Audio Engineering Society

Convention Paper 6738

Presented at the 120th Convention
2006 May 20–23 Paris, France

Adaptive Audio Equalization of Rooms based on a Technique of Transparent Insertion of Acoustic Probe Signals*

Ariel F. Rocha[1], António Leite[2], Francisco Pinto[3] and Aníbal J. S. Ferreira[4]

[1] INESC Porto, Porto, Portugal
arocha@inescporto.pt

[2] INESC Porto, Porto, Portugal
"<Click here and type Author[2] Email Address>"

[3] INESC Porto, Porto, Portugal
fmiguelpinto@gmail.com

[4] University of Porto, Portugal; INESC Porto, Porto, Portugal
ajf@fe.up.pt

ABSTRACT

This paper presents a new method performing real-time adaptive equalization of room acoustics in the frequency domain. The developed method obtains the frequency response of the room by means of the transparent insertion of a certain number of acoustic probe signals into the main audio spectrum. The opportunities for the insertion of tones are identified by means of a spectral analysis of the audio signal and using a psychoacoustic model of frequency masking. This enhanced version of the adaptive equalizer will be explained as well as its real time implementation on a TMS320C6713 DSP based platform. Finally the results of the acoustic tests and conclusions about its performance will be presented.

1. INTRODUCTION

The problem of room acoustics has been widely investigated essentially with the purpose to improve certain quality parameters, when capturing sound that propagates in a closed acoustical space such as a room or an auditorium, independently of the place where the sound is captured. The most conditioning aspects concerning room acoustics are its geometry, volume and

*This work has been supported by the Portuguese FCT-POSI and European FSE/FEDER research programs, under project POSI/CPS/48517/2002.

type of surfaces, besides other interfering elements in its interior such as persons, chairs, curtains, etc. Due to the fact that the room frequency response can be modified by time varying factors such as the number of persons inside the room at each instant, the temperature of the room, etc, one possible approach in acoustical room treatment is related with the dynamic monitoring of the sound reproduction automatically during the audio performance.

In this paper we present a room acoustics equalizer based on digital signal processing. The developed system is able to estimate and compensate the asymmetries in the room response when these asymmetries depend on time varying acoustic conditions.

This work represents the continuation of a previously developed adaptive 20-band room equalizer by Leitão, Fernandes and Ferreira [1]. A subsequent optimization of that system was performed and presented by Leite and Ferreira in [2], introducing several improvements into its performance by means of the optimization of the adaptation rule and some parameters of the adaptive process. That system performs adaptive but not transparent equalization, limiting its application to a stage of acoustic correction previous to the audio program execution. The present work is the conclusion of the previous developments and represents an extension of the adaptive filtering technique towards a more powerful version able to compensate the asymmetries of the room response in real-time and in an adaptive way with the innovation that these asymmetries are evaluated in a transparent way to the listeners, without interrupting the audio presentation.

The paper is organized as follows: in Section 2 the problem of a time varying room response is characterized as well as the limitation of previous developments in addressing this problem. Section 3 describes the technique to perform adaptive equalization proposed in this work by means of the transparent insertion of acoustic test signals in the audio material in order to estimate the current room response we want to compensate. In Section 4, a technique to produce proper test signals is explained. Section 5 presents the main concepts about the psycho-acoustic model of frequency masking applied to identify conditions adequate for the transparent insertion of the test signals. In Section 6, the technique used for the insertion of tones is explained. Section 7 describes the method to estimate the current room response from the inserted test signals. In section 8 the algorithm to perform the adaptation of the equalizer parameters is presented. Section 9 refers to the DSP platform applied for the real time implementation of the system. In Section 10, a graphical interface developed to visualize in real time several internal signals is described. Section 11 presents a few results obtained from different acoustic tests as well as other indicators of the performance achieved by the system. Finally, Section 12 concludes the paper.

2. PROBLEM CHARACTERIZATION

As it was already mentioned, this work is based on a previously developed digital 20-band adaptive equalizer that performs the compensation of the room response in real-time and in an adaptive way by means of a technique of fast FIR filtering in the frequency domain. A simplified block diagram is shown in Figure 1 for convenience.

This version of the adaptive equalizer uses white noise as a room excitation signal in order to estimate its frequency response. For this reason the equalization process is not transparent to the listeners. In other words, that means that it can not be applied during the presentation of the audio program. Therefore its application remains limited to a stage of acoustic correction previous to the performance ("sound checking").

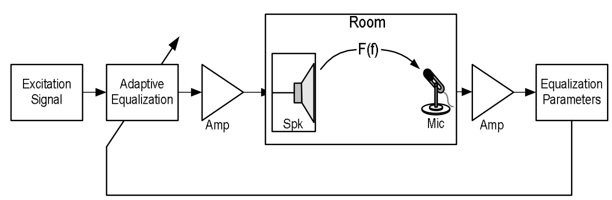

Figure 1: Simplified block diagram of the previous development

Even though this is a very useful procedure that allows establishing an initial setting for the room equalization, it is well known in practice that several parameters affecting the room acoustics usually change as a consequence of variations in the audience size and distribution, air temperature and humidity, etc. This problem makes interesting the possibility of an automatic detection of these dynamic variations of the room acoustics and the reconfiguration of the equalization parameters during the course of the audio program.

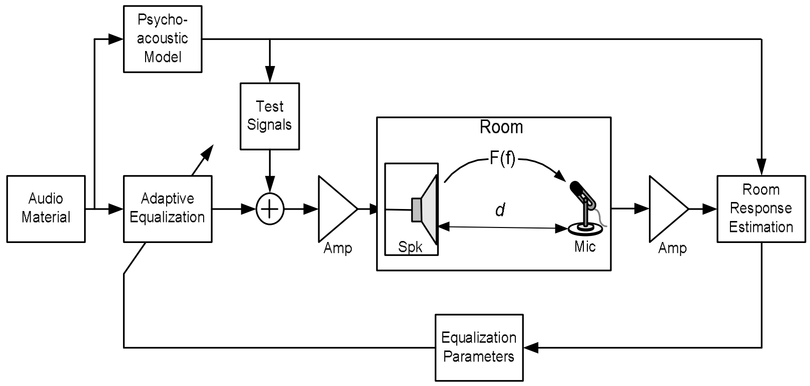

Figure 2: Simplified Diagram of the Proposed Idea

3. A NEW TECHNIQUE FOR REAL TIME ADAPTIVE EQUALIZATION

The main goal of the new implementation is to obtain a periodical estimation of the room frequency response in order to adapt the FIR filter coefficients with the objective of compensating in real time the asymmetries between the room response and the desired response, which can vary along the time because of possible changes in the room acoustics.

The central idea is to obtain the room response estimation by means of the insertion of acoustic test signals generated artificially, in such a way that they remain unperceived to the listeners.

The system uses pure sinusoids (tones) as test signals because of they can be generated easily during a short time and are clearly identifiable when coming inside the main audio spectrum. Moreover, sinusoids are naturally similar to the original audio components and therefore they can be integrated to the acoustic signal without noticeable disturbances.

In order to make this process transparent for the audience the proposed technique takes advantage of the frequency masking effect that is a characteristic of the human hearing. This psycho-acoustic effect is used to find proper regions in the main audio spectrum in which an inserted tone will be masked by components of the original audio signal as long as the tone amplitude remains under a certain threshold level.

This method requires the production of a psycho-acoustic model of the original audio signal as we will see in the next sections. This development is also based on the study of room acoustics characteristics performed by G. Fernandes in [3].

Figure 2 presents a general block diagram of the developed equalization system that can be compared with the previous version of the adaptive equalizer.

At this moment we have not found other scientific works that implement in practice a system based on this idea of a transparent insertion of sinusoids in the acoustic signal, by applying a psycho-acoustic model, in order to estimate the room response in real-time as a base for the adaptive equalization.

However, looking at the commercial market of professional audio equipments we found a model of adaptive equalizer, developed by Sabine ™ that implements a similar concept of transparent insertion of sinusoids to obtain the room response [4]. After analyzing the specifications of this commercial equalizer we believe that our method is more flexible since it is able to insert test tones in any location of the discrete audio spectrum and with more convenient amplitude levels.

Moreover the technique for identification of proper insertion points based on the masking effect allows a better use of the opportunities provided by the original audio material.

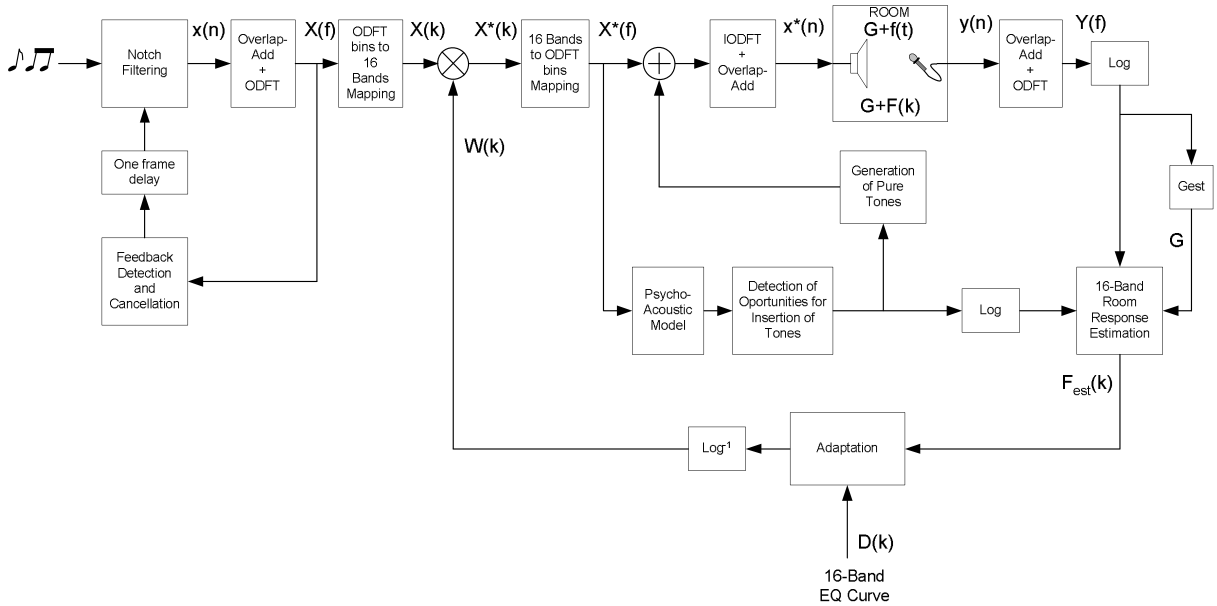

Figure 3: General Block Diagram of the Adaptive Equalization System by Transparent Insertion of Tones

Figure 3 presents a block diagram of the system performing adaptive equalization by transparent insertion of test tones that has been implemented on a DSP platform for real time operation.

In that figure, it can be observed, at the initial part of the processing, a set of blocks performing the detection and cancellation of eventual interferences caused by positive acoustic feedbacks between the loudspeaker and the test microphone. This additional feature was developed and presented in [5] and is able to detect and suppress automatically multiple disturbances produced by simultaneous feedbacks.

3.1. Analysis Framework

The method of adaptive equalization implemented in this work is based on techniques of fast FIR filtering in the frequency domain [1]. The real-time implementation of the developed system applies a scheme of time-frequency analysis by means of a very efficient algorithm of linear convolution [6], based on the well known methods *Overlap-and-Add* (OLA) and *Overlap-and-Save* (OLS) [7]. These methods of circular convolution of a vector are applied, in this case, to perform the linear convolution of the audio sequence with the equalization filter.

Several reasons related to convenient symmetry properties, computational efficiency and coherence with the previous developments on which this work is based, justify the use of the Odd Discrete Fourier Transform (ODFT) [8] [9] as a method to perform the transformation of the audio signal from time domain to frequency domain.

According to the implemented model, the input audio sequence $x(n)$ is divided in successive segments. Each segment is multiplied by an analysis window $h(n)$ and transformed to the frequency domain by computing a direct ODFT of 1024 points. Then, the spectral coefficients are mapped into the 16 equalization bands and multiplied by the coefficients of the adaptive filter, $w(k)$. Once performed the filtering in frequency domain, the 16 compensated bands are converted to 512 frequency coefficients and the inverse ODFT is calculated to obtain the temporal output segment. Finally, successive segments are multiplied by a synthesis window $h(n)$ and overlapped and added to obtain the output audio sequence.

3.2. Equalization Bands Scheme

The previous version of the adaptive equalizer implemented a non-linear division of the audible spectrum in 20 equalization bands (also referred as sub-bands) based on the model of critical bands of the human auditory system [6]. This model tries to imitate the frequency resolution of the human hear and is useful to characterize perceptual phenomena like loudness perception and masking effects. The implementation of this model produces transition bandwidths between each two equalization bands with different sizes according to the region considered of the audible spectrum.

In this work, the original scheme of 20 bands was reduced to a format of 16 non-uniform bands due to several reasons related to the number of spectral bins occupied for each band (according to the spectral resolution determined by the ODFT length and the sampling frequency) and the probability of detecting an sufficient number of opportunities of insertion of test tones inside each band. Those reasons will be better explained in the following sections. The central frequency and width of each one of the 16 bands are shown in Table 1.

Band	Central Frequency (Hz)	Width (Hz)	Lowest Bin	Highest Bin	Width (bins)
0	70,3	187,5	0	3	4
1	257,8	187,5	4	7	4
2	445,3	187,5	8	11	4
3	632,8	187,5	12	15	4
4	867,2	281,25	16	21	6
5	1195,3	375	22	29	8
6	1617,2	468,75	30	39	10
7	2179,7	656,25	40	53	14
8	2929,7	843,75	54	71	18
9	3867,2	1031,25	72	93	22
10	5226,6	1687,5	94	129	36
11	7101,6	2062,5	130	173	44
12	9445,3	2625	174	229	56
13	12445,3	3375	230	301	72
14	16289,1	4312,5	302	393	92
15	21210,9	5531,25	394	511	118

Table 1: Definition of the 16 Equalization Bands

The characteristic of spectral selectivity of the equalizer was defined for the new scheme according to the same model of critical bands and is depicted in Figure 4.

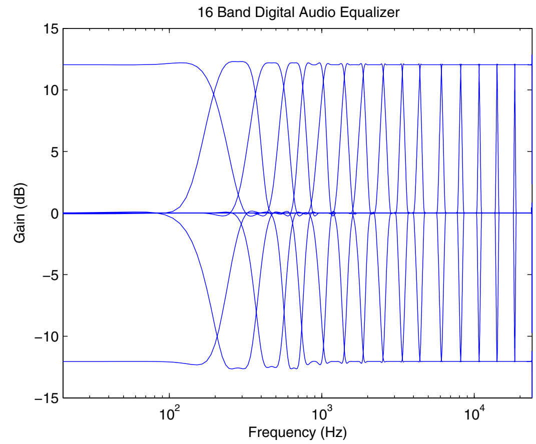

Figure 4: Spectral Selectivity Characteristic of the 16 Band Adaptive Equalizer

4. GENERATION OF PURE SINUSOIDS

The artificial generation of the test sinusoids is performed by means of a method of synthesis from the ODFT representation of the wanted tone. The method is explained in the following paragraphs.

4.1. Method of Spectral Generation of Sinusoids

The time analysis/synthesis window in our framework, *h(n)*, is a symmetric window defined as the square root of a *Hanning* window [9]:

$$h(n) = \sin\frac{\pi}{N}\left(n+\frac{1}{2}\right);\ 0 \le n \le N\text{-}1. \qquad (1)$$

This window is commonly used in audio coding since it satisfies the requirement of a perfect reconstruction of the original signal and is analytically tractable.

As it is demonstrated by Ferreira in [9], when this analysis window is combined with the ODFT filter bank a number of interesting results are obtained. For instance, when the input audio signal, *x(n)*, is a pure tone whose frequency is such that, due to the contribution of *h(n)*, it corresponds to a sampled frequency of the ODFT, $\ell\frac{2\pi}{N}$, with ℓ in the range $\left[1, \frac{N}{2}-1\right]$, only two spectral lines in the ODFT

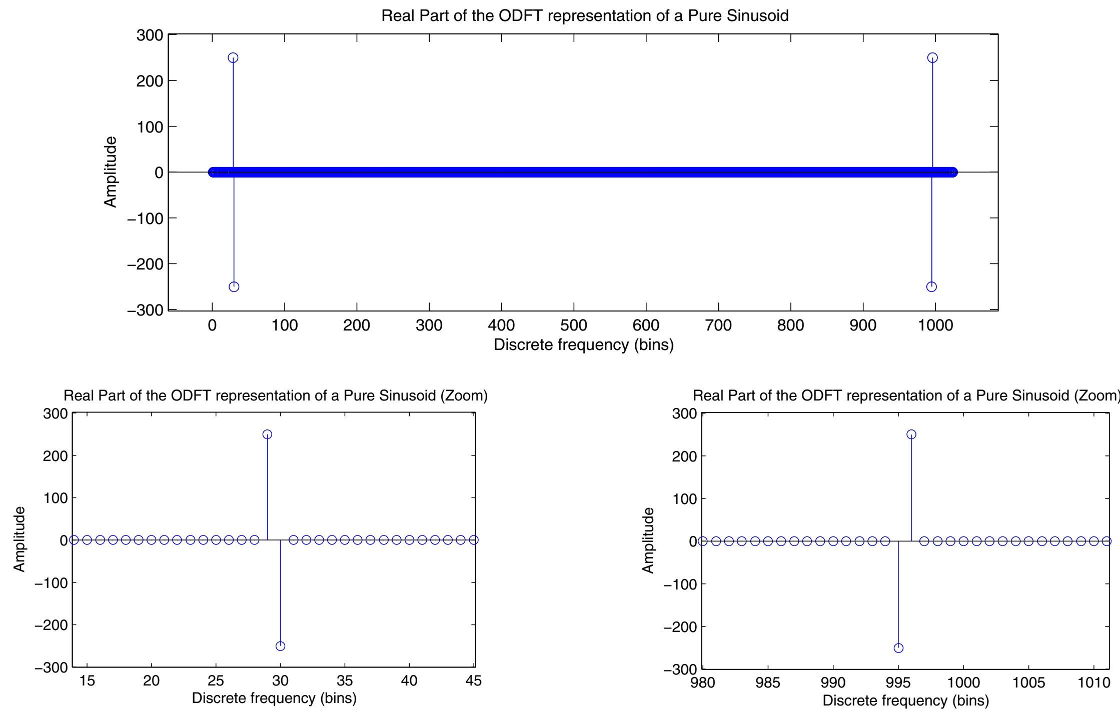

Figure 5: Real part of the ODFT representation of a pure sinusoid (Above). Zoom at the spectral lines corresponding to the sinusoid, in the first and second half of the spectrum (Below).

magnitude spectrum and in the Nyquist range, are different from zero.

In fact, if *x(n)* is given by:

$$x(n) = A\sin\left[\frac{2\pi}{N}\ell n + \phi\right] \qquad (2)$$

Then:

$$X_O(k) = \frac{NA}{4}\left[e^{-j\left(\frac{\pi}{2N}-\phi\right)}\delta(k-\ell+1) + e^{j\left(\pi-\frac{\pi}{2N}-\phi\right)}\delta(k-\ell)\right]$$

$$0 \le k \le \frac{N}{2}-1 \qquad (3)$$

The meaning of this result is that in the set of unique Odd-DFT coefficients, an exactly sampled pure tone is mapped to a two consecutive spectral lines. This is the result of a (suppressed carrier) double-side-band modulation effect caused by the sine window *h(n)*.

Further analysis of the phase of these spectral lines reveals that the fixed phase ϕ in Eq. (*x(n)* = …) is available from the phase of the spectral line $k = \ell - 1$ after adding $\frac{\pi}{2N}$. Moreover, the difference between the phase at $k = \ell$ and the phase at $k = \ell - 1$ is exactly $\pi\left(\frac{1}{N}-1\right)$, regardless of ϕ.

Based on the previous considerations, we implemented a technique for the generation of pure sinusoids from its ODFT spectral representation.

The generation process starts with the definition of the bins corresponding to the desired frequency, as if we would "draw" spectral lines at these points. Then, after the application of the IODFT, a highly precise sinusoidal signal is obtained in the time domain. The practical rule for the pure sinusoids generation algorithm can be summarized as:

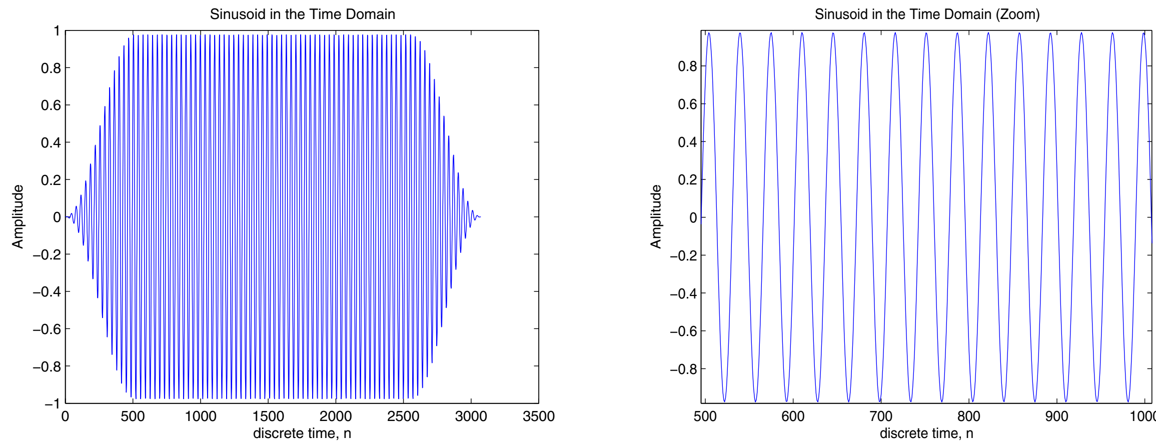

Figure 6: Sinusoid of f = 1400 Hz generated by the algorithm in the discrete time domain, covering its total duration (Left). Expansion in time of the upper graph to verify the precision of the sinusoidal shape (Rigth).

- From the wished frequency f_t, the value of ℓ is defined as the integer value closest to: $\frac{f_t}{f_s}N$, where f_s is the sampling frequency.

- The lowest half of the ODFT Real Part sequence, in the range [1, N/2 – 1], has two spectral lines at $\ell - 1$ and ℓ, with identical amplitude but opposite sign because of the approximate phase difference of π between them.

- Due to the symmetry property of the ODFT regarding N/2, the highest half of the ODFT Real Part, in the range [N/2, N-1], has two spectral lines at $N - \ell - 1$ and $N - \ell$ whose respective amplitudes are conjugated to the corresponding lines in the first half. The rest of elements of the Real Part sequence are zeros.

- If the value of $\ell - 1$ lies in an odd number, the sign of the ODFT spectral lines must change to its opposite value in each consecutive frame.

- The Imaginary Part of the pure sinusoid ODFT spectrum is always a vector of N zeros.

This method is able to produce highly pure sinusoids in a "clean" way and with a frequency resolution established by the discrete ODFT bin-scale. In this case, using a sampling frequency f_s = 48000 Hz and an ODFT of length 1024, the frequency resolution is 46.875 Hz.

In order to validate this technique, several tests were performed to generate sinusoidal tones of different frequencies. As a result, a set of graphs of the generated signals in the time and frequency domains were obtained.

Figure 5 shows the Real Part of the ODFT corresponding to a sinusoid of 1400 Hz, having spectral lines at the bins 29 and 30, in the first half and 995 and 996 in the second half. It can be observed the conjugated values of the amplitudes with respect to N/2.

Figure 6 represents the discrete time signal corresponding to the sinusoid of 1400 Hz generated from its ODFT during a period of approx. 60 ms where it is possible to observe the smoothing effect of the time analysis window.

4.2. Insertion of Test Tones in the Audio Spectrum

The technique for generation of pure sinusoids is applied in order to insert tones in the audio material during a brief period of time. This is the foundation of the method proposed for the estimation of the room response. The insertion of the sinusoids is performed in the discrete frequency domain by the addition to the original audio spectrum of the probe sinusoid spectrum.

With the objective of validate this method, an algorithm was implemented that captures a digital audio signal and inserts during a certain period of time a sinusoid artificially generated at one selected point of its spectrum. In order to smooth the insertion process the algorithm previously generates a thin gap in the spectral region near to the insertion point. This is carried out by applying a highly selective notch filtering at that position. The notch filter is implemented by a second order IIR model, whose selectivity has been verified as sufficient [5].

The implemented algorithm was evaluated taking several audio signals from wave files and inserting probe tones in different spectral positions. For every case a new wave file was obtained, containing the audible evidence of the presence of the inserted tone.

A region of the audio spectrum before and after the tone insertion is depicted in Figure 7. Only a difference in the shape of the original spectrum is observed in the second graph, at the proximity of the inserted tone, due to the notch filtering, but the effect is not perceived by listening to the resulting audio file.

It was verified that the implemented technique is able to produce a thin hole in the main audio spectrum (whose width is defined by the zero-pole relation parameter of the IIR model [5]) and insert a pure sinusoidal tone in that hole in an accurate way.

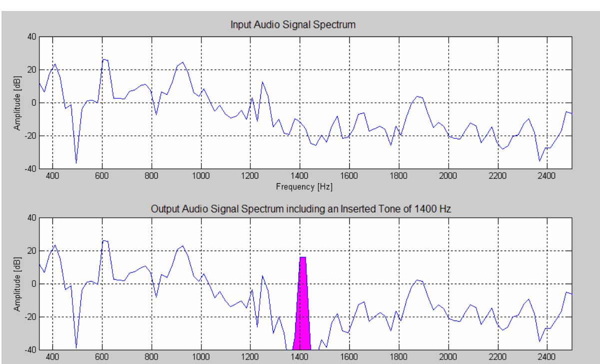

Figure 7: A portion of the main audio spectrum with its original shape (Above), and the same portion after the notch filtering and the insertion of a sinusoidal tone (Below).

However, because of the non uniform division of the ODFT spectrum in the equalization sub-bands, the bands corresponding to the region of low frequencies cover significantly less bins than those for higher frequencies. (For instance, the lowest sub-band covers 4 bins while the highest one covers 118 bins). As a consequence, a notch filter applied at the lowest bands can produce a noticeable degradation of the sound quality. Moreover, the continue activation and deactivation of notches in those bands can also produce disturbances. A “clean” notch activation/deactivation process would require higher computational complexity and more DSP resources. For this reason, the pre-filtering notch to create a “gap” in the spectrum was discarded and a simple technique of substitution of bins was implemented that replace the original couple of bins with the two bins corresponding to the inserted tone.

5. FREQUENCY MASKING EFFECT

The masking effect is a characteristic of the human auditory nervous system and is derived from its physiological properties. In the following paragraphs, some important concepts describing this phenomenon are presented, which have been extracted from [9] and [10].

The masking effect can be understood like a kind of interference with the audibility of a sound (called probe or “maskee”) caused by the presence of another sound (called masker), when both these sounds are close enough to each other in frequency and occur simultaneously or closely to each other in time.

If a lower level probe sound is inaudible because of a simultaneous higher level masker, the effect is referred to as the *simultaneous masking* or *frequency masking*. Masking is typically described by the minimum shift if the probe intensity level above its threshold of audibility in quiet, necessary for the probe to be heard in the presence of masker. The *Threshold of Masking* corresponds to a limit case when the masking conditions are such that the “maskee” is in the threshold between audibility and inaudibility. When many simultaneous maskers occur together, the overall masking effect as a function of frequency can be determined by the *Global Threshold of Hearing* [10]. In this paper we will refer to it as “Masking Threshold”.

This interesting phenomenon of masking has been widely studied by the Psychoacoustics and is extremely important for the design of digital perceptual audio coders because it can be exploited to make the quantization noise inaudible. The simultaneous masking effect is characterized by the following aspects observed from practical experiences:

- The higher the level of the masker, the greater is the masking.
- Masking decreases as probe frequency gets closer to that of the masker.
- Masking is greater on frequencies above the masker frequency than on frequencies below it.
- Due to a nonlinearity of the human hearing system, the masking curve has similar shape for various masker harmonics.

In this work we take advantage of the masking effect to "hide" the test sinusoids in the audio signal. The spectral regions of the audio material that are most suitable to the insertion of test tones are identified based on the analysis of the Masking Threshold curve when this curve reveals the presence of some masker in the original audio spectrum. The amplitude of each test tone is also calculated from the Masking Threshold level at the test position in order to make the tone masked by its adjacent spectrum and therefore imperceptible for the listeners.

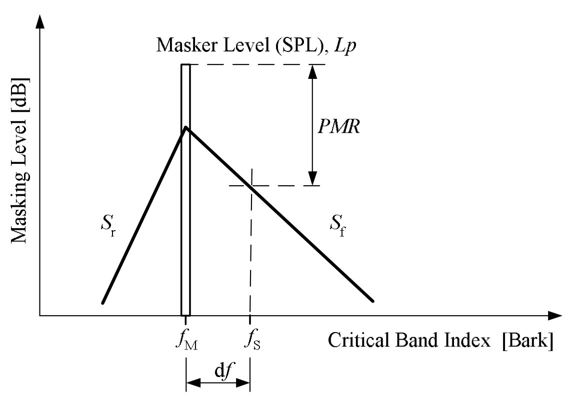

Figure 8: Simplified model of the masking threshold level

A simplified model of the Masking Threshold is used to represent the masking effect produced by each frequency component of the audio signal when considered as a masker. As shown in Figure 8, that model is based on a triangular shape, using two different slopes for the masking level: a "rising slope", *Sr*, on the left of the masker and a "falling slope", *Sf*, on its right.

PMR means the "Probe-to-Masker Ratio" at the frequency *fs* and is calculated in dB taking account into the *Absolute Threshold of Hearing*, which represents the minimum sound intensity of a pure tone making it just audible [9]. The abscissa scale used in this model is Bark which represents the natural organization of frequency as perceived by the human auditory system in *Critical Bands* (related with the hearing physiology) [9]. By definition, one Bark corresponds to the width of one critical band.

From the referred bibliography, we know that while the rising slope S_r of the masking threshold triangle is approximately constant, the falling slope S_f depends on the masker level (SPL) in dB and its center frequency in KHz, as expressed by the following equation [10]:

$$Sf = -24 + 0.2L_p - 0.23 f_c^{-1} \quad (4)$$

Where: S_f = Falling Slope [dB/Bark]
L_p = Masker tone level SPL [dB]
f_M =Center frequency of the masker tone [KHz]

By means of the application of this psycho-acoustic model, the system obtains the Masking Threshold curve in each frame. The algorithm analyzes the masking effect at each point of the audio ODFT spectrum, generically identified in Figure 8 as f_S. An analysis band is defined in the vicinity of f_S, covering a certain number of bins at both sides of that point. Then, each bin in this band is considered as a masker tone and the model of Figure 8 is applied in order to obtain the PMR at f_S from the power level of the masker tone (at the frequency f_M).

Once obtained the PMR values from all the bins of the analysis band, these contributions are added. As a result, the masking threshold level at f_S is obtained as a consequence of the "addition of masking" effect [9] produced by all the adjacent bins in the analysis band. This procedure is repeated for all the bins of the ODFT spectrum. Thus, the masking threshold curve is obtained for the current frame.

As an example of this, Figure 9 shows a portion of the PSD curve corresponding to an audio signal with its masking threshold curve superimposed.

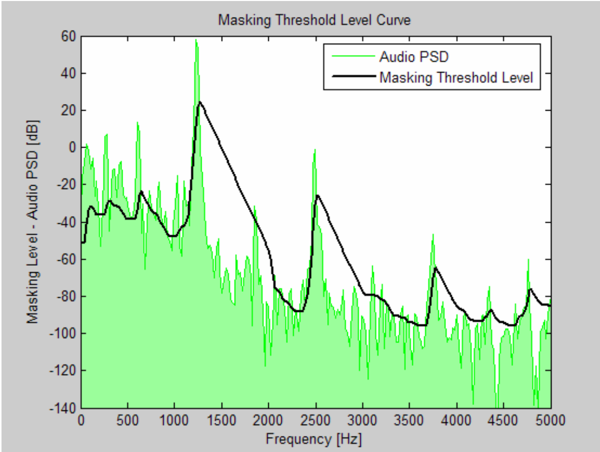

Figure 9: Masking Threshold Level curve over the PSD representation of the audio signal

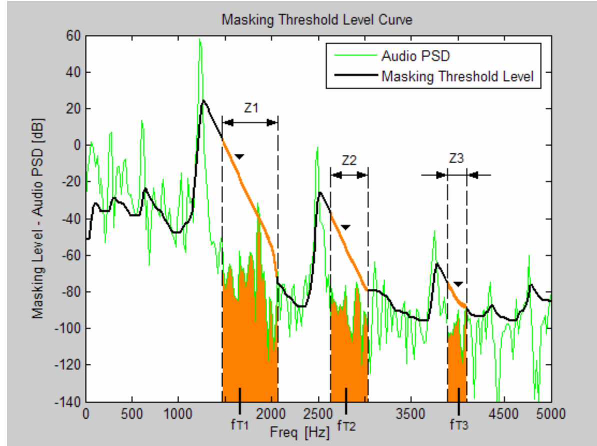

Figure 10: A portion of the ODFT audio spectrum with the corresponding masking threshold curve, indicating the zones favorable for tone insertion identified by the algorithm (highlighted areas) and the most appropriate insertion points inside each favorable zone (with a small triangle).

6. IDENTIFICATION OF OPPORTUNITIES AND INSERTION OF TEST TONES

The opportunities for the proper insertion of probe tones in the main audio spectrum are identified by means of the analysis of the masking threshold curve. The algorithm that performs this process was developed in an experimental way based on conclusions extracted from practical tests.

6.1. Identification of Spectral Conditions Favorable to the Tone Insertion

As it was already observed, the presence of masker tones in the audio signal produces local maxima in the masking threshold curve. The spectral regions that are most appropriates for the transparent insertion of a test tone, are located at the right of the masker tones and therefore they can be identified by a falling flank at right of a local maximum in the masking threshold curve.

Once an appropriate spectral zone is identified, the most adequate bin for insertion inside that zone is determined taking account into the shape of the spectrum that makes possible the highest separation in amplitude between the future test tone and its lateral bins in such a way that the tone be affected by the adjacent spectrum as weakly as possible.

Figure 10 shows the same portion of the audio spectrum considered in Figure 9, but now with three favorable zones highlighted (called Z1, Z2 and Z3) and the most suitable insertion points identified inside each favorable zone by applying the explained criterion (f_{T1}, f_{T2} and f_{T3}).

6.2. Test Tone Amplitude Calculation Rule

The rule that establishes the amplitude of the probe tones to be inserted in the audio spectrum is based on the relation between masking threshold level at the test frequency and the value of the mean energy corresponding to a narrow band centered at the same frequency. This rule was implemented and optimized based on the listening to the resulting audio signal in order to verify that the insertion of probe tones is not perceived in an audible way (due to the frequency masking effect). In the following, the rule is presented.

Calling f_T the frequency of a generic insertion point detected by the algorithm, let's define:

$M(f_T)$ = Masking Threshold level at the test frequency bin f_T.

E_{BAND} = Average Energy level in a band of the original input signal spectrum, formed by 6 bins, centered at the bin f_T. It is computed using the following equation:

$$E_{BAND} = 10 \times \log_{10}\left[\frac{\sum_{k=L-2}^{L+3} |X(k)|^2}{6}\right], \quad (5)$$

Then, the value of the test tone Power, P_T, can be obtained by using the following rule:

$$\begin{cases} E_{BAND} + 10 \leq M(f_T) & \Rightarrow \quad P_T = M(f_T) \\ E_{BAND} \leq M(f_T) < E_{BAND} + 10 & \Rightarrow \quad P_T = E_{BAND} + 10 \\ E_{BAND} - 10 \leq M(f_T) < E_{BAND} & \Rightarrow \quad P_T = E_{BAND} \\ M(f_T) < E_{BAND} - 10 & \Rightarrow \quad P_T = M(f_T) + 10 \end{cases} \quad (6)$$

where all the magnitudes are expressed in dB.

Once obtained the power magnitude of the test tone, its amplitude A_T is calculated as:

$$A_T = \sqrt{10^{\frac{P_T}{10}}} \quad (7)$$

Figure 11 depicts a segment of the original audio spectrum in which a number of test tones were inserted based on the analysis of the Masking Threshold curve.

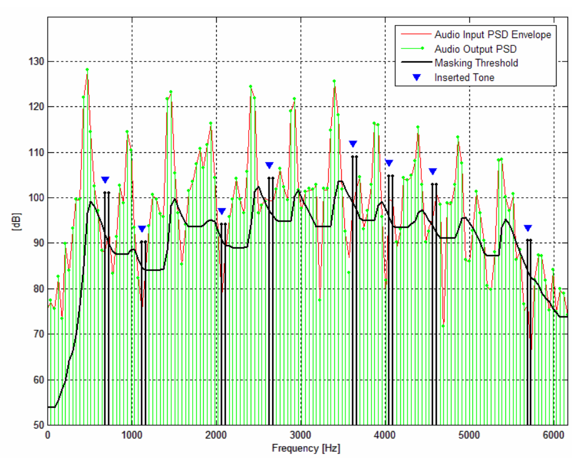

Figure 11: A portion of the original ODFT audio spectrum with several tones inserted in the locations detected by means of the implemented technique.

6.3. Activation of the Test Tones

Whenever a zone favorable to the possible insertion of a tone is identified in the spectrum, the continuity in time of that condition is verified along a minimum period of approx. 10 ms. If the condition remains after that time, the insertion of a test tone is activated at the most appropriate bin inside the detected zone, which is established by the algorithm. Each tone is treated as an independent test object having one of several possible states. These states are related to the insertion of the tone according the spectral condition verified and are defined as:

- **"Available":** The test object is disabled and waiting for a new opportunity of being inserted in the spectrum

- **"Candidate":** A new opportunity of insertion was identified. The test object was assigned to that opportunity and is waiting for the minimum permanence in time of that condition is verified. The tone frequency is assigned to the most suitable position inside the favorable region.

- **"Active":** The test tone was inserted into the audio spectrum and is present in the acoustic signal.

The algorithm that controls the activation, permanence and deactivation of each test tone can be represented by the state diagram depicted in Figure 12.

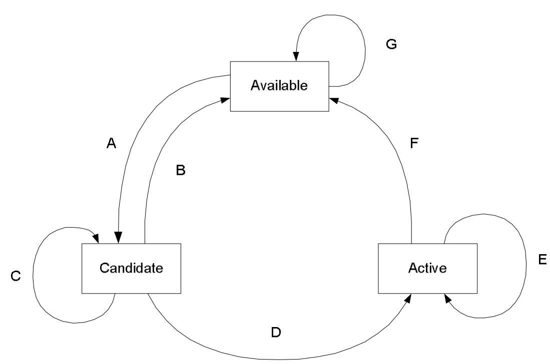

Figure 12: State Diagram of the test tones activation/deactivation process

References for Figure 12:

A. A new opportunity of insertion was detected in the current audio spectrum and the test object is assigned to that possibility.

B. The permanence in time of the condition favorable to the insertion is being evaluated and the minimum period of time is not yet achieved. The test object remains waiting for its insertion.

C. The opportunity does not remain after the minimum period of time.

D. The minimum continuity in time of the opportunity was verified and the test tone will be inserted in the audio spectrum.

E. The test tone is present in the acoustic signal and the condition favorable is verified once again. Therefore, the tone will remain inserted.

F. The condition favorable is no longer verified so as the tone must be removed from the audio signal and the test object will be disabled.

G. There is no opportunity of insertion associated to the object.

7. REAL TIME ESTIMATION OF THE ROOM FREQUENCY RESPONSE

Each tone inserted in the audio material is delivered to the room by the loudspeaker, as a part of the acoustic signal. The test tone is attenuated by the room acoustics during its propagation through the room and the magnitude of the same tone once captured by the microphone will be different from its original value. This attenuation represents the room response at the tone frequency for the current instant.

In order to obtain the attenuation in the tone magnitude, whenever a test tone is inserted in the audio signal the algorithm waits until the propagation time is completed along the distance *d* between the speaker and the test microphone (see Fig. 2). The DSP implementation of the adaptive equalizer is frame oriented so that the time resolution is measured in frames. According to the frame length and the sampling frequency used, the processing time for one frame is approx. of 10,7 ms.

If we call P_{To} the power level in dB of the original test tone and P_{Tr} the power level in dB when it is received, then the attenuation is calculated by means the following equation:

$$Att_T = P_{To} - G_o - \left(P_{Tr} - G_r\right) \qquad (8)$$

where G_o is the volume, in dB of, the audio signal outputted by the system and G_r is the volume, in dB, of the signal received in the microphone. Both these volumes are considered, in the attenuation calculation, in order to make the system independent to the amplification level produced by the sound reinforcement chain. This calculation is repeated for all the tones inserted in the audio signal captured by the microphone at the current instant. Then, the mean value of all the attenuations corresponding to tones inserted in a same sub-band is computed. Thus, the average attenuation is obtained for every sub band, as:

$$\overline{Att(k)} = \frac{1}{N_k}\sum_{i=1}^{N_k} Att_{Ti} \qquad k = 1,2,\ldots,16 \qquad (9)$$

Where the index i identifies each tone received in the same sub-band and N_k is the total number of tones in the band k at the current frame.

In order to obtain a more stable value of the room response, a criterion of temporal mean is applied for the calculation of the attenuation in each sub-band. Thus, the estimated room response *F(k)*, in the generic sub-band *k,* is obtained by computing the expected value of the attenuation for that sub-band considering the last *m* frames, as expressed by the following equation.

$$F_{est}(k) = E_m\left\{\overline{Att(k)}\right\} \qquad (10)$$

The room frequency response curve can be represented by 16 segments, each one corresponding to the value of $F_{est}(k)$ at each band, as shown in Figure 22. This curve is refreshed periodically by the algorithm, considering periods of time between 5 and 10 seconds. During this time, the value of the room response is considered unvarying.

8. ADAPTIVE FILTERING IN THE FREQUENCY DOMAIN

Once the attenuation values for each sub-band have been determined, the reciprocal filter that compensates the frequency response of the room is updated accordingly. The compensation is performed through a frequency-domain adaptive filter, described by:

$$W_{n+1}(k) = W_n(k) + \mu(k)\left[D(k) - F_{est}(k) - L_{EQ}(k)\right]$$

$$k = 1,2,\ldots,16 \qquad (11)$$

where $W_n(k)$ and $W_{n+1}(k)$ represent the old and updated filter coefficients, $\mu(k)$ the adaptation step, $L_{EQ}(k)$ is the current equalization level, $D(k)$ the desired equalization level, and $F_{est}(k)$ the estimated room response. All these magnitudes are expressed in dB. In order to clarify the result of Eq. 11 we should remember that the test tones are not affected by the equalization filtering (as shown in the block diagram of Figure 3) and consequently, the term $F_{est}(k)$ does not depend on the value of $W(k)$. Moreover, $F_{est}(k)$ remains constant during a time sufficient so that the convergence of the adaptive process of $W(k)$ can be attained (5 seconds, from the experiments).

This method, initially proposed by Leite and Ferreira in [2] and based on the adaptive filters theory [11], is an improved version that allows different adaptation steps for each sub-band. The performance of this adaptation method was successfully verified in the referred bibliography. Experiments suggest that, for the filter to converge properly, $\mu(16)$ must be set to at least one order of magnitude higher than $\mu(1)$.

9. IMPLEMENTATION OF THE EQUALIZER ON THE DSP PLATFORM

The previous version of the adaptive equalizer was implemented on a platform based on a floating point 32 bit digital signal processor with a 150 MHz clock rate. Given that the new version of the system requires more processing capacity as well as other real time data transmission features, the developed work also included a migration of the system towards a new DSP platform in order to be able of adding a more complex set of functions to the original code and in a stereo mode. The new system was implemented in real time on the TMS320C6713 platform [12], based on a floating point 32 bit DSP from TI with a 225 MHz clock rate.

10. VISUALIZATION THROUGH RTDX

Real-Time Data Exchange (RTDX) enables us to monitor all activity in real-time without interrupting the DSP process. To this effect, we created a Qt graphical application, in the host computer, that periodically requests an array of data from the target DSP and draws its contents in a 2D plotter. This way, we can promptly visualize all relevant data within the adaptation process - namely, the time-domain, frequency-domain and sub-band representations of the input and output signals. A print screen of the application is depicted in Figure 22.

11. EVALUATION OF THE DEVELOPED SYSTEM

11.1. Efficiency of the Method of Detection of Opportunities of Transparent Insertion of Tones

In order to evaluate the performance of the method of detection of opportunities based on the masking threshold analysis, we made a statistical count of the number of opportunities detected per sub-band along a certain period of time. To this effect, we used a CD player to generate the original audio signal.

- Audio Source: *Vivaldi - The Four Seasons (Felix Ayo)*, Song Nr. 3: *Spring*.

- Considered period: 20 seconds from the beginning of the track.

The statistical results of this test are presented in Table 2.

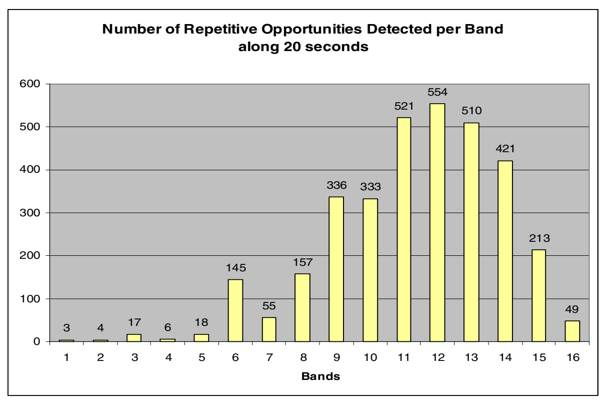

Figure 13: Distribution of repetitive opportunities along the 16 sub-bands.

When an opportunity for tone insertion is detected in the same frequency spot for two or more consecutive frames, we call it a repetitive opportunity. This condition is necessary to guarantee that the identified spot is in fact a consistent opportunity, and not just impulsive, for the insertion of the tone.

Band	Number of Consecutive Detections					Detected Opps.	Repetitive Detections	Frequency of the Repetitive Detections [Opps/sec]	Period between two consecutive Repetitive Detections [sec]
	1	2	3	4	5				
1	16	2	1	0	0	19	3	0,15	6,67
2	69	2	2	0	0	73	4	0,2	5
3	187	13	4	0	0	204	17	0,85	1,18
4	114	5	1	0	0	120	6	0,3	3,33
5	153	15	2	1	0	171	18	0,9	1,11
6	618	95	30	13	7	763	145	7,25	0,14
7	290	32	15	4	4	345	55	2,75	0,36
8	762	105	28	17	7	919	157	7,85	0,13
9	558	177	81	46	32	894	336	16,8	0,06
10	685	189	75	48	21	1018	333	16,65	0,06
11	1084	307	110	70	34	1605	521	26,05	0,04
12	1270	328	126	71	29	1824	554	27,7	0,04
13	1459	336	108	44	22	1969	510	25,5	0,04
14	1698	297	86	33	5	2119	421	21,05	0,05
15	1616	179	25	8	1	1829	213	10,65	0,09
16	1166	45	4	0	0	1215	49	2,45	0,41

From Table 2, we can conclude that the number of repetitive opportunities detected in all the bands, along the test period, is sufficient to estimate the room response in a consistent way.

Table 2: Statistics of detected opportunities of tone insertion

It can be observed that the bands corresponding to the lowest frequencies produce fewer opportunities than the other bands. This is a natural consequence of the fact that lowest bands have a smaller bandwidth and, consequently, a small probability of providing favorable conditions for tone insertion. For example, only 3 repetitive opportunities were detected in the first band in a period of 20 seconds. Nevertheless, tone insertion could be performed, in this band, with a periodicity of approximately 7 seconds. This value is very acceptable when considering the refresh ratio necessary for obtaining a continuous estimation of frequency response of the room.

Table 2 provided us a graphical representation - depicted in Figure 13 – of the distribution of repetitive opportunities detected in each sub-band. This chart clearly demonstrates the difference between the number of repetitive detections in lower and higher bands. As the width of the sub-band increases, the number of repetitive opportunities for tone insertion also increases. However, it is also noticeably that above 12th band the number of opportunities decays. This is explained by the fact that the audio signal has lower spectral components at very high frequencies, and thus less favorable conditions for frequency masking.

11.2. Prior Estimation of the Frequency Response of the Room

In order to evaluate the feasibility of the tone insertion method in determining the frequency response $F(f)$ of the room, we performed a prior estimation of the response using two different approaches: excitation by white noise and through insertion of individual tones. In both approaches, the signal was outputted through the speakers and then compared with the one received by the microphone. Each frame of the received signal provided an estimation of $F(f)$. The results were then averaged along 1500 frames.

11.2.1. Estimation through white noise

The first estimation of $F(f)$ was obtained by outputting white noise $N_w(f)$ through the speakers. Since the spectrum of the white noise is flat, the relative attenuations in each frequency of the received signal *X(f)* are equivalent to those that describe the frequency response of the room. Thus,

$$X(f) = F(f) \cdot N_w(f) = F(f) \cdot N_w \tag{12}$$

Figure 14 shows the frequency response of the room obtained by using this method. 16-band representation of *F(f)* is depicted in Figure 15. Since the curve of the estimated frequency response is extremely stable when averaged along 1500 frames, it represents a good approximation to the actual response of the room.

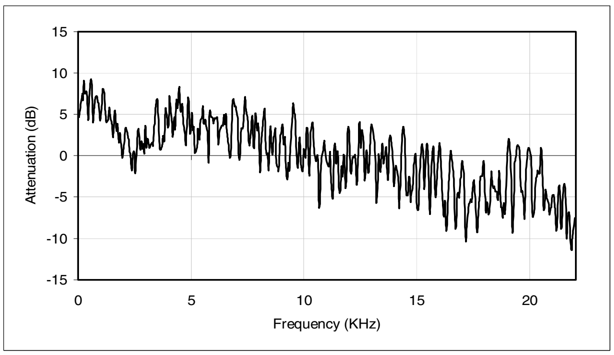

Figure 14: $F(f)$ of the room.

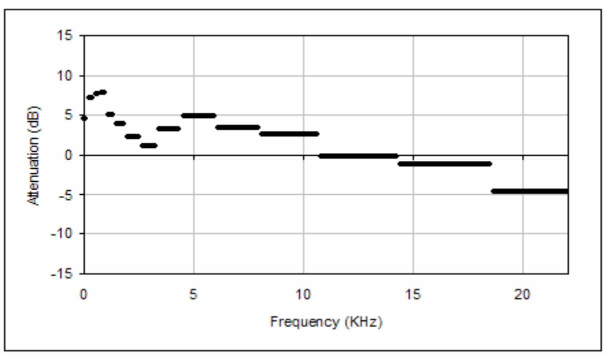

Figure 15: Sub-band representation of $F(f)$

11.2.2. Estimation through insertion of tones

The second estimation of *F(f)* was obtained by shifting one individual tone along the 512 bins of the output spectrum, and measuring the respective attenuation in each position. The complete room response representation is depicted in Figure 16 and the corresponding 16-band profile is shown in Figure 17.

By means of the comparison of the results obtained by the two previous methods (and considering eventual changes in the room acoustics between each test) we can conclude that the use of individual tones as a room excitation signal is an adequate technique able to estimate the room response with a sufficient approximation.

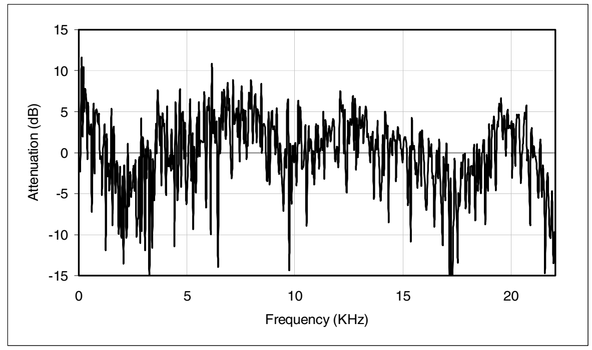

Figure 16: $F(f)$ of the room.

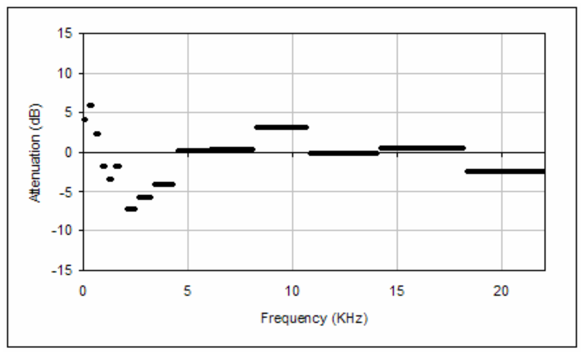

Figure 17: Sub-band representation of $F(f)$.

11.3. Comb Filtering Effect

The oscillations observed in the *F(f)* curve are a result of the comb filtering introduced by the room, which happens when the acoustic signal is added to its own reflection before reaching the microphone, creating periodic notches in the spectrum. If the room is reverberant, the effect of comb filtering is more intense. To evaluate the frequency distortion introduced by the comb filtering in each estimation method, we took the first derivative of *F(f)* and plotted the absolute values.

Figures 18 and 19 show that the estimation of *F(f)* by insertion of pure tones is more vulnerable to the effect of comb filtering than the white noise method, since it creates deeper notches into the spectrum.

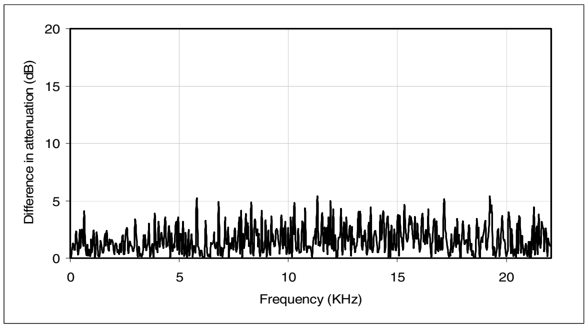

Figure 18: $|dF(f)/df|$ (White noise method).

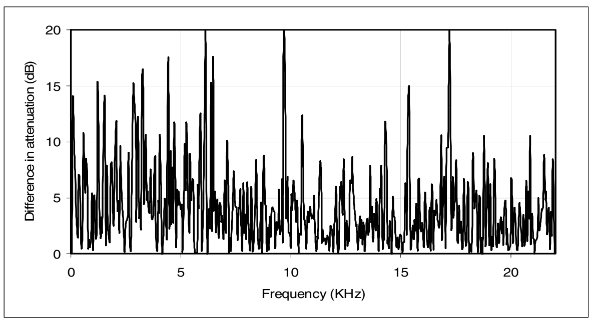

Figure 19: $|dF(f)/df|$ (Individual tones method).

11.4. Standard Deviation Analysis

As described in section 4, a probe tone is inaudible as long as it remains below the masking threshold curve. Most of the times, this requires the tone to be at the same amplitude level as its adjacent frequencies (or even bellow), making it more vulnerable to the unsteadiness of the near spectrum.

To evaluate the stability of a tone under these conditions, we replaced the acoustic signal by white noise, and then measured the standard deviation in the amplitude of the received tone along 1500 frames. We finally obtained the curve in Figure 20 by varying the amplitude ratio between the probe tone and the white noise (TNR).

These results confirm that the tone is more fluctuant for low TNR values, and more stable for high TNR values. The difference, however, is more significant when the TNR varies between -10dB and +10dB approximately. As Figure 21 shows, increasing the ratio above 10dB does not translate into a significant improvement in stability.

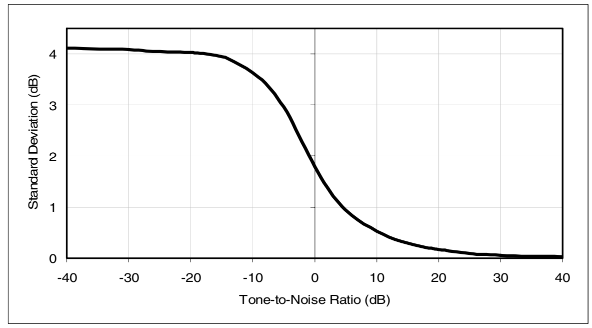

Figure 20: Effect of TNR in standard deviation

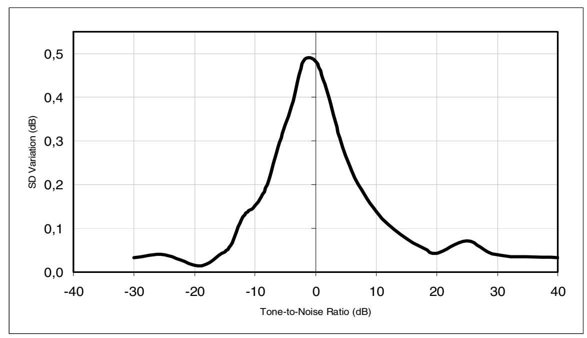

Figure 21: Difference in standard deviation

Through this study, we conclude that the opportunity for insertion of a probe tone is optimal if the masking threshold is at least 10dB above the spectral curve in that spot.

11.5. Acoustic Tests

The complete system, as depicted in Figure 3, was finally applied to a practical situation of equalization in real-time, and the graphical application was used to monitor the equalization progress. The characteristics of the test room were the same as described in [1] and [2]. For the initial tests, the "flat" condition (0 dB for all the sub-bands) was selected for the desired equalization curve. Later, other curves were tested as well. Along the experiments, we verified that the technique of transparent insertion of tones is able to closely estimate the frequency response of the room in a consistent way, while keeping the inserted tones unperceivable to the audition of expert listeners. It was also observed that the convergence of the adaptive filter coefficients become less accurate as the desired curve deviates from the flat.

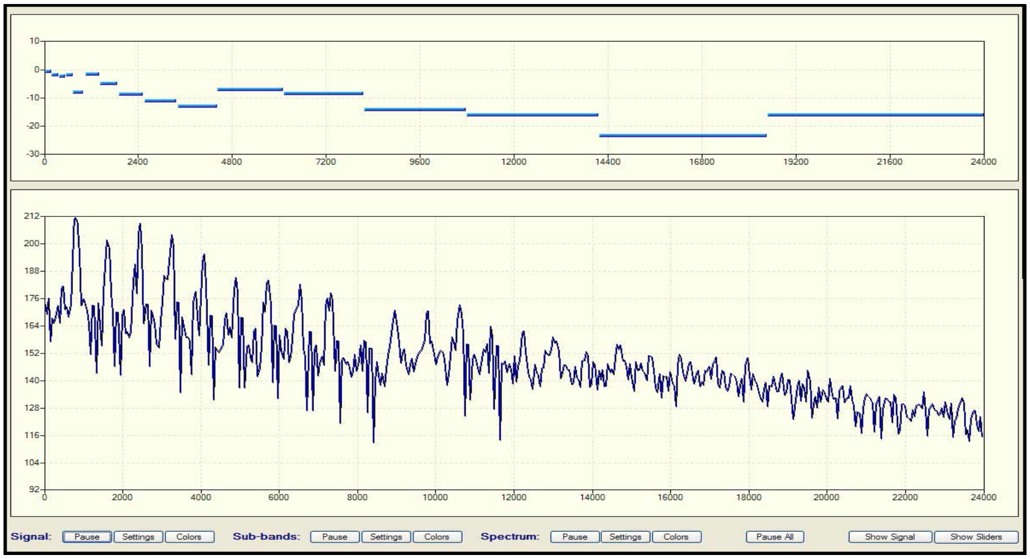

Figure 22: Screen of the graphical interface. In the upper window the current estimation of the room response curve is represented. In the lower window the spectrum of the audio signal, is depicted in real time.

Additionally, the filter exhibits an unsteady behavior in the lower-frequency bands due to the influence of background noise in the received signal. The random nature of the additive noise and the narrowness of the lower bands are the main reasons for this behavior.

Figure 22 presents the estimated frequency response of the room and the spectrum of the audio signal.

12. CONCLUSION

In this paper, it was verified that the developed method based on the transparent insertion of pure sinusoids can be used successfully to estimate in a few seconds the frequency response of the room in a transparent way to the listeners.

From the experiences, we observed that the application of the adaptive filtering in the frequency domain is able to compensate the asymmetries caused by the room acoustics in the audio signal. The technique for the adaptation of the filter coefficients, implemented in this work, exhibits a suitable degree of convergence.

12.1. Suggestions for Future Developments

As a proposal for future works related to this field, we can suggest that, due to the fact that the technique developed for the detection of opportunities for transparent insertion of tones is a first version, it should be improved, performing a more detailed study of the characteristics of the masking threshold curve that reveal appropriate spectral conditions.

The developed system can be applied to a higher number of microphones distributed by the room in order to obtain the frequency response in different points of the auditorium. This represents a main goal in the field of adaptive equalization of room acoustics.

13. REFERENCES

[1] Jorge Leitão, Gabriel Fernandes and Aníbal Ferreira, *Adaptive Room Equalization in the Frequency Domain*, 116th AES Convention Paper, Berlin, Germany, May 2004.

[2] António Leite and Aníbal Ferreira, *An Improved Adaptive Room Equalization in the Frequency Domain*, 118th AES Convention Paper, Barcelona, Spain, May 2005.

[3] Gabriel F. P. Fernandes, *Implementação em DSP de um Sistema Real-Time Analyzer – Aplicação na Igualização Adaptativa*, Master Thesis, School of Engineering of the University of Porto, June 2002.

[4] Sabine Inc., *Real Q2$_{TM}$ Adaptive Real Time Equalizer Operation Guide*, Version 7.5.

[5] Ariel F. Rocha and Aníbal Ferreira, *An Accurate Method of Detection and Cancellation of Multiple Acoustic Feedbacks"*, 118th AES Convention Paper, Barcelona, Spain, May 2005.

[6] Aníbal Ferreira and José Vieira, *An Efficient 20 Band Digital Audio Equalizer*, 98th AES Convention, Frebruary 1995.

[7] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Prentice-Hall International, London, UK, 1999.

[8] Maurice Bellanger, *Digital Processing of Signals*, John Willey & Sons, New York, 1989.

[9] Aníbal J. S. Ferreira, *Spectral Coding and Post-Processing of High Quality Audio*, PhD Thesis, School of Engineering of the University of Porto, May 1998.

[10] A. Dabrowsky – T. Marciniak, *The Computer Engineering Hand Book* - Chapter 27: *Audio Signal Processing*, CRC Press LLc, 2002.

[11] Simon Haykin, *Adaptive Filter Theory*, 4th. Edition, Prentice-Hall International, September 2001.

[12] Texas Instrument, Dallas, *TX, TMS320 C6000 CPU and Instruction Set*, 2000, SPRU189F.