# APPROXIMATION OF REAL IMPULSE RESPONSE USING IIR STRUCTURES

A. Primavera, S. Cecchi, L. Romoli, P. Peretti, F. Piazza

A3Lab - DIBET - Università Politecnica delle Marche Via Brecce Bianche 1, 60131 Ancona Italy a.primavera@univpm.it www.a3lab.dibet.univpm.it

#### ABSTRACT

In this paper, we propose a new approach to the approximation and simulation of a real impulse response. Starting from a preliminary analysis of the mixing time, the impulse response is decomposed in the time domain considering the early and late reflections. Therefore, an IIR structure composed of a cascade of second-order sections and four all-pass filters is employed to synthesize the first part of the impulse response, using a parametric optimization process in the frequency domain. Then, a recursive structure composed of comb and all-pass filters is used to synthesize the late reflections, exploiting a minimization criterion in the cepstral domain. Several results are reported taking into consideration a real impulse response, confirming the validity of the proposed approach.

## 1. INTRODUCTION

In the last 30 years, the problem of simulating real impulse responses (IRs) by means of digital reverberation systems has been deeply investigated [1, 2, 3, 4]. Even if the signal processors technology has been improved rapidly allowing real-time convolution with long impulse responses, many structures for the approximation of the impulse response have been proposed in the literature. A good approximation using shorter filters allows a reduction of the computational cost, which is always remarkable, as well as the number of memory access, taking into consideration an embedded implementation. The most common approach is to simulate the socalled early reflections by FIR filters and the exponentially decaying reverberant part by recursive structure like comb filters and all-pass filters. Therefore, all approaches involve the problem of finding appropriate model parameters of the recursive structure taking into account the IR time domain behaviour. In [5, 6] an adaptive processing is presented in order to define suitable reverberator coefficients for the Moorer's and Schoereder's model with reference to a selected IR. However, other different approaches can be found in the literature taking into consideration different domains of analysis. In [4], a wavelet analysis using logarithmic multirate filter banks is used to define the parametric models. A similar approach is presented in [7] where the Short Time Fourier Transform is applied to decompose the signal, while Steigliz-McBride algorithm is used to evaluate the coefficients of the IIR structure. With regards to the decomposition into early and late reflections, the IR is generally truncated at an arbitrary point in time, considering that the early reflections are usually included within the first 80ms; obviously, this does not imply a good choice for the definition of the early reflections depending on the environment characteristics. The selection of a truncation point that is determined by specific statistical characteristics of an individual IR was explored for the first time in [8], and then exploited in [9].

Differently from the most common approaches, the proposed method is totally based on IIR structures. Starting from a preliminary analysis of the mixing time, the impulse response is decomposed in the time domain considering the early and late reflections. Therefore, an IIR structure composed of a cascade of second-order sections and four all-pass filters, is employed to synthesize the first part of the impulse response, using a parametric optimization process in the frequency domain. Then, a recursive structure composed

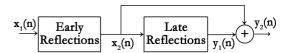


Figure 1: General scheme of the algorithm.

of comb and all-pass filters is used to synthesize the late reflections, exploiting minimization criterion in the cepstral domain.

In Section 2 an accurate description of the entire algorithm is reported, focusing on the automatic parameters setting procedure for early and late reflections (Sections 2.1 and 2.2, respectively) and on the used minimization criterion (Section 2.3). Section 3 reports the obtained experimental results in terms of objective and subjective measures. Finally, conclusions are drawn in Section 4.

## 2. ALGORITHM DESCRIPTION

The first step of the algorithm is the partitioning of the early reflections from the late reverberation, defining the mixing time, i.e., the time elapsed from early to late reflections. As mentioned before, the mixing time is generally fixed at an arbitrary point in time considering that early reflections are included within the first 80 ms. However, since they depend on the environmental characteristics, this choice does not imply a good definition of the mixing time. Therefore, different methods have been developed to estimate this value without knowing the physical characteristics of the environment. The method used in this work is based on gaussianity estimators [9] and it evaluates when the temporal distribution of the signal tends towards a gaussian distribution. Starting from the evaluation of the mixing time, two different procedures are then applied to evaluate the parameters of each structure (i.e., early and late reflections) as shown in Fig. 1: the former is defined in the frequency domain, while the latter in the cepstral domain.

## 2.1 Early reflections approximation

The main characteristic of the proposed approach for the approximation of the first part of the IR resides in the fact that the main structure is based on a second order sections (SOSs) chain of parametric IIR filters. The presented approach derives from a technique applied to the loudspeaker equalization [10]. As shown in Fig. 2, the final approximation filter with *N* sections can be written as follows:

$$H_{\text{er}}(z) = \prod_{p=1}^{N} H_p(z),$$
 (1)

where  $H_p$  is in the form

$$H_p(z) = \frac{a_2 z^{-2} + a_1 z^{-1} + a_0}{b_2 z^{-2} + b_1 z^{-1} + 1},$$
 (2)

with the parameters defined in Table 1.

Let us consider that the filter to be approximated varies around the 0-dB line (Fig. 3), it is possible to approximate the real IR using the proposed structure, compensating the difference between the

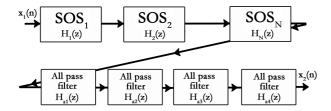


Figure 2: Block diagram of early reflections device.

real frequency response and the reference line. Taking into consideration H, i.e., the Fast Fourier Transform (FFT) of the real IR, an error function is calculated as follows:

$$e(w) = 1 - \frac{H(w)}{H_{\text{er}}(w)}.$$
 (3)

A frequency warping is applied by defining a new non-uniformly spaced frequency axis  $w_{log} = [w_0, w_1, \dots, w_{N-1}]$ . This is done in order to respect the logarithmic behavior of the human ear [4, 10]. In this way a new error vector is defined as follows:

$$e(w_k)_{\mathrm{dB}} = H(w_k)_{\mathrm{dB}} - H_{\mathrm{er}}(w_k)_{\mathrm{dB}},\tag{4}$$

where  $w_k$  represents the k-th element of the new logarithmically spaced frequency axis  $w_{log}$ . Four are the steps of the algorithm:

- 1. Initialization of the algorithm:  $H_{\text{er}}^{(1)}(w_k) = 1$ , p = 1.
- 2. The error vector is calculated as reported in Eq. (4): the biggest area is found by searching for the biggest peak/dip in  $e(w_k)$ . Fig. 3 shows an example of the areas to be searched.
- 3. The initial parameters for the parametric filter  $H_p(z)$  are selected: the central frequency  $f_p$  is the mean between the zerocrossing point of the selected area; the gain  $G_p$  is the value of the error at  $f_p$ ;  $Q_p$  is chosen between 1.5 and 3 for each p filter.
- 4. Then, an optimization of the initial parameters (i.e.,  $f_p$ ,  $G_p$ ,  $Q_p$ ) is performed as suggested in [10], using a heuristic approach. It consists in performing iterative random variations of the parameters close to the initial values computed previously using a new cost function defined as:

$$e_{\log} = \frac{1}{n_f - n_i + 1} \sum_{k=n_i}^{n_f} |e(w_k)_{dB}|$$
 (5)

where  $n_f$  is the final frequency index and  $n_i$  the initial one of the  $w_{log}$  vector.

The last three steps are repeated N-1 times for the other SOSs considering:

$$H_{\text{er}}^{(p+1)}(w_k) = H_p(w_k)H_{\text{er}}^{(p)}(w_k)$$
 (6)

and p = p + 1 for each iteration.

Since the SOSs chain is capable of identifying just the magnitude of the frequency response, four all-pass filters have been added

Table 1: Parametric filter coefficients:  $A_p$  is the square root of the filter linear gain  $G_p$ ,  $Q_p$  is the quality factor of each filter,  $w_0 = \tan(\pi f_c/f_s)$  and  $f_s$  is the sampling frequency.

<i>J</i> C <i>j</i>	/c/ J3/ J3 J.			
	$a_0$	$a_1$	$a_2$	
	$(1 + \frac{A_p}{Q_p}w_0 + w_0^2)$	$(2w_0^2-2)$	$(1 - \frac{A_p}{Q_p}w_0 + w_0^2)$	
	$b_0$	$b_1$	$b_2$	
Г	$(1 + \frac{Q_p}{A_p}w_0 + w_0^2)$	$(2w_0^2-2)$	$(1 - \frac{Q_p}{A_p}w_0 + w_0^2)$	

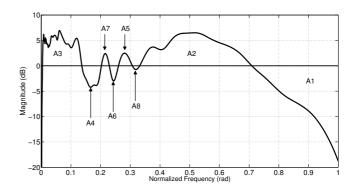


Figure 3: Search of biggest area and design of first SOS.

in order to supply the correct phase of the artificial filters. The frequency response of an all-pass filter is given:

$$H_{a_l}(z) = \frac{-g + z^{-D}}{1 - gz^{-D}} \tag{7}$$

with  $l=1,\cdots,4$  and g,D the parameters to be set. Then, a minimization procedure is applied to the entire structure using the method described in Sec. 2.3, taking into consideration the following loss function:

$$L_1 = \frac{1}{N} \sum_{n=0}^{N-1} (\phi_r(n) - \phi_a(n)), \qquad (8)$$

where  $\phi_r$  is the phase response of the real IR and  $\phi_a$  is the phase response of the structure described in Fig. 2.

## 2.2 Late reflections approximation

The late reflection approximation is realized by an IIR filters network simulating the reverberation tail. In details, it is composed of four all-pass filters in series and eight parallel LBCF (Low-pass feedBack Comb Filters) for each audio channel, as shown in Fig. 4. The comb filter is implemented using a delay-line whose output is low-pass filtered and added to the input. The overall LBCF transfer function is:

$$H_{LBCF}(z) = \frac{1}{1 - g_c L_p(z) z^{-n}} = \frac{1}{1 - g_c \frac{1 - d}{1 - dz^{-1}} z^{-n}},$$
 (9)

where  $g_c$  and d are two different parameters that permit to change the generated reverberation effect. The variable gain  $(g_c)$  sets the reverberation time (T60) while the damp (d) sets the air damping at high frequencies.

The frequency response of an all-pass filter is given by:

$$H_{\rm AP}(z) = \frac{-1 + (1 + g_a)z^{-n}}{1 - g_a z^{-n}}.$$
 (10)

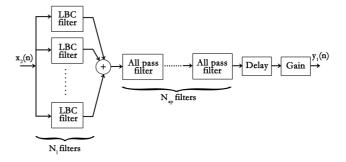


Figure 4: Block diagram of late reflections device.

An all-pass filter is obtained only for  $g_a \cong 0.618$ .

Obviously, the number of comb and all-pass filters can be increased as pointed out in [3] where a feedback delay network is used; this leads to a more precise approximation but also to a more complex structure with a great computational complexity. Therefore, a compromise between computational efficiency and approximation quality has to be achieved: the structure used in the proposed approach (e.g.,  $N_l = 8$  and  $N_{ap} = 4$ ) is sufficient to well identify late reflections, as will be demonstrated in Sec. 3.

Starting from Eq. 9, Eq. 10, and Fig. 4, 17 parameters ( $g_c$ , d,  $g_a$ , one delay, one gain, eight delay lines of the LBCF filters and four delay lines of the all-pass filters) have to be set in order to make the artificial IR as similar as possible to the desired IR. Therefore, an adaptation procedure has to be used for iteratively finding these parameters, taking into consideration the following loss function:

$$L_{2} = \max \left\{ \max \left\{ \sqrt{\sum_{i=1}^{R} \sum_{j=1}^{M} [T_{r}(i,j) - T_{a}(i,j)]^{2}} \right\} \right\}, \quad (11)$$

where  $T_r$  is a matrix which represents the Mel-Frequency Cepstral Coefficients (MFCC) derived from the real IR,  $T_a$  is the MFCC matrix of the approximated IR [11], R is the length of the analysis window, and M is the number of MFCC. The MFCC cepstral coefficients are derived from the IR as suggested in [12]: the FFT of the windowed impulse response is firstly computed to derive the energy spectrum and the result is warped onto the mel scale using a filter bank; then, the MFCC coefficients are calculated as the inverse FFT of the logarithm of the power spectrum. The minimization procedure applied to this loss function is described in Sec. 2.3, taking into consideration M = 26 MFCC, a window of length R = 256 with an overlap of 50%.

#### 2.3 Minimization procedure

The minimization criterion is based on Simultaneous Perturbation Stochastic Approximation (SPSA) [5, 9] as shown in Fig. 5. Taking into account a minimization problem, where  $L(\theta)$  is the loss function and  $\theta$  is the p-dimensional vector of parameters, the steepest descent method can be used to find the set of parameters  $\theta^*$  that minimize  $L(\theta)$ :

$$\left. \frac{\partial L(\theta)}{\partial \theta} \right|_{\theta = \theta^*} = 0. \tag{12}$$

However, when the system is characterized by a large number of parameters (high value of p), it is really hard to determine the gradient value. In this case, SPSA procedure allows to obtain an approximation of the gradient with two measures of the loss function, considering a large number of variables to be optimized. The general recursive form of a classic minimization problem is:

$$\theta_{k+1} = \theta_k - a_k \frac{\partial L(\theta_k)}{\partial \theta_k},$$
(13)

where k is the iteration number. By using the SPSA technique, Eq. 13 can be modified as:

$$\theta_{k+1} = \theta_k - a_k g_k(\theta_k), \tag{14}$$

where  $a_k$  is an arbitrary gain sequence and  $g_k$  is the gradient estimation  $\frac{\partial L(\theta_k)}{\partial \theta_k}$  on the k-th iteration. Considering a simultaneous perturbation, the value of  $g_k$  can be computed as

$$g_k(\theta_k) = \frac{y(\theta_k + c_k \Delta_k) - y(\theta_k - c_k \Delta_k)}{2c_k} \begin{bmatrix} \Delta_{k_1} \\ \vdots \\ \Delta_{k_p} \end{bmatrix}^{-1}, \quad (15)$$

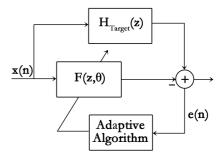


Figure 5: General scheme of the adaption procedure:  $F(z, \theta)$  is the transfer function of whole late reflections structure while  $H_{\text{Target}}$  represents the late reflections of the real impulse response.

where  $y(\cdot)$  denotes a measure of the loss function,  $c_k$  is a gain sequence, and  $\Delta(k)$  is a p-dimensional random vector.

The following three conditions have to be respected ensuring the system stability:

$$E\left\{\left|\Delta_{k_{i}}\right|^{-1}\right\} < \infty \quad \forall i = 1, \dots, p$$
 (16)

$$\lim_{k \to \infty} a_k = 0; \tag{17}$$

$$\lim_{k \to \infty} c_k = 0; \tag{18}$$

where  $\Delta_k$  is a vector with a Bernoulli statistics and value +1 or -1. The sequences  $a_k$  and  $c_k$  can be written as follows:

$$a_k = \frac{a_0}{(A+k+1)}\alpha\tag{19}$$

$$c_k = \frac{c_0}{(k+1)}\gamma,\tag{20}$$

with A,  $\alpha$ ,  $\gamma$ ,  $a_0$  and  $c_0$  arbitrary constants. The following values are used as proposed in [13]: A=1,  $\alpha=0.602$ ,  $\gamma=0.101$ , and  $c_0=0.01$ .  $a_0$  is a vector of constants that varies for each parameter to be found in the optimization process.

## 2.4 Computational complexity

In order to evaluate the performance of the proposed algorithm, the computational complexity in terms of the needed number of multiplications and additions, used for the filtering operations, has been

Table 2: Comparison between the proposed method and the FIR filter approach[14] in terms of computational complexity for filtering operation. N is the number of SOS filters and  $N_a$  is the number of all-pass filters for the early reflection part.  $N_l$  is the number of comb filters and  $N_{ap}$  is the number of all-pass filters in the late reflection part. K is the frame size, L = 2K is the overlap while P is the number of partitions for FIR filtering approach.

	Proposed approach	FIR implementation
Add	$4N + 2N_a + 2N_l + 3N_{ap}$	$\frac{\frac{1}{K}[2L\log_2(L)+2L(P-1)]+}{2\frac{1}{K}(LP)}$
Mul	$5N + 2N_a + 3N_l + 2N_{ap}$	$\frac{\frac{1}{K}[L\log_2(L)]+}{4\frac{1}{K}(LP)}$

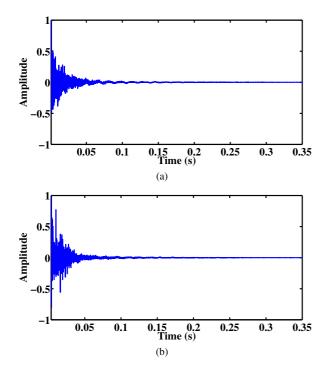


Figure 6: (a) Real impulse response of a room  $(f_s = 48000Hz)$ . (b) Artificial impulse response derived with the proposed approach.

calculated. Table 2 reports the computational load of the proposed technique compared with the FIR filter basic approach using an uniform partitioned overlap and save technique [8, 9, 14]. It is clear that the use of IIR filters improves the performance in terms of required multiplications and memory employed for the memorization of the filters taps. Taking into consideration the numbers used for the experimental results (e.g., N = 20,  $N_a = 4$ ,  $N_l = 8$ ,  $N_{ap} = 4$ , K = 64, L = 128, P = 584) and Table 2, we have that the overall number of multiplications for the proposed approach is 140 while for the FIR approach is 4686. Considering the overall number of additions, we have 116 additions for the proposed approach and 4696 for the FIR approach. If we consider the memory usage which is an important parameter in the case of embedded system implementation, the proposed approach has a lower request (e.g., 10000 samples) than the other method (e.g., 300000 samples). Therefore, the proposed approach has lower computational complexity which can increase taking into consideration a more complicated structure: there is a compromise between the requested accuracy and the maximum computational load. However, we will show through the experimental results that it is possible to have a good approximation of the IR taking into consideration a simple structure with low computational complexity.

## 3. SIMULATION RESULTS

Different tests have been carried out in order to evaluate the approximation quality, in terms of subjective evaluation and objective measures, comparing the approximated signal with the real one. The automatic procedure for the parameters setting has been tested with different real IRs, but for the sake of brevity just the results for a room environment is reported: Fig. 6(a) shows the time behaviour of the real IR with a T60 of 683ms, represented as a FIR filter of 32768 samples. First of all, the mixing time has been estimated using two gaussian estimators (e.g., Kurtosis and Mean Absolute Deviation/Standard Deviation ratio) as described in [9]: for the used real IR we have a mixing time of 30ms. Then, the proposed approach has been applied taking into consideration a structure with N=20 filters and  $N_a=4$  all-pass filters for the early reflection

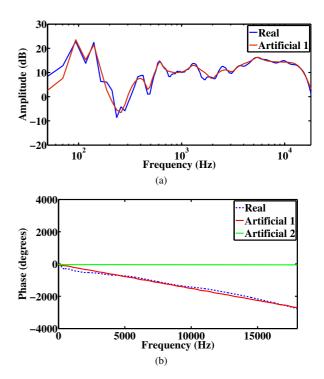


Figure 7: (a) represents the magnitude response of the real IR and of the artificial one and (b) reports the phase of the real IR and of the artificial response without the all-pass filters in the early response approximation (Artificial 2) and with the all-pass filters (Artificial 1).

structure and  $N_l = 8$  comb filters and  $N_{ap} = 4$  all-pass filters for the late reflections part. Regarding the early reflections structure, Fig. 7(a) and Fig. 7(b) show a comparison of the magnitude and phase response of the real and artificial IRs, respectively. The artificial response is quite similar to the original one but the differences are due to the number of filters used for the structure. It is possible to have a more accurate artificial IR but we have tested that this approximation is sufficient to have the same reverberation perception of the real one. Fig. 6(b) shows the time response of the overall artificial IR: it looks very similar to the real IR (Fig. 6(a)), as it was confirmed by the listening tests. Fig. 8(a) shows the Energy Decay Relief (EDR) [9] of the real IR, while Fig. 8(b) represents the EDR of the approximated IR: they result very similar and this is supported by the listening tests. The behaviour of the approximated IR is slightly smoothed since it is calculated considering the smoothed version of the real IR. Moreover, it is worth nothing that as reported in [15] the human ear cannot detect spectrum changes of  $\pm 2dB$  with respect to the level in the neighbouring bands; therefore, the difference between Fig. 8(a) and Fig. 8(b) can be neglected as also confirmed by the listening tests.

Listening tests have been carried out in order to assess the quality of the proposed approximation. The subject is required to compare the stimuli with the original signal and to score it according to a discrete quality scale divided into five intervals (i.e., Bad, Poor, Fair, Good, and Excellent) [16]. The scores are then normalized in the range between 0 and 100, where 0 corresponds to the bottom of the scale (Bad). A grade of 100 is assigned when the stimulus is identified as equal to the reference signal. Three test sessions were carried out taking into consideration a different number of the SOSs (i.e., N), reproducing the sound track through a PC with professional headphone (i.e., AKG). The number of subjects involved was 10 (8 males and 2 females), as suggested in [16], with ages ranging from 21 to 35. The data analysis is performed as an average across subjects of the score associated with each stimulus. Fig. 9 reports the results considering 15, 30, and 50 SOSs, keeping fixed the structure of the late reflections: it is evident that for all the

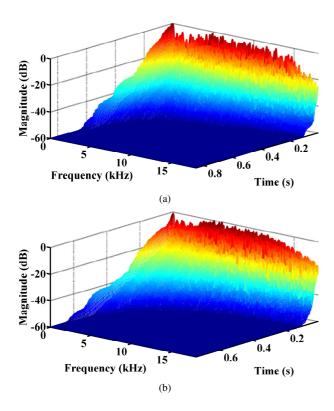


Figure 8: EDR: (a) Real impulse response, (b) approximated impulse response with N=20,  $N_a=4$ ,  $N_l=8$ ,  $N_{ap}=4$ .

tests, the listeners did not find significant differences between the approximated IR and the real one, since all the scores are contained between 85 and 95, i.e., Good and Excellent intervals. Furthermore, increasing the number *N* of SOSs, the mean score increases too; this is due to the fact that the real IR is better approximated. However, good results can be achieved also considering less sections, with a clear computational advantage.

## 4. CONCLUSION

A novel approach for the approximation and simulation of a real impulse response has been presented. After a preliminary analysis of the mixing time, the impulse response is decomposed in the time domain considering the early and late reflections. The early reflections approximation is achieved taking into consideration an IIR structure composed of a cascade of second-order sections, employing a frequency domain parametric optimization process and four all-pass filters to better approximate the phase response of the real IR, using a minimization criterion. Then, the late reflections approximation is based on a recursive structure composed of comb and all-pass filters, exploiting a minimization criterion in the cep-

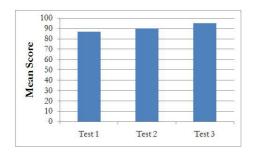


Figure 9: Listening tests: mean score obtained over each test.

stral domain. Several results are reported taking into consideration objective and subjective tests for real impulse responses. The results obtained in the objective tests in terms of energy decay relief have been confirmed by the subjective listening tests: the listeners did not find differences between the approximated IR and the real one. Future works will be oriented to the real time implementation of the approach considering an embedded platform.

## REFERENCES

- [1] M.R. Schroeder, "Natural Sounding Artificial Reverberation," J. Audio Eng. Soc., vol. 10, pp. 219–223, 1962.
- [2] J.A. Moorer, "About This Reverberation Business," *Computer Music Journal*, vol. 3, no. 2, pp. 13–28, 1979.
- [3] J. Jot, "Digital Delay Networks for designing artificial reverberators," in *Proc. 90th Audio Engineering Society Convention* (AES'91), Paris, Feb 1991.
- [4] M. Schoenle, N. Fliege, and U. Zoelzer, "Parametric approximation of room impulse responses by multirate systems," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'93)*, Minneapolis, MN, apr. 1993, vol. 1, pp. 153–156 vol.1.
- [5] G. Constantini and A. Uncini, "Real-Time Room Acoustic Response Simulation by an IIR Adaptive Filter," *Electronics Letters*, vol. 39, pp. 330–332, 2003.
- [6] A. Uncini, G. Costantini, and G. Barbasso, "Adaptive Filter for Real-Time Room Acoustic Response Simulation," in Proc. Second International Conference "Understanding and Creating Music", Caserta, Nov 2002.
- [7] J.C. Sarris and G.E. Cambourakis, "Time frequency analysis and parametric approximation of room impulse responses," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'03)*, Hong Kong, apr. 2003, vol. 6, pp. 297–300.
- [8] R. Stewart and D. Murphy, "A Hybrid Artificial Reverberation Algorithm," in *Proc. 122nd Audio Engineering Society Convention (AES'07)*, Vienna, Austria, May 2007.
- [9] A. Primavera, L. Palestini, S. Cecchi, F. Piazza, and M. Moschetti, "A Hybrid Approach for Real-Time Room Acoustic Response Simulation," in *Proc. 128th Audio En*gineering Society Convention (AES'10), London, UK, May 2010.
- [10] G. Ramos and J.J. Lopez, "Filter Design Method for Loud-speaker Equalization Based on IIR Parametric Filters," *J. Audio Eng. Soc.*, vol. 54, no. 12, pp. 1162–1178, Dec. 2006.
- [11] S. Heise, M. Hlatky, and J. Loviscach, "Automatic Adjustment of Off-the-Shelf Reverberation Effects," in *Proc. 126th Audio Engineering Society Convention (AES'09)*, Munich, Germany, May 2009.
- [12] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Trans. Audio, Speech and Lang. Process.*, vol. 28, no. 4, pp. 357 366, aug. 1980.
- [13] Spall J.C., "Implementation of the simultaneous perturbation algorithm for stochastic optimization," *IEEE Trans. Audio, Speech and Lang. Process.*, vol. 34, pp. 817 823, 1998.
- [14] W. G. Gardner, "Efficient Convolution without Input-Output Delay," J. Audio Eng. Soc., vol. 43, no. 3, pp. 127–136, Mar. 1995
- [15] J. Lopez and G. Ramos, "Low Order IIR Parametric Loudspeaker Equalization, A Psychoacoustic Approach," in Proc. 118th Audio Engineering Society Convention (AES'05), Barcelona, Spain, May 2005.
- [16] ITU-R BS. 1534, "Method for subjective listening tests of intermediate audio quality," 2001.