau| \leq 1ms$$



La Ecuación que mejor ajusta las muestras obtenidas es:

$$SFD = -2{,}9 \cdot (IACC)^{\frac{3}{2}}$$

Trabajo de investigación

Descripción de las salas evaluadas:

Teatro del Colegio Euskal Echea:
Es una sala construida en 1940 fue proyectada para conciertos de orquestas Filarmónicas, con capacidad para 1000 personas entre la PB y el Pullman. El techo presenta concavidades que focalizan la energía sonora hacia ciertas franjas de 2m de ancho aproximadamente sobre la audiencia, paralelas a la línea frontal del escenario.

Sala AB del Centro Cultural San Martín:
Sala construida durante la década de 1960, con capacidad para 1000 personas sentadas, todas en Planta baja.

Sala principal del Teatro Colón:
Es la casa de Ópera más importante de Argentina. El Teatro Colón fue construido en 1908, con capacidad para 2487 personas sentadas.

Salón Dorado del teatro Colón:
Dentro del teatro Colón, en el primer piso sobre la entrada de la calle Libertad, se encuentra este salón el cual se utiliza para conciertos vocales y musicales (usualmente no más de 5 o 6 instrumentos) con capacidad para 100 personas aproximadamente.

Equipamiento de Medición:

- Un par (2) de micrófonos Crown PZM en configuración binaural espaciados 17cm entre sí (llamado "*Binaural Wood Head*").
- El post procesamiento fue realizado mediante 2 canales de FFT utilizando una ventana tipo Hanning, analizadores de espectro y de oscilogramas, softwares de correlación y procesadores de integral reversa de Schroeder.
- 1 Micrófono omnidireccional Eartworks M30.
- 2 Micrófonos Shure SM81.
- 1 Micrófono AKG414 ULS.
- 1 Micrófono Audio Technica AT4033.
- Un reproductor de CD.
- Un grabador de ADAT de 8 canales.
- 7 Preamplificadores de micrófono de una consola Soundcraft de 16 canales de entrada.
- Un amplificador de Audio de 2 canales marca CREST FA901.
- Un altavoz marca JBL modelo 4425, a menos que se especifique lo contrario.

Disposición microfónica:

Se planteó una nueva disposición microfónica para la medición de los parámetros acústicos de una sala. La misma cuenta con 7 micrófonos de diferentes características polares, 6 de los cuales se encuentran en la posición de medición.

Sobre la fuente sonora:
Cardioide:

El mismo capta lo emitido por esta, de manera de conocer cómo está siendo excitada la sala. La distancia de este micrófono hasta la fuente será > $\frac{L^2}{\lambda}$, para cumplir con la condición de *Fraunhoffer* de campo lejano para fuentes puntuales. Debido a que el SFD es un parámetro subjetivo de difusividad de los recintos, es recomendable (para lograr una buena mensura) que la fuente posea la direccionalidad que tendría una fuente sonora real de uso común dentro de la sala; en su defecto, la aplicación de las leyes del buen arte sugieren el uso de fuentes omnidireccionales, con lo que el resultado será más universal pero menos real.

En cada posición del oyente a ser evaluada:
Omnidireccional:

El mismo capta la totalidad de la respuesta al impulso $h_T(t)$.

Cardioide frontal:

El mismo capta la información directa, proveniente del frente a partir de la posición de medición.

Cardioide trasero:

El mismo capta solamente lo emitido por aquella porción de sala existente detrás de la posición de medición.

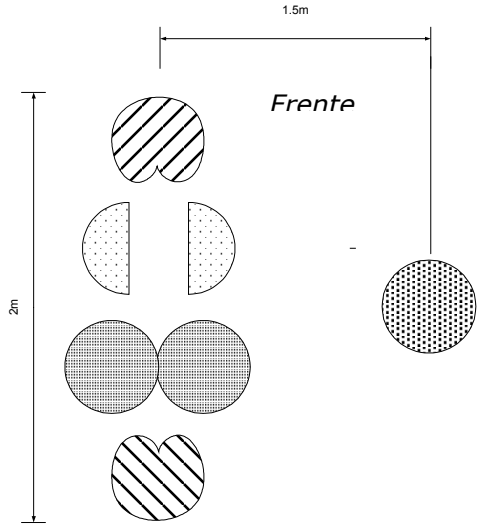
Figura de ocho:

El mismo capta sólo la energía lateral.La misma puede ser comparada con la energía total (micrófono omnidireccional) y/o con la energía frontal (micrófono cardioide).

"Binaural Wood Head" (2 micrófonos hemiesféricos separados 17cm uno de otro):

Este set de 2 micrófonos realiza una captación binaural en la posición de medición. Separadamente se sugiere evaluar sus captaciones tomando como entrada lo captado por el micrófono de fuente, en la medida que así lo requieran los resultados que arroje la evaluación del SFD (*RT60 binaural*).

La disposición de los micrófonos que minimiza la interferencia entre ellos es:



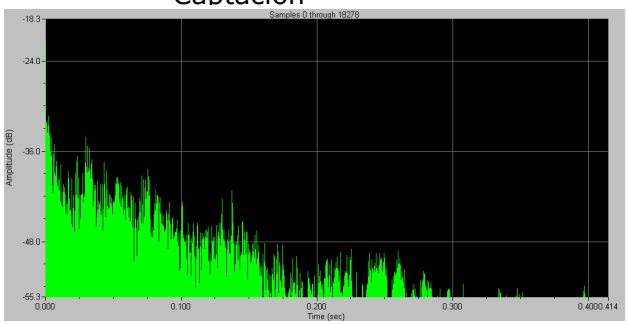
RT60 Binaural:

Se encontró que el análisis tradicional de la respuesta al impulso de un recinto en una posición cualquiera de medición enmascara resultados que se evidencian en las respuestas al impulso binaurales, tomando como excitación del sistema lo captado en cercanía a la fuente sonora y como salida del sistema lo propio de los micrófonos de la "*Binaural Wood Head*". Como de toda respuesta al impulso, también se pueden obtener los valores de RT60 (en este caso valores L y R), EDT, Claridad 50 y 80.

Los valores resultantes del mapeo del SFD motivan el precedente análisis encontrándose así los comportamientos particulares de cada posición de evaluación.

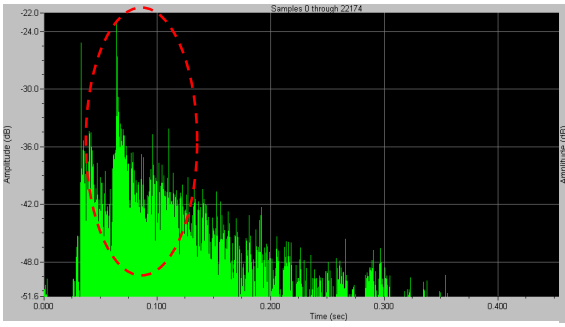
En las siguientes figuras se puede observar el resultado de la captación monofónica de la posición 2 del CCGSM que enmascara los fenómenos observados en las respuestas al impulso *binaurales*, como ser reflexiones, ancho de banda de las mismas, los decays lateralizados y las diferencias de claridad entre L y R.

Captación



RT60 = 1.02

Oído Izquierdo

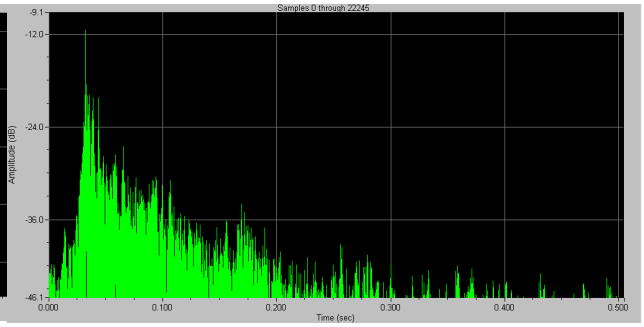


RT20 = 0.64

C50 = 3.38dB

C80 = 6.83dB

Oído Derecho



RT20 = 0.58

C50 = 6.99dB

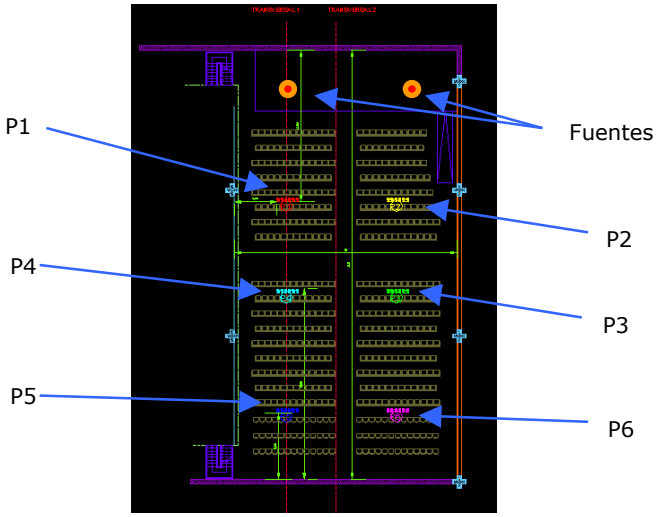
C80 = 11.53dB

Algunos resultados:

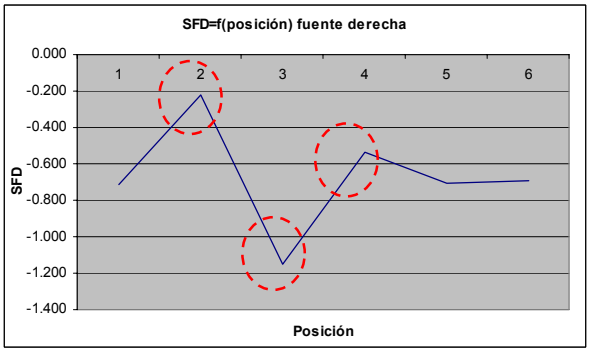
<u>Centro Cultural General San Martín:</u>

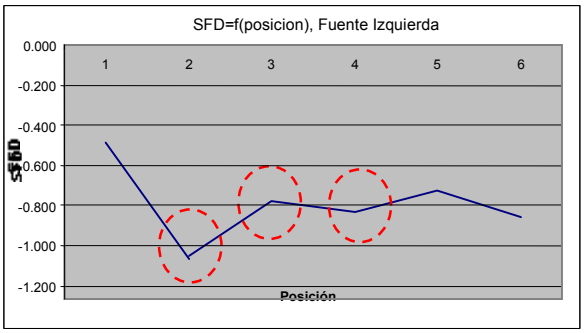
La señal de excitación utilizada fue pseudo-ruido MLS sin ningún tipo de ecualización.
Las fuentes sonoras fueron 2 cajas acústicas marca Meyer, modelo , las que tienen incorporado un sistema de protección contra sobre excitación de señal. Se tuvo la precaución de no trabajar con es dicho sistema accionado.
Se utilizaron estas cajas en sus respectivas posiciones (L y R) debido a que son las fuentes sonoras de uso habitual en dicho recinto. Esto generó 2 (dos) mediciones en cada posición, una con cada caja.

Se detalla a continuación el plano con las posiciones de las mediciones realizadas:



Confeccionando los gráficos del SFD=f(posición, Fuente) es posible analizar la simetría axial del recinto. Los resultados fueron:





Se encontraron grandes diferencias de SFD en las posiciones 2, 3 y 4 entre la excitación de la caja izquierda y la derecha (*asimetría axial de difusión*). Esto motivó el posterior análisis de los *parámetros binaurales* anteriormente mencionados encontrando los motivos de tales discrepancias.

Además a partir de los gráficos anteriores es posible analizar el campo difuso hallando el promedio y el desvío de los valores encontrados.

Teatro Colón, Sala principal:

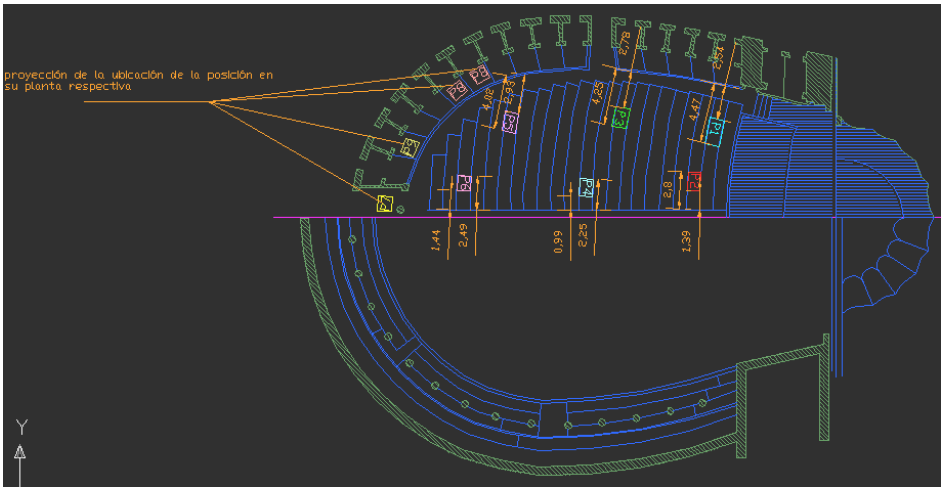
La señal de excitación utilizada fue pseudo-ruido MLS con ecualización de tipo Rosa (pendiente = -3dB/Oct) con el objeto de cuidar la integridad funcional de la fuente sonora, ya que la misma no posee protección alguna.

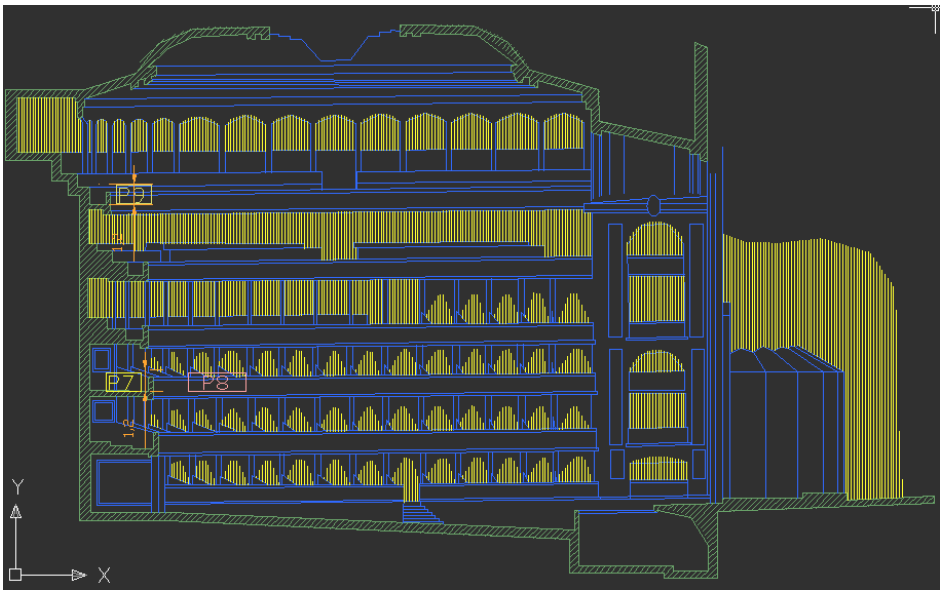
La fuente sonora utilizada fue la caja JBL detallada anteriormente.
La ubicación de la misma fue en el centro del escenario, presuponiendo que en el uso cotidiano las fuentes se ubican en dicha posición.

Se detalla a continuación el plano con las posiciones de las mediciones realizadas:

1) Vista de Planta con las posiciones de medición 1, 2, 3, 4, 5 y 6:

2) Sección vertical donde se muestran las posiciones de medición en los palcos (P7 y P8) y en la Cazuela (P9):

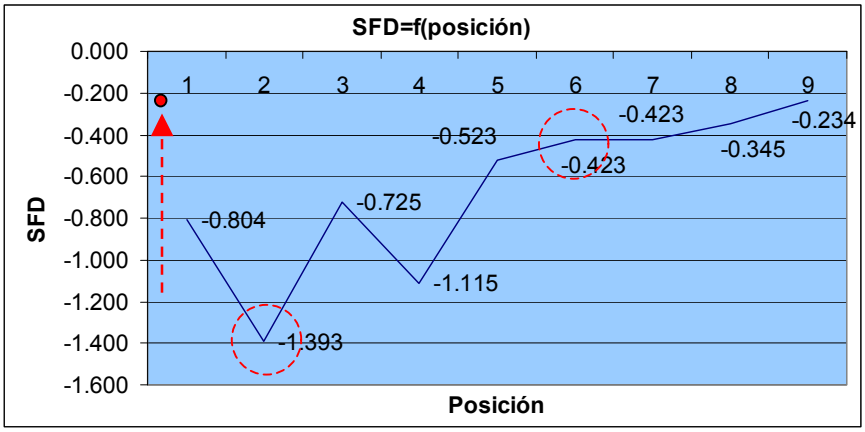




En este caso la confección del gráfico SFD=f(posición) nos permite realizar un *análisis de zonas de percepción sonora* de la sala. Se puede comparar el valor del SFD obtenido en la única posición medida dentro del Salón Dorado del mismo edificio.

El punto rojo apuntado por una flecha del mismo color indica el valor de SFD del Salón Dorado del Teatro Colón, el cual tiene un RT60 monoaural de 1,1 segundos a frecuencias medias. Se evidencia un valor de difusividad similar al que se encuentra en la Cazuela (P9) de la Sala principal, razón por la cual es posible justificar las excelentes cualidades sonoras subjetivas con las que el público oyente siempre pondera del lugar.

Del gráfico del SFD=f(Posición) y de las respuestas al impulso binaurales podemos explicar la tendencia de emisión vocal en diagonal al público (y no frontal) por parte de las voces pequeñas. La causa radica en la reducida difusividad presente en el centro de la sala (posiciones 2 y 4) lo que se evidencia como menor nivel sonoro (pudiendo ser fácilmente enmascarada la fuente) y baja sensación estéreo en dichos lugares respecto de otras zonas con mayor SFD. La solución a esta característica de todas las grandes salas es orientar las fuentes sonoras de menor nivel sonoro hacia el lateral más cercano, generando así las reflexiones que en caso de emitir directamente hacia el frente no tendría, las que le proporcionarán mayor nivel sonoro y sensación de espacialidad y envolvimiento.



Sensibilidad y minimización de los posibles errores sistemáticos en la medición del SFD:

La información binaural para el post procesamiento pueden ser:

1) Directamente las respuestas al impulso $h(t)_L$ y $h(t)_R$ obtenidas con los métodos *Deconvolución de MLS* y *Log Sine Sweep*.
2) La captación binaural de una señal de extremadamente baja autocorrelación para $t \neq 0$. Esta condición la cumple nuevamente una señal *MLS White*.

El SFD presenta gran sensibilidad a (en orden descendente de importancia):

- La Autocorrelación de la señal de excitación. Razón por la cual se debe utilizar MLS White o utilizar un método apropiado de restitución de la señal plana a partir del MLS rosa.
- La estabilidad del Clock del formato de registro en sistemas de deconvolución de señal MLS.
- La S/N alcanzable en el recinto por el Nivel de Presión Sonora de la fuente.
- Los niveles relativos entre L y R, lo que implica una calibración apareada inicial de las ganancias de los preamplificadores de los micrófonos binaurales.

- La alteración *por medios electrónicos* de los tiempos de arribo L y R en el método de medición.
- La alteración *física* de tiempos de arribo L y R por lo que es fundamental la correcta elección de la posición de la cabeza que contiene los micrófonos binaurales.El procesamiento estadístico de la FFT (DFT) al hallar la respuesta al impulso de un sistema.
- El ruido propio resultante del Sistema de Medición (ej.: 2 ch FFT).
- Al intervalo temporal de análisis (en este trabajo se estableció en 60 segundos).
- Los ruidos de procedencia lateral.
- La invariancia en el tiempo del sistema (gente, corrientes de aire, etc.).

Conclusiones:

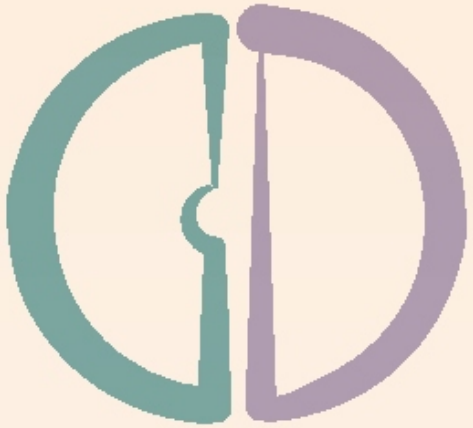
- Se observó que el SFD es un parámetro ortogonal al RT60.
- Se observó una gran sensibilidad del SFD ciertos parámetros los que deberán ser controlados en toda medición para minimizar los errores sistemáticos.
- Se desarrolló un método de medición para completar el posterior procesamiento y la adquisición de datos acústicos complementarios.
- Se comparó el análisis de recintos en base a la respuesta al impulso tradicional con el estudio del mapeo del *SFD=f(posiciones)* y el posterior análisis del *RT60 binaural*, dando por resultado un mucho mayor conocimiento del funcionamiento acústico de los recintos por medio del último.
- También se observó la gran practicidad de conocer el SFD en las etapas de análisis y proyecto acústico.
- El SFD es el grado de densidad de reflexiones existentes dentro de una respuesta al impulso.
- El SFD mide, en cierta forma, la calidad de la reverberación presente en los puntos de medición dentro de un recinto.

Agradecimientos:

Dr. Angelo Farina.

Bibliografía:

Yoichi Ando, “Architectural Acoustics”.
Leonid I. Makrinenko, “Acoustics of Auditoriums in Public Buildings”.
Malcolm Crocker, “Handbook of Acoustics”.
Leo beranek, “Concert and Opera halls, how they sound”.
Alejandro Bidondo, “Difusores Acústicos I” y “Difusores Acústicos II”, memorias del Congreso mexicano de Acústica años 2000 y 2001 respectivamente.



CLICDATA
MULTIMÍDIA

Projeto e produção do CD

Fone/fax: (48) 232.8690

clicdata@clicdata.com.br

www.clicdata.com.br