DIRECT METHOD WITH RANDOM OPTIMIZATION FOR LOUDSPEAKER EQUALIZATION USING IIR PARAMETRIC FILTERS

Germán Ramos[*], *José J. López*[**], *Jaime Lloret*[**]
Electronics Department[*], Communications Department[**]
Universidad Politécnica de Valencia
Camino de Vera s/n, 46022 VALENCIA, SPAIN

ABSTRACT

A direct method that uses a random optimization process is proposed for the design of parametric IIR (Infinite Impulse Response) filters for audio equalization. The method uses the loudspeaker response as starting point in order to obtain the desired electro-acoustical target. The algorithm uses a bank of parametric second order filters, known as peak filters in the audio field, to equalize and approximate the filtered response to the desired one. When the recursive process is finished, the parameters that define the filters (frequency, gain and Q) are obtained in correction order of importance; first the ones that provide deep correction in the response. This characteristic allows the implementation of scalable systems with different degrees of complexity and correction.

1. INTRODUCTION

Digital audio filtering is today a low cost solution for improving loudspeakers response and make them more close to an ideal one. It is straightforward to develop a digital filter placed before the amplifier and the loudspeaker, fig.1, in order to perform linear response equalization and correct both, magnitude and phase responses. Additionally, it is possible to decrease the nonlinear distortion produced by a loudspeaker using nonlinear modeling techniques [1] but this will not be the topic of this paper.

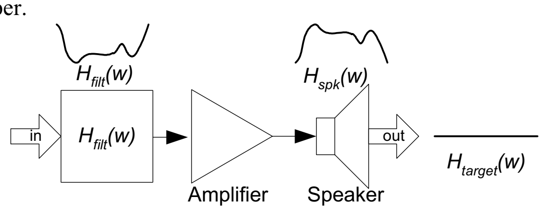

Fig 1. Loudspeaker equalization system

For full-band loudspeakers, the designed digital filter $H_{filt}(\omega)$ must correct the difference between its frequency response (magnitude and phase) $H_{spk}(\omega)$ and the desired electro-acoustical target response $H_{target}(\omega)$.

$$H_{target}(\omega) \approx H_{spk}(\omega) \cdot H_{filt}(\omega) \quad \rightarrow \quad H_{filt}(\omega) \approx \frac{H_{target}(\omega)}{H_{spk}(\omega)}$$

If the target is a flat response, $H_{target}(\omega)=1$, then $H_{filt}(\omega)$ will be a causal approximation of the inverse of the loudspeaker transfer function: $H_{filt}(\omega) \approx H_{spk}(\omega)^{-1}$.

The design of the inverse filter using a finite impulse response (FIR) approximation is straightforward, obtaining always stable solutions and correcting magnitude and phase. However, there are also some drawbacks: the order of the FIR filter, and therefore the computational costs, could be excessive large if low frequencies are needed to be corrected, as well as introduces large delays that could not be permissible in live applications. Besides, pre-echo or pre-hiss could appear been noticeable on transient stimulus. Techniques like warped filters (WFIR) are useful to reduce the length of the filters [2].

On the other hand, IIR filters are more efficient computationally and have no problems for correcting low frequencies as FIR. However, the design of these filters is not so easy as FIR and can become unstable. Another problem of these filters is the quantization effect, which at low frequencies reduces the dynamic range of the system and can also cause oscillations. To prevent this, different filter topologies as parallel filter decomposition, noise-shaping or warped IIR (WIIR) techniques have been employed [2,3,4,5]. At [2], Karjalainen demonstrates that the use of warped FIR or IIR could be computationally efficient and reduce by a factor up to 5 the filter order incrementing only 2.5 to 3 times the computational cost. Methods for the design and optimization of IIR filters could be found at [2,6]. Hawksford and Greenfield [6] perform this approximation first in magnitude with the minimum phase response of the loudspeaker, and second (optional) with excess-phase correction using all-pass filters, representing an useful approach to test the importance of phase equalization. Another interesting approach was done by Rimel and Hawksford using Genetic Algorithms [7] as optimization engine.

The previous commented methods for FIR and IIR filter design should produce good results. But in all of them, a little modification of the target response after design (i.e. to perform some subjective modification) will request a new and complete redesign starting form the beginning. All of them perform the approximations from a coefficient point of view, searching for the sets of a_i and b_i coefficients that minimize the error function. Therefore, there is no direct information about where (frequency) and how (gain and Q) is correcting the response. In the case of FIR filter this information is spread in the designed impulse response $h_{filt}[n]$. For IIR filters (implemented often as second order sections, SOS), the effect are difficult to evaluate individually. Also in the IIR methods, the filters are no designed in order, first the ones that correct more the response.

In order to solve the difficulties before commented, we propose in this paper a method that uses conventional IIR parametric audio filters, so the result of the optimization is easy to evaluate and modify because the optimized parameters are the filter's frequencies, gains and Q. Therefore, there is automatic feedback about what is doing each filter and besides the parameters to be optimized goes down from 5 coefficients by filter to only 3. More additionally advantages will be commented at the conclusions.

2. DESCRIPTION OF THE METHOD

The algorithm proposed to perform the filter optimization can be classified as a direct method. Such type of methods minimizes the error function from a heuristic point of view. Despite these methods are not as mathematically rigorous as conventional ones, from a practical point of view they have advantages as: avoid the use of derivatives or gradient calculus in its optimization; guided initial solutions; straightforward implementation of the iterative loops in the optimization process; stability assured; and possibility to be constrained (the solutions must fulfill restrictions like maximum gain or Q).

As defined previously, $H_{filt}(\omega)$ is the digital filter to be designed that minimizes the difference between the desired response $H_{target}(\omega)$ and the loudspeaker response $H_{spk}(\omega)$. Next equation shows the error $e(\omega)$ which is frequency dependent. Considering a sampled version in the frequency, the mean error could be computed as a summation of squares or absolute values

$$e(\omega) = H_{target}(\omega) - H_{spk}(\omega) \cdot H_{filt}(\omega)$$

$$e_1 = \sum_{i=if}^{ef} e(\omega_i)^2 \qquad \text{or} \qquad e_2 = \sum_{i=if}^{ef} \left| e(\omega_i) \right|$$

where w_i is the frequency index on a discrete version of the Fourier transform, *if* is the initial frequency index to evaluate and *ef* the end frequency index as shown in fig. 2. By this way it will be possible to optimize (equalize) between *if* and *ef*, avoiding extra low or high frequencies or limiting the optimization to a certain frequency band.

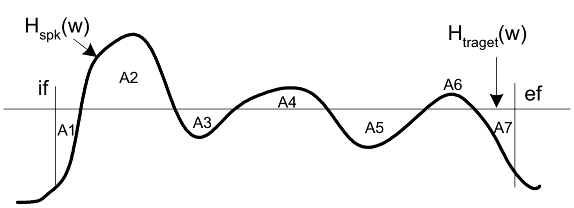

Fig 2. $H_{spk}(\omega)$, $H_{target}(\omega)$ and error areas

A weighting vector $W(\omega)$ could be inserted at the error function to emphasize or attenuate the importance across the frequency axis. The use of $W(\omega)$ allows to avoid the equalization of certain frequency band, or minimize the error more efficient at other one. With this weighting vector, the error will be:

$$e_1 = \sum_{i=if}^{ef} \left(W(\omega_i) \cdot e(\omega_i) \right)^2 \qquad \text{or} \qquad e_2 = \sum_{i=if}^{ef} \left| W(\omega_i) \cdot e(\omega_i) \right|$$

The structure of the correction filter $H_{filt}(\omega)$ used in our algorithm is based on a chain of SOS as shown at figure 3. Each section is a second order IIR filter with minimum-phase and derived from $H(s)$ to $H(z)$ by bilinear transformation. It is also possible to use directly $H(s)$ to design the filters and evaluate the error in s-domain to perform a posterior analog implementation. The normalized analog prototype used in this work is:

$$H_{param}(s) = \frac{s^2 + \dfrac{A}{Q} \cdot s + 1}{s^2 + \dfrac{s}{A \cdot Q} + 1} \qquad \text{Parametric (Peak) Filter}$$

where the parameter A is A=10^(gain_dB/40). It is also possible to use any other type of parametric filters.

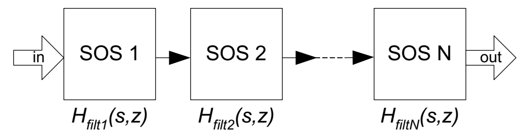

Fig. 3. Implementation of the correction filter

The algorithm is divided in two steps. First a good initial solution is set up for each filter based on error-area criterion, and second, this initial solution is optimized randomly modifying the values of the parameters (frequency, gain, Q) from the initial solution. If the new random filter decreases the error respect to the target, it will be the new initial solution, other ways the initial solution remains the same. The process finish after N iterations or until the error does not minimize any more in M consecutive iterations. In [9] a similar idea is used to optimize analog filters, taking initial guided values and performing a local exploration close to it. The bests new parameters obtained are used as the new initial values.

As said before, the first step is to find a good initial values for the first filter. The main idea is to find the biggest peak or hole in the frequency to compensate it with a parametric filter (each peak/hole on the frequency response of a speaker could be modeled as a parametric filter). This will be done searching the biggest area on $e(\omega)$. To do this, the error between its zero-crossing points is accumulated. The error areas could be seen on figure 2, where A2 is the biggest one, and therefore, the first to compensate. The initial values for $H_{filt1}(\omega)$ will be selected as follows: the frequency *f_i* will be the logarithmic mean between the zero-crossing points of the area A2; the *gain_i* will be the value of the error at *f_i*; and *Q_i* could be defined looking for the –3dB points if they exits, or selected directly between a value from 1.5 to 3 and let the next optimization step will adjust it. The initials values could also be entered by the user avoiding this initial process.

Once the initial values are defined, starts the optimization process using random variations close to it. If the new random filter is better (*e* decreases), then, these new values will be the new initial ones. Thanks to the use of only 3 parameters to define each filter instead of the 5 coefficients, this method is reasonable in term of computing cost. Using variations of the parameters up to 5% and taking 100 to 200 iterations the result will be always useful and really close to the minimum error. Figure 4 shows a sample error surface for a 2 parameter function displaying the evolution of the optimization process. If the new random values are worse (black circle) then will be ignored, if they are better (white circle) then will be the new initials. Due to the first (guided) initial solution is close to the minimum, there is no much risk about falling on a local minimum. Restrictions are imposed to the values of gain (maximum 12 dB to avoid internal saturation on the implementation) and Q (maximum 10 to avoid ringing in the time domain).

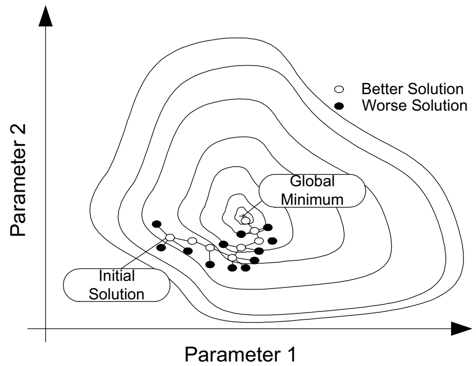

Fig. 4. Error surface and random optimization process

Once the first filter is designed and optimized, it is fixed. The initial values *f_i*, *gain_i*, *Q_i* have been optimized to *f_o*, *gain_o*, *Q_o*. Then starts the same process (search initial conditions and optimization) for the second filter, third, etc. starting from the new filtered $H_{spk}(\omega)\cdot H_{filt1}(\omega)\cdot H_{filt2}(\omega)\cdot\cdot H_{filti}(\omega)$ response. It is convenient that each 5 to 10 filters designed, to reoptimize them starting from their values (no initials are searched again) in order to improve the interaction effects between them, and then to fix them again. Optionally, at the end, the same procedure could be applied to all filters reducing the range of variation to 2% and let iterate more times (about 500-1000), but the improvement is in general insignificant.

Some practical aspects and recommendations to improve the results of the method are commented in the following paragraphs.

$H_{target}(\omega)$ could be generated numerically from $H_{target}(s,z)$, from a table with frequency and amplitude/phase values, or using the response of other loudspeaker in order to simulate its sound with the loudspeaker to be equalized.. Some aspects about $H_{target}(\omega)$ selection are at [2]. Selecting a flat response as a target to obtain an ideal response could be solved theoretically, but this will introduce too much boost at low frequencies trying to extend the band where the loudspeaker does not work, generating distortion and decreasing the dynamic range. Also the speaker can be damaged. A high pass filter with cutoff *fc* near the resonance frequency of the loudspeaker in the box will be better, or to perform the optimization with *if=fc* to exclude frequencies below *fc*. Also, subjective criteria could be introduced like "loudness", "rock", "jazz", or "classic" curves.

To obtain $H_{spk}(\omega)$, maximum length sequences (MLS) can be employed to discard reflections. It is desirable to perform a weighted mean between measures done at 0 degrees and at different angles from the loudspeaker axis (ie. +15º, -15º) to include information about the directivity of the loudspeaker. Then, a resample process should be done on the frequency axis. The linear frequency resolution obtained by the discrete Fourier transform should be resampled to a new logarithmic scale (300-400 points are enough) in order to reduce the frequencies where the error function will be later evaluated, and also following the logarithmic response of the ear and the filters. Finally, a smoothing of the data will be desirable (1/6 to 1/12 octave) to eliminate narrow and inaudible peaks or dips on the response that will also difficult the optimization process because of the increase in the number of error areas.

Additionally, it is convenient to center $H_{spk}(\omega)$ over 0 dB between frequencies *if* and *ef*, other ways, the mean difference between the $H_{target}(\omega)$ level and $H_{spk}(\omega)$ will be corrected during the optimization with the first filter, using the necessary gain and a low Q value. The level of the measured $H_{spk}(\omega)$ is always relative, and there is no need for global gain inside the designed $H_{filt}(\omega)$ that could decrease the dynamic range at the final DSP implementation.

3. EXAMPLE OF APPLICATION

As a sample of application, a non-flat response bass-reflex box with a 6 inch woofer and a half inch tweeter with cut-off at 3kHz is optimized. The frequency response $H_{spk}(\omega)$ of the loudspeaker is shown at figure 5 and also the flat target $H_{target}(\omega)$ and the initial error areas. The optimization will be applied between *if*=80 Hz and *ef*=20kHz in order to avoid excessive boosting of the low frequency band as mentioned before.

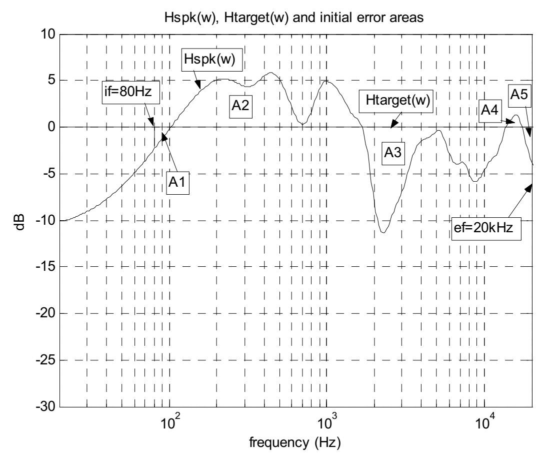

Fig. 5. Loudspeaker response, target function and error areas

The greatest area is A2, then the initial solution is *f_i*=410 Hz, *gain_i*=-5.2dB, and *Q_i*=0.6. After the optimization process the filter moves to *f_o*=310 Hz, *gain_o*=-5.46dB, *Q_o*=0.77. The influence of this first filter $H_{spk}(\omega)\cdot H_{filt1}(\omega)$ is shown at figure 6.

Figure 7 shows the result after 30 parametric filters. At the end of the process, three vectors contain the solutions: frequencies, gains and Qs. As the target is flat, the response of the filter $H_{filt}(\omega)$ (minimum-phase, invertible) is an approximation to $H_{spk}(\omega)^{-1}$ (80 Hz to 20kHz), so $H_{filt}(\omega)^{-1}$ is also a parametric model of $H_{spk}(\omega)$. This last idea to perform models of loudspeakers with minimum-phase parametric, low-pass and high-pass filters was used by Waldman [9] and Schurck [10].

The absolute mean error decreases from 4 dB to less than 0.2 dB. One important aspect is that the filters are designed in correction order, so always the first ones will be the ones that correct more. Only the first 10 filters are really important, the others 20 only decrease the error less than 0.5 dB as seen on figure 8a, where the evolution of the absolute mean error is shown (30 filters and 200 iterations per filter).

In figure 8b there is a detail of the error evolution for the second filter where the two processes of the algorithm are clearly differentiated: the correction obtained by the search of the initial solution, and the effect of the random optimization process where the last 80 iterations do not find better solution.

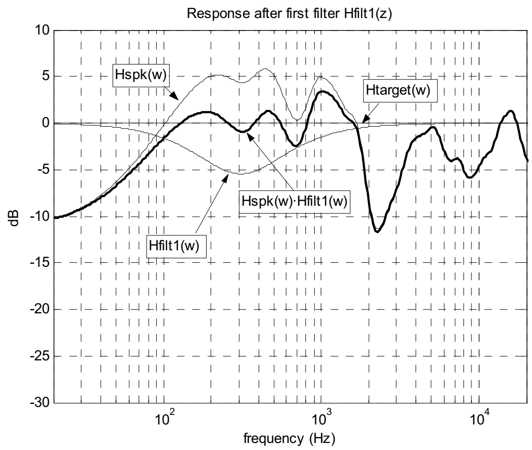

Fig. 6. Effect of the first filter $H_{flt1}(\omega)$

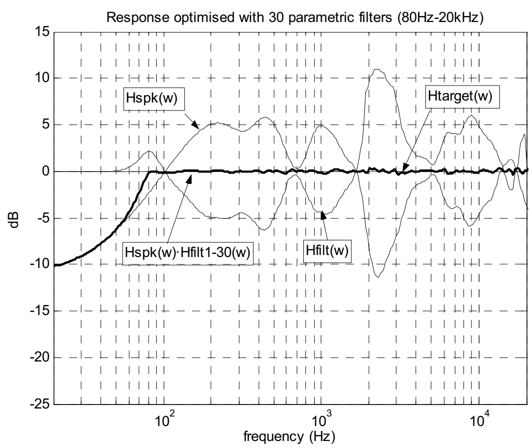

Fig. 7. Equalization with 30 filters. $H_{filt}(\omega)$ response

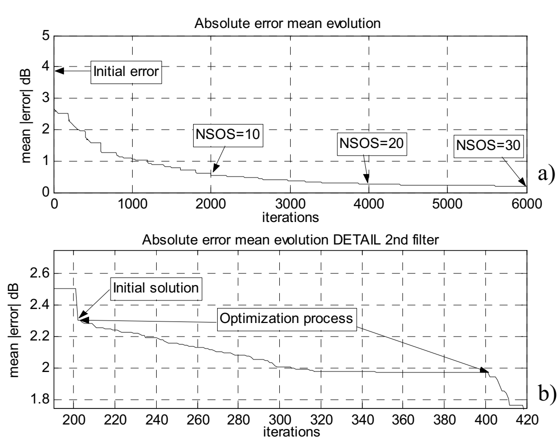

Fig. 8. Error evolution and detail

4. CONCLUSIONS

A novel direct method to equalize audio systems with parametric filters is proposed. The algorithm starts from the system response $H_{spk}(\omega)$ to achieve the desired electro-acoustical response $H_{target}(\omega)$. To perform this correction, a bank of SOS parametric filters $H_{filt}(\omega)$ is designed in order to minimize the difference between them.

The method is always stable and could be constrained, forcing to the new solutions to fulfill restrictions like maximum gain or maximum Q to prevent overflow and ringing.

The equalization is performed in two steps sequentially. First, a good initial solution is searched for each filter, and secondly, an optimization of this initial solution is done using random variations of the parameters close to it.

The use of parametric filters defined by its frequency, gain and Q in the filter bank reduces the number of parameters from the 5 of a generic IIR filter to 3. The result of the optimization is easy to interpret and easy to post-modify (ie. for subjective adjustments) without the need to perform a new total re-design like in other methods.

The design of the filters is in correction order importance. First the ones that perform deep corrections in the response. Thanks to this, if there are computational restrictions to N filters on the final DSP implementation, always the first N will be the best.

This method could also be used to obtain a parametric model of the loudspeaker selecting as target a flat response and inverting the filter (minimum-phase).

Practical equalizations with different loudspeakers have been carried on, always with good results, achieving residual errors lower than 0.5 dB using, in general, only 10 filters. The resulting filters are directly portable to DSP systems, to commercial digital equalizers and also to the analog domain.

5. REFERENCES

[1] U. Horbach, "Design of High-Quality Studio Loudspeaker Using Digital Correction Techniques", *109th Conv. AES 2000*

[2] M. Karjalainen, E. Piirilä, A. Järvinen, and J. Huopaniemi. "Comparison of Loudspeaker Equalization Methods Based on DSP Techniques", *J. AES Vol 47*, 1999

[3] G. Ramos, J. López, "Low Computational Cost Method for Solving Round-off Error at Low Frequency in Digital Audio Equalization", *Forum Acoustic*, Sevilla 2002

[4] J. Dattorro, " The Implementation of Digital Filters for High Fidelity Audio", *J. AES 7th Int. conf.,* 1989

[5] W. Rhonda, " Filter Topologies" *AES UK Conference*, 1992

[6] R. Greenfield, M. Hawksford "Efficient Filter Design for Loudspeaker Equalization",*86th Convention of AES*, Hamburg, preprint 2764 (M-8), March 1989. Revised 1991

[7] A. Rimell, M. Hawksford, " The Application of Genetic Algorithms to Digital Audio Filters", *98th conv. AES 1995*

[8] A. Kaizer, "Numerical Optimization of the CrossOver Filter in a Multiway Loudspeaker", *75th Conv. AES* 1984

[9] W. Waldman, "Simulation and Optimization of Multiway Loudspeaker Systems Using Personal Computer", *J. AES Vol 36, nº 9*, September 1988

[10] P. Schuck, "Design of Optimized Loudspeaker Crossover Networks Using a Personal Computer" *J. AES Vol 34, nº 3*, March 1986