A subband implementation of a multichannel and multiple position adaptive room response equalizer

A Subband Implementation of a Multichannel and Multiple Position Adaptive Room Response Equalizer

S. Cecchi[a,1,*], A. Terenzi[a,1], V. Bruschi[a,1], A. Carini[b,1], S. Orcioni[a,1]

[a] *Department of Information Engineering, Università Politecnica delle Marche, Via Brecce Bianche, 60131 Ancona, Italy.*
[b] *Department of Engineering and Architecture, Università degli Studi di Trieste, Via Alfonso Valerio, 10 34127 Trieste, Italy*

Abstract

This paper deals with a subband implementation of an adaptive multichannel and multiple position room response equalization system capable of improving the listening experience in a real environment. The main focus of the approach is the introduction of a subband identification procedure capable of estimating in real-time the impulse responses of a system with a great accuracy and a quick convergence. These aspects are very important also for the equalization procedure to obtain a good accuracy and quick development of the equalization filters as function of the environment variations. Several experiments have been performed in a real scenario to test the performance of the algorithm and to show the beneficial effects of the subband identification on the equalization procedure.

Keywords: adaptive equalization, room response equalizer, channel decorrelation, subband analysis, impulse response identification

*Corresponding author
Email addresses: s.cecchi@univpm.it (S. Cecchi), a.terenzi@univpm.it (A. Terenzi), v.bruschi@univpm.it (V. Bruschi), acarini@units.it (A. Carini), s.orcioni@univpm.it (S. Orcioni)
[1] Phone/Fax Number: +390712204453

1. Introduction

The perception of sound reproduction is modified by the listening environment, which can introduce undesired artifacts (e.g., standing waves, frequency band extensions, nonlinearities) to the original sound. The room response equalization (RRE) allows us to improve the quality of the sound reproduction in real environments (such as cinema, home, theater, car). A wide variety of RRE methods have been studied over the last 40 years with the aim of improving the listening experience by compensating the room transfer function (RTF) from the loudspeaker system to the listener [1]. Different RRE approaches have been proposed in the literature: many of them consider fixed equalizers, designed a priori before the filtering operation [1, 2, 3, 4, 5]. However, the room can be considered as a time-varying environment (a "weakly non-stationary" system as defined in [3]), so the room response varies with position [2] and with time [3]. In audio processing, this is a very challenging task since many issues have to be examined [6]. Therefore, adaptive solutions should be adopted in order to track and correct slow variations in the room response due to temperature changes or movement of people or other obstacles. In particular, two types of room response (RR) adaptive equalizers can be found: single-point (single-input/single-output SISO, multiple-input/single-output MISO) [4] and multi-point (single-input/multiple-output SIMO, multiple-input/multiple-output MIMO) [1, 7] room equalizers. The single-point room equalization filter is obtained from a measurement of the room response in a single position, so the room equalization is achieved only in a reduced zone around the measured point. Differently from single position case, the multi-point room equalizer is designed starting from several measurements of the RTF at different locations in order to expand the area of equalization. While there is a huge amount of approaches to RRE [8], in what folllows we review only the most relevant ones related to the proposed solution.

A first adaptive RRE system was proposed by Elliot et al. in [9], where the equalizer is designed by adaptively minimizing the sum of the squared errors between the equalized responses and the delayed input signal. A biased

adaptive algorithm has been lately proposed in [10], in order to increase the convergence speed and improve the robustness of the adaptive identification algorithm in the case of a low signal to noise ratio. Unfortunately, this method requires an a-priori estimation of the impulse responses and it is not easy to obtain in many practical applications. In [6, 11], a technique suitable for RRE in current commercial hi-fi audio products is discussed. This method is derived from the technique presented in [12], where a single position RRE approach in the frequency domain is considered. The system of [12] allows the equalization by dividing the loudspeaker signal and the microphone signal in subbands and updating the filter weights in these subbands. This approach is interesting because it combines the robustness towards peaks and notches of the room responses, the simplicity of the implementation, and the capability of tracking RR variations. In [13, 14], the method of [12] was refined by first developing a multiple position solution and then by considering a warped frequency domain in order to expand equalization to the lower frequencies. However, the adaptive room response equalization approaches proposed in [12, 13, 14] investigate the equalization of a single sound source, i.e., of a single audio channel. Contrariwise, the technique in [11] presents a multichannel solution that contemplates also the non-uniqueness problem. This problem happens in the stereophonic acoustic echo cancellation [15] because of the ill-conditioned covariance matrix. To solve the non-uniqueness problem, a method to minimize the interchannel coherence is usually applied [16]. For example, in [17], a simple method that uses positive and negative half-wave rectifiers is employed to reduce the coherence between the channels. Unfortunately, many of the techniques used to weaken the channel cross-correlation often introduce considerable distortions, which are unacceptable in the case of high quality sound reproduction systems [18]. Therefore, an appropriate technique, capable of decorrelating the input signals without modifying the audio quality, must be examined. In [11], an adaptive multichannel and multiple position RRE system has been briefly introduced. In [6] the same method has been completely detailed and expanded in order to provide a real-time implementation on commercial hi-fi products

suitable for a stereo reproduction. The approach of [6] allows us to estimate the room responses with great accuracy by reducing the inter-channel coherence through a technique that produces only a modest degradation of the sound quality. The RR equalizer of [6] is designed starting from a prototype that is derived in the warped frequency domain to increase the equalization in the low frequency spectrum and to simultaneously reduce the computational cost of the filters design. Unlike other techniques presented in the literature, the system discussed in [6] is able to solve the problem of multichannel and multiple position room response equalization for time varying environments by applying a simple and computationally efficient solution and without introducing artefacts. In [19], an adaptive multichannel and multiple position room response equalization system is reported. In particular, an adaptive and accurate estimation of the room responses is provided introducing a normalized least mean square optimization approach with a variable step-size, and taking advantage of a multichannel interchannel coherence reduction technique based on the missing fundamental phenomenon. Then, the equalizer is designed with the introduction of a novel prototype function derived from the combination of quasi-anechoic impulse responses with the impulse responses recorded in the real environment to be equalized.

In this paper, an evolution of the technique proposed in [6, 19] is reported, improving the identification procedure. In particular, a multichannel subband adaptive structure has been adopted for the system identification to obtain a better resolution of the impulse response (IR) at low frequencies and thus a better resolution in the equalization procedure. The subband structure allows us also to achieve a quicker convergence of the algorithm. Starting from the subband filter bank proposed in [20], an efficient structure has been developed and combined with a multi-point equalization approach which at this time represents in the state of the art a useful technique to extend the effect of the equalization procedure [8]. Results have been reported taking into consideration a real scenario.

The rest of the paper is organized as follows. Section 2 describes the proposed

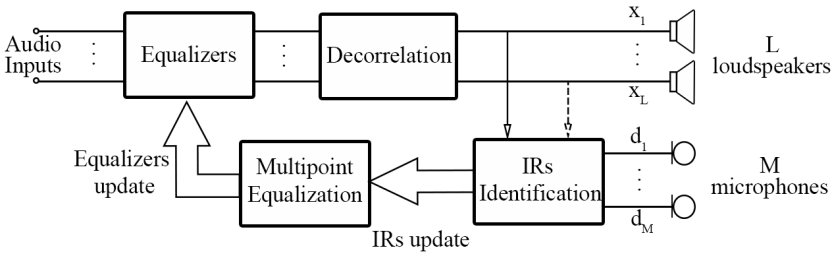


Figure 1: Block diagram of the proposed solution.

algorithm for RRE, taking into consideration the different parts of the algorithm, i.e., the input signals decorrelation, the subband room response estimation, and the equalizer design. Section 3 provides different experimental results on the proposed subband RR equalizer. Eventually, concluding remarks are given in Section 4.

2. Proposed Approach

The proposed system is designed for providing an iterative estimation of the room impulse responses (RIRs) and, at the same time, a multipoint equalization. The system repeats the approach presented in [19] with a variation in the room response identification that is performed considering a multirate subband approach in order to reduce the computational cost and to improve the convergence speed.

The block diagram of the proposed approach is shown in Figure 1. In the case of more than one loudspeaker, convergence problems can occur due to the channels correlation. These problems are very common in multichannel acoustic echo cancellation because of the possible errors in the identification of the acoustic paths [21, 22]. In this context, a method to reduce the inter-channel coherence must be exploited, as described in [23]. Furthermore, the identification of room responses is achieved by a subband adaptive filtering using the structure presented in [20]. In this way, $L \cdot M$ room responses between the L loudspeakers and the M microphones are estimated and then exploited

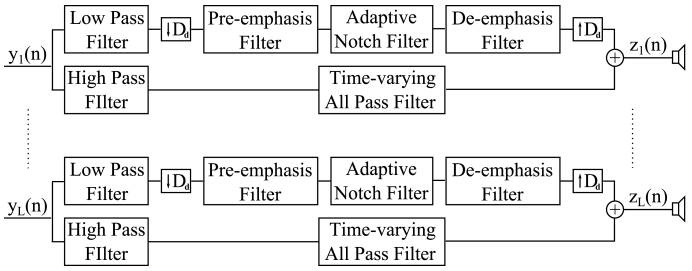


Figure 2: Multichannel decorrelation procedure.

for the equalizer design. These three steps of the algorithm, i.e., input signals decorrelation, subband room response identification, and multipoint equalizer design are described in the following subsections.

2.1. Input decorrelation

The multichannel input signals decorrelation is obtained by the psychoacoustic criteria of the missing fundamental [24], as reported in [25]. This psychoacoustic phenomenon is associated with the human capability of perceiving the fundamental frequency although it is not actually present in the signal. Moreover, second-order time-varying all-pass filters are included in this approach so as to expand the solution to the entire frequency spectrum [25]. The scheme of the algorithm is reported in Figure 2, in which each input channel $y_p(n)$ of the decorrelation block, being $p = 1, \cdots, L$, is divided into two subbands [19].

Hence, after a decimation operation of D_{d}, an adaptive notch filter $H_p(z, n)$ is applied in the lower frequencies while a second-order time-varying all-pass filter $F_p(z, n)$ is applied in the high-frequency spectrum causing an alteration of the signal phase. This method allows us to accurately identify each processed channel $z_p(n)$.

In the low-frequency band, the notch filters are created by L second-order lattice structures in order to remove the L fundamental frequencies. The p-th notch filter is described as follows:

$$H_p(z, n) = \frac{1 + 2k_p(n)z^{-1} + z^{-2}}{1 + k_p(n)[1 + \alpha_p(n)]z^{-1} + \alpha_p(n)z^{-2}}, \tag{1}$$

where $p = 1, \cdots, L$ represents the channel index, $k_p(n)$ is the adaptive coefficient connected to the tracked frequency $f_p(n)$ and $\alpha_p(n)$ is the pole-zero contraction factor that controls the bandwidth of the filter [26]. The cut-off frequency of the notch filter controlled by $\alpha_p(n)$ must change in time so as to identify and delete the fundamental frequency. Thus, decorrelation is guaranteed in the whole low-frequency band providing results similar to a set of time-varying all-pass filters [27].

The contraction factor $\alpha_p(n)$ is related to each channel and the time-varying vector $\boldsymbol{\alpha}(n) = [\alpha_1(n), \cdots, \alpha_L(n)]$ provides disparity among channels even if the fundamental frequency is the same. In particular, a right circular shift of one sample $s(\cdot, 1)$ is applied to the vector $\boldsymbol{\alpha}(n)$ every Q samples [25], as explained in the following equation:

$$\boldsymbol{\alpha}(n) = \begin{cases} s[\boldsymbol{\alpha}(n-1), 1] & \text{if } \left(n - Q\left\lfloor \frac{n}{Q} \right\rfloor\right) = 0 \\ \boldsymbol{\alpha}_p(n-1) & \text{otherwise.} \end{cases} \tag{2}$$

The adaptive coefficient $k_p(n)$ is included in the interval $(-1, 1)$ to avoid the divergence of the filter and represented by the following sigmoid function:

$$k_p(n) = \frac{2}{1 + e^{-\gamma_p(n)}} - 1, \tag{3}$$

being $\gamma_p(n) \in R$. The tracking of the fundamental frequency is obtained by finding $\gamma_p(n)$ that minimize the output energy of the filter in (1), as described in [28]. In this way, the filter of (1) is completely determined and it is capable of removing the fundamental frequency. It is possible to derive this fundamental frequency $f_p(n)$ with the knowledge of the sampling frequency f_s and of the down-sampling factor D_d, considering the following equation:

$$f_p(n) = \frac{f_\mathrm{s}}{D_\mathrm{d}} \cdot \frac{1}{2\pi} \cos^{-1}[-k_p(n)]. \tag{4}$$

Moreover, the low-frequency band could contain also some harmonics, so a pre-emphasis filter $H_\mathrm{pre}(z)$ is used in order to improve the method in the low frequencies and a de-emphasis filter $H_\mathrm{de}(z)$ is applied to annul the effect of the

first. These filters are described as follows:

$$H_{\mathrm{pre}}(z) = \frac{1}{1-\nu z^{-1}} \tag{5}$$

$$H_{\mathrm{de}}(z) = 1-\nu z^{-1}, \tag{6}$$

being $0 < \nu < 1$.

Considering the high-frequency range, the phase of the input channels is changed by the application in each channel of L second-order time-varying all-pass filters. The transfer function of the p-th all-pass filter is described by the following equation [29]:

$$F_p(z,n) = \frac{k_p^2(n) - 2k_p(n)z^{-1} + z^{-2}}{1 - 2k_p(n)z^{-1} + k_p^2(n)z^{-2}}, \tag{7}$$

so, it is identified by a pole with a multiplicity of 2 connected to the coefficient $k_p(n)$ of Eq. (3). This characterization of the all-pass filter allows us to maintain the spatial perception of the speech [27] because the restriction $|k_p(n)| < 1$ guarantees stability and causality of the filter and ensures that the inter-aural time delay difference between the two ears is lower than the well-known "just noticeable inter-aural delay" [30]. As described in [25], the alteration in sound direction is negligible as the maximum variation in the time of arrival is about $40\,\mu$s for all frequencies.

2.2. Subband Room Response Identification

The architecture used for the room response identification is based on a subband adaptive filtering structure with critical sampling [20]. Considering D the number of subbands and $h_{\mathrm{p}}(n)$ the impulse response of a prototype filter of order N_h that guarantees perfect reconstruction when used in a cosine modulated analysis-synthesis system [20], the analysis and synthesis filters are obtained as follows,

$$g_k(n) = 2h_{\mathrm{p}}(n)\cos\left[\frac{\pi}{D}(k+0.5)\Big(n-\frac{N_h}{2}\Big) + \vartheta_k\right], \tag{8}$$

$$f_k(n) = 2h_{\mathrm{p}}(n)\cos\left[\frac{\pi}{D}(k+0.5)\Big(n-\frac{N_h}{2}\Big) - \vartheta_k\right], \tag{9}$$

where $\vartheta_k = (-1)^k \frac{\pi}{4}$, for $0 \leq k \leq D-1$ and $0 \leq n \leq N_h$. Figure 3 shows the subband structure considering $L = 2$ loudspeakers and one of the microphones. In Figure 3, **GG** is the double analysis filter-bank, which is derived from the uniform cosine modulated filter-bank **G**, and **F** is the corresponding synthesis filter-bank. The double analysis filter-bank **GG** is composed by the D filters $G_k(z)G_k(z)$ for $k = 0, \ldots, D-1$, and the $D-1$ filters $G_k(z)G_{k+1}(z)$ for $k = 0, \ldots, D-2$, with the following impulse responses [31]:

$$g_k(n) * g_k(n) \approx 2[h_{\mathrm{p}}(n) * h_{\mathrm{p}}(n)] \cos\left[\frac{\pi}{D}(k+0.5)\Big(n - \frac{N_h}{2}\Big) + 2\vartheta_k\right], \tag{10}$$

$$g_k(n) * g_{k+1}(n) \approx 2q_0(n) \cos\left[\frac{\pi}{D}(k+0.5)\Big(n - \frac{N_h}{2}\Big)\right], \tag{11}$$

where

$$q_0(n) = \left[h_{\mathrm{p}}(n)e^{j\frac{\pi}{2D}n}\right] * \left[h_{\mathrm{p}}(n)e^{-j\frac{\pi}{2D}n}\right]. \tag{12}$$

As a consequence, considering an input signal $x_l(n)$ with $l = 1, ..., L$ the the loudspeaker index, the outputs of the filter-bank after the downsampling operation are $2D-1$ signals derived as follows in Z-domain:

$$X_{l,k,k}(z) = X_l(z^{\frac{1}{D}})G_k(z^{\frac{1}{D}})G_k(z^{\frac{1}{D}}) \tag{13}$$

for $k = 1, \ldots, D-1$, and

$$X_{l,k,k+1}(z) = X_l(z^{\frac{1}{D}})G_k(z^{\frac{1}{D}})G_{k+1}(z^{\frac{1}{D}}) \tag{14}$$

for $k = 1, \ldots, D-2$.

These signals constitute the inputs of a bank of adaptive filters. Let us denote with ξ the downsampled time index, with $\mathbf{x}_{l,k,k}(\xi)$ and $\mathbf{x}_{l,k,k+1}(\xi)$ the input vectors formed by the samples of signals in (13) and (14), respectively, and with $\mathbf{w}_{m,l,k}(\xi)$ the vectors collecting the coefficients of the adaptive filters from input l to microphone m. Each adaptive sub-filter should have at least $K = (\frac{N+N_h+2}{D} + 1)$ coefficients [20] with a uniform frequency bandwidth of π/D and a center frequency of $\pi/(2D)$. The sub-filters are obtained through the minimization of the sum of the instantaneous subband squared-errors, represented by:

$$J_m(\xi) = \sum_{k=0}^{D-1} e_{m,k}^2(\xi), \tag{15}$$

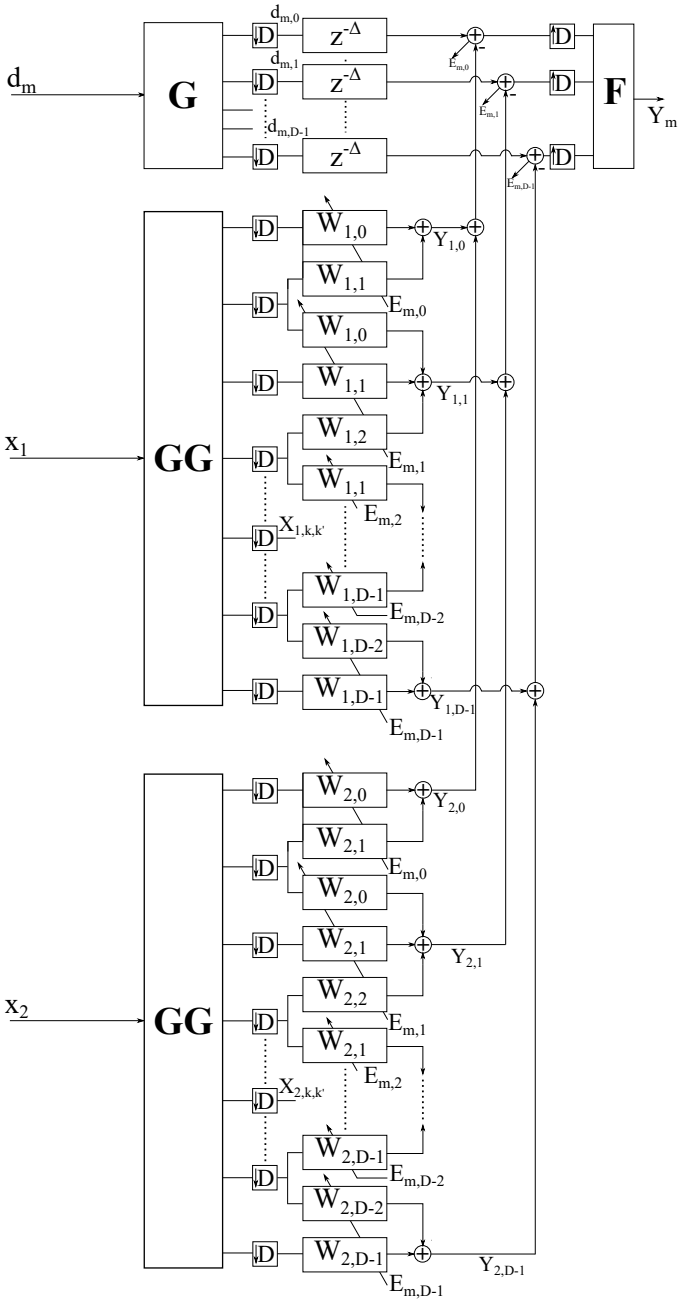


Figure 3: Subband IRs identification procedure with $L = 2$ and $M = 1$.

where $m = 1, ..., M$ and M is the number of microphones. The error signals are obtained by the following relation

$$e_{m,k}(\xi) = d_{m,k}(\xi - \Delta) - \sum_{l=1}^{L} y_{m,l,k}(\xi), \tag{16}$$

where $d(\cdot)$ is the desired signal, $\Delta = \frac{N_h+1}{D}$ is the delay introduced by the filter bank $\mathbf{G}$, and

$$y_{m,l,k}(\xi) = \mathbf{x}^T_{l,k,k}(\xi)\mathbf{w}_{m,l,k}(\xi) + \mathbf{x}^T_{l,k-1,k}(\xi)\mathbf{w}_{l,m,k-1}(\xi) + \mathbf{x}^T_{l,k,k+1}(\xi)\mathbf{w}_{l,m,k+1}(\xi). \tag{17}$$

The filters $\mathbf{w}_{m,l,k}$ are adapted according to the following formula:

$$\begin{aligned}\mathbf{w}_{m,l,k}(\xi+1) = \mathbf{w}_{m,l,k}(\xi) + \mu_{l,k}(\xi)[&\mathbf{x}_{l,k,k}(\xi)e_{m,k}(\xi) \\ &+ \mathbf{x}_{l,k-1,k}(\xi)e_{m,k-1}(\xi) + \mathbf{x}_{l,k,k+1}(\xi)e_{m,k+1}(\xi)].\end{aligned} \tag{18}$$

The step sizes $\mu_{l,k}(\xi)$ are normalized by the sum of instantaneous powers of the signals involved in the coefficients adaptation, i.e.,

$$\mu_{l,k}(\xi) = \frac{\mu}{\delta + P_{l,k,k}(\xi) + P_{l,k-1,k}(\xi) + P_{l,k,k+1}(\xi)}, \tag{19}$$

where δ is a small positive constant to avoid division by zero and the power is estimated as follows:

$$P_{l,k,k'}(\xi+1) = \beta P_{l,k,k'}(\xi) + (1-\beta)X^2_{l,k,k'}(\xi), \tag{20}$$

with β a constant in the range $(0, 1)$. Finally, the reconstructed frequency response between the l-th loudspeaker and the m-th microphone $\widehat{H}_{m,l}(z)$ is computed as the sum of the interpolated sub-filters $W_{m,l,k}(z^D)$, filtered by the synthesis filter-bank $\mathbf{F}$, i.e.,

$$\widehat{H}_{m,l}(z) = \sum_{k=0}^{D-1} W_{m,l,k}(z^D)F_k(z). \tag{21}$$

The computational complexity of the subband identification algorithm is lower than a full band LMS approach as reported in [20]. In particular, the overall number of multiplications per input sample required for the filtering and adaptation of the sub-filters $W_{m,l,k}(z^D)$ is computed as follows:

$$\frac{2(3D-2)N_p}{D^2} + \frac{2(3D-2)N_h}{D^2}, \tag{22}$$

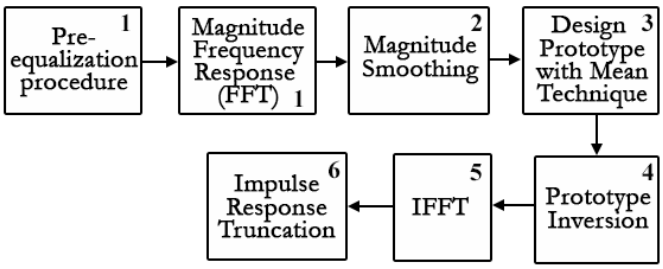


Figure 4: Room response equalization procedure.

where the first term corresponds to the filtering operations and the second term corresponds to the adaptation procedure with N_p the length of the full band system. For high-order adaptive filters, the dominant term in the above expression is $6N_p/D$, which is about $D/3$ times smaller than the number of multiplications required by the fullband LMS algorithm ($2N_p$) as reported in [20].

2.3. Multipoint Equalizer Design

The proposed identification procedure has been tested with a multipoint equalization technique [5]. This approach has been considered due to its capability of enlarging the listening sweet spot taking into consideration different microphones positions. The considered multipoint technique takes into consideration a quasi-anechoic approach [32], thus producing also a general equalization of the used loudspeakers. The prototype function is derived from the combination of quasi-anechoic IRs, derived from a gated version (up to the first reflection) of the responses, with the IRs recorded in the real environment.

Figure 4 shows the equalization approach used in the presented work that follows these steps:

1. Starting from a set of impulse responses derived in the zone to-be-equalized, a pre-processing is considered exploiting the quasi-anechoic IR spectrum for frequency greater than a certain transition frequency and the original (ungated) IR spectrum below the same transition frequency. For simplicity of representation in this sub-section we work in the DFT domain and the operation is performed

by applying the following equation to each RIR:

$$H_{m,l}(\kappa) = \widehat{H}_{m,l}(\kappa) \cdot w_{\mathrm{lf}}(\kappa) + \widetilde{H}_{m,l}(\kappa) \cdot w_{\mathrm{hf}}(\kappa), \tag{23}$$

where $\kappa = 0, \cdots, K-1$ indicates the frequency beam, $\widehat{H}_{m,l}(\kappa)$ is the frequency response of the original estimated RIR, $\widetilde{H}_{m,l}(\kappa)$ is the frequency response of the gated RIR, $w_{\mathrm{lf}}(\kappa)$ and $w_{\mathrm{hf}}(\kappa)$ are the half Hann windows used for selecting the low-pass and high-pass frequency bands, respectively. The linear combination in Eq. (23) is used to equalize only the direct sound in mid-high frequency range, which determines localization and most of timbre perception, while full equalization is applied in the modal frequency range [11].

2. Then, starting from the magnitude spectra of the IRs, a smoothing operation is applied, simulating the poorer frequency resolution at higher frequencies of the human auditory system. The approach of [33] has been used to obtain a non uniform frequency magnitude spectrum smoothing of the frequency response $H_{m,k}(\kappa)$ for all m and l. In this way, a broader equalized zone is achieved exploiting a less precise equalization at higher frequencies resulting from the non-uniform resolution, which decreases increasing the frequency.

3. At this point, a representative response of the considered acoustic environment is derived taking into account all smoothed IRs. The prototype frequency response is obtained using an arithmetic mean of the zero-phase smoothed frequency responses, as follows:

$$H_{\mathrm{pr}_l}(\kappa) = \frac{1}{M} \sum_{m=1}^{M} H_{\mathrm{sm}_{m,l}}(\kappa), \tag{24}$$

with $\kappa = 0, \cdots, K-1$ and $H_{\mathrm{sm}_{m,l}}(\kappa)$ the smoothed frequency response from loudspeaker l to microphone m.

4. Finally, a frequency domain inverse filter is obtained through the use of a frequency deconvolution with regularization technique [34] applied to the prototype as follows:

$$H_{\mathrm{inv}_l}(\kappa) = \frac{H^*_{\mathrm{pr}_l}(\kappa)}{\left|H_{\mathrm{pr}_l}(\kappa)\right|^2 + \beta}, \tag{25}$$

being $\kappa = 0, \cdots, K-1$, $H^*_{\mathrm{pr}_l}(\kappa)$ the complex conjugate of $H_{\mathrm{pr}_l}(\kappa)$, and β the

regularization factor, which allows us to avoid excessive gains often appearing at high frequencies. In the experimental results presented in Section 3, a small regularization factor with value 10^{-5} is considered. The equalization filter frequency response of length W is computed in the unwarped domain by interpolating with a cubic spline [14] the K values of $H_{\mathrm{inv}_l}(\kappa)$.

5. After that, the inverse FFT of the interpolated frequency response is performed.

6. Finally, the resulting sequence is truncated in order to determine the equalization filter.

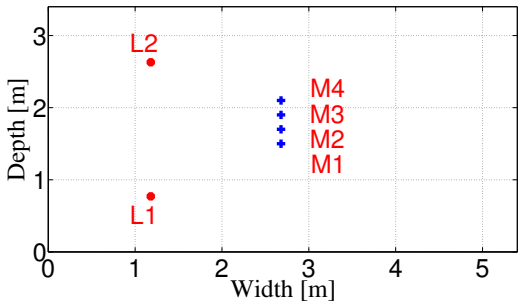
Therefore, it results a valid approach for a real-time application since its computational complexity is mostly determined by the inverse FFT, that is an $O(W \log W)$ algorithm [34].

3. Experimental results

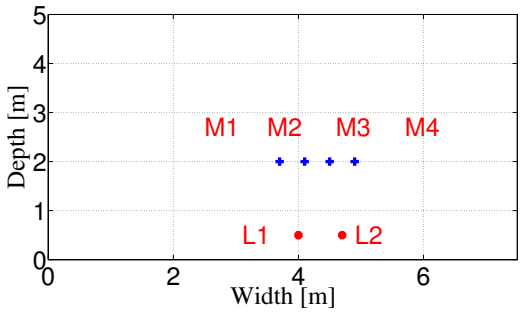
Some experimental results are reported in this section to show the effectiveness of the proposed approach. The performance of the decorrelation algorithm has not been included for space limitations and also because its effectiveness was fully demonstrated in [25]. Also the effectiveness of the multi-point approach has been demonstrated in several papers of the same authors, e.g., [6, 8, 19].

Several tests have been done considering impulse responses measured two real rooms. Rooms size with loudspeakers and microphones positions are shown in Figures 5 and 5ii: the microphones distance was set to 30 cm and they were placed at 1.2 m height. Professional equipment was used following the procedure described in [19]. More in details, measurements have been performed using a professional ASIO sound card and microphones with an omnidirectional response. A personal computer running NU-Tech platform has been used to manage all I/Os [35]. The impulse responses have been derived using a logarithmic sweep signal excitation [36] at 48 kHz sampling frequency. These responses are then used as terms of comparison in the identification procedure.

For the adaptation procedure a filter length L_f of 4096 samples has been



i



ii

Figure 5: Loudspeakers and microphones positions for (i) experiments 1 and 2 and (i) experiments 3 and 4.

considered, working on blocks of $N = 8192$ samples, with $D = 256$ and $f_s = 48$ kHz.

The equalizer is designed in the warped domain considering $W = 8192$ frequency points and the final length of the equalizer is 1024 samples. The equalized frequency range goes from 10 Hz to 20 kHz with the same sampling frequency of 48 kHz. The stereo input signals used for the presented results are the following sound tracks:

- "International Geophysical Year" from Donald Fagen,
- "I Sat by the Ocean" from Queens Of The Stone Age,

corresponding to Experiment 1/3 and 2/4 respectively. Two songs have been chosen as input signals in order to show the algorithm performance considering

a real scenario and variable inputs. This is a very important aspect since the identification algorithm must work during the sound reproduction without altering the sound perception. Some audio samples of these experiments can be downloaded at [37].

3.1. Results on Subband Identification procedure

Figure 6 shows the four impulse responses relative to the right loudspeaker channel considering the first sound track and room 1 (e.g., Experiment 1), Figure 7 shows the four impulse responses relative to the left loudspeaker channel for the second sound track and room 1(e.g., Experiment 2), Figure 8 shows the four impulse responses relative to the right loudspeaker channel considering the first sound track and room 2 (e.g., Experiment 3), and finally Figure 9 shows the four impulse responses relative to the left loudspeaker channel for the second sound track and room 2 (e.g., Experiment 4) . The impulse responses identified with a logarithmic sweep signal procedure are compared with the responses obtained with the subband identification procedure considering $D = 1$ and $D = 256$ subbands. The good performance obtained in terms of identification are also confirmed by the results reported in Figures 10, 11, 12, 13 where a frequency difference of the impulse responses identified with the subband procedure and the impulse responses obtained with the sweep identification are reported.

It is evident that the subband structure is capable of identifying the impulse responses and, increasing the number of subbands, it is possible to obtain a very accurate estimation of the responses. This is evident for all experiments thus considering different inputs. Furthermore, the use of the subband structure allows us to obtain a quicker converge increasing the number of subbands D, as shown in Figure 14.

This is due to the fact that increasing the number of subbands, the signal is divided in small frequencies parts where it is more stationary and where it is possible to use a more suitable stepsize exploiting Eqs. (19) and (20). However, increasing the number of subbands could lead to an increase of the memory usage of the hardware system, increasing the parallel computational load. A

compromise between identification performance and hardware capability should be evaluated.

3.2. Results on Equalization procedure

Figure 15, 16, 17, and 18 show the results for the four experiments in terms of smoothed real room magnitude responses identified with the subband procedure, the prototype responses and the equalization curves obtained from the single band identification and the subband identification. It is evident that there are several differences due to a different accuracy in the identification of the responses considering the subband procedure. This implies also a different resolution of the equalization curve that can increase the quality of final equalization procedure. This is also demonstrated considering the spectral deviation (SD) measure as reported in Table 1 and 2. The SD gives a measure of the deviation of the magnitude response from a flat one [38], considering each IR before and after the equalization as also reported in [8]. It is evident that using the subband identification procedure, it is still possible to obtain a reduction of the SD, especially considering the results obtained in the low frequency range, i.e., in Table 2.

4. Conclusions

In this paper, a subband implementation of a multichannel multiple position adaptive room response equalizer has been presented. The novelty of the algorithm is focused on the introduction of a subband identification procedure capable of identifying the multichannel impulse responses with a great accuracy and with an increase in terms of convergence rate since the approach is based on a multirate implementation. The accuracy of the identification is fundamental for the generation of the equalization filters and the quick convergence is very important especially for those environments that are very unstationary in terms of acoustic response. Experimental results carried on real environments have underlined the positive aspects of the introduction of a subband identification procedure taking into consideration the final equalization procedure.

Table 1: SD evaluation considering the single band identification ($D = 1$) and the subband identification ($D = 256$) for both experiments with a frequency range of $10 - 20000$ Hz.

Experiments Subbands Number Frequency Range	Initial SD	Final SD $D = 1$ 10Hz-20kHz	Final SD $D = 256$ 10Hz-20kHz
EX1 - right channel	10.96	2.60	**2.59**
EX1 - left channel	9.50	2.66	**2.65**
EX2 - right channel	10.96	2.60	**2.59**
EX2 - left channel	9.50	2.66	**2.65**
EX3 - right channel	10.41	2.65	**2.64**
EX3 - left channel	12.17	2.59	**2.58**
EX4 - right channel	10.41	2.65	**2.64**
EX4 - left channel	12.17	2.59	**2.58**

Table 2: SD evaluation considering the single band identification ($D = 1$) and the subband identification ($D = 256$) for both experiments with a frequency range of $10 - 200$ Hz.

Experiments Subbands Number Frequency Range	Initial SD	Final SD $D = 1$ 10Hz-200Hz	Final SD $D = 256$ 10Hz-200Hz
EX1 - right channel	4.09	3.98	**3.79**
EX1 - left channel	3.91	2.81	**2.64**
EX2 - right channel	4.09	3.90	**3.79**
EX2 - left channel	3.91	2.65	**2.63**
EX3 - right channel	3.80	3.43	**3.20**
EX3 - left channel	3.78	3.25	**2.95**
EX4 - right channel	3.80	3.36	**3.20**
EX4 - left channel	3.78	3.33	**3.00**

References

[1] S. Bharitkar, C. Kyriakakis, Immersive Audio Signal Processing, Springer, 2006.

[2] J. Mourjopoulos, On the Variation and Invertibility of Room Impulse Response Functions, J. Sound Vibr. 102 (2) (1985) 217–228.

[3] P. D. Hatziantoniou, J. N. Mourjopoulos, Errors in Real-Time Room Acoustics Dereverberation, J. Audio Eng. Soc. 52 (9) (2004) 883–899.

[4] S. T. Neely, J. B. Allen, Invertibility of a Room Impulse Response, J. Acoust. Soc. Amer. 66 (1) (1979) 165–169.

[5] A. Carini, S. Cecchi, I. Omiciuolo, F. Piazza, G. L. Sicuranza, Multiple Position Room Response Equalization in Frequency Domain, IEEE Trans. Speech and Audio Processing 2 (1) (2012) 122–135.

[6] S. Cecchi, L. Romoli, A. Carini, F. Piazza, A Multichannel and Multiple Position Adaptive Room Response Equalizer in Warped Domain: Real-time Implementation and Performance Evaluation, Applied Acoustics 82 (2014) 28–37.

[7] L.-J. Brannmark, A. Bahne, A. Ahlen, Compensation of loudspeaker–room responses in a robust mimo control framework, IEEE transactions on audio, speech, and language processing 21 (6) (2013) 1201–1216.

[8] S. Cecchi, A. Carini, S. Spors, Room Response Equalization - A Review, Applied Sciences (Switzerland) 8 (1).

[9] S. J. Elliott, P. A. Nelson, Multiple-Point Equalization in a Room Using Adaptive Digital Filters, J. Audio Eng. Soc. 37 (11) (1989) 899–907.

[10] L. Fuster, M. de Diego, M. Ferrer, A. Gonzalez, G. Pinero, A Biased Multichannel Adaptive Algorithm for Room Equalization, in: Proc. 20th European Signal Processing Conference, Bucharest, Romania, 2012, pp. 1344–1348.

[11] S. Cecchi, L. Romoli, F. Piazza, A. Carini, A Multichannel and Multiple Position Adaptive Room Response Equalizer in Warped Domain, Proc. 8th Int'l Symposium on Image and Signal Processing and Analysis (Sep. 2013).

[12] A. J. S. Ferreira, A. Leite, An Improved Adaptive Room Equalization in the Frequency Domain, in: Proc. 118th Audio Engineering Society Convention, Barcelona, Spain, 2005.

[13] S. Cecchi, A. Primavera, F. Piazza, A. Carini, An Adaptive Multiple Position Room Response Equalizer, in: Proc. EUSIPCO 2011, Barcellona, Spain, 2011, pp. 1274–1278.

[14] S. Cecchi, A. Carini, A. Primavera, F. Piazza, An Adaptive Multiple Position Room Response Equalizer in Warped Domain, in: Proc. EUSIPCO 2012, Bucharest, Romania, 2012, pp. 1955–1959.

[15] J. Benesty, D. R. Morgan, M. M. Sondhi, A Better Understanding and an Improved Solution to the Specific Problems of Stereophonic Acoustic Echo Cancellation, IEEE Trans. Speech Audio Process. 6 (2) (1998) 156–165.

[16] J. Benesty, C. Paleologu, T. Gansler, S. Ciochina, A perspective on stereophonic acoustic echo cancellation, Vol. 4, Springer Science & Business Media, 2011.

[17] C. Stanciu, C. Paleologu, J. Benesty, S. Ciochina, F. Albu, Variable-forgetting factor rls for stereophonic acoustic echo cancellation with widely linear model, in: 2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO), IEEE, 2012, pp. 1960–1964.

[18] L. Romoli, S. Cecchi, P. Peretti, F. Piazza, A Mixed Decorrelation Approach for Stereo Acoustic Echo Cancellation Based on the Estimation of the Fundamental Frequency, IEEE Trans. Audio, Speech and Language Processing 20 (2) (2012) 690–698.

[19] S. Cecchi, L. Romoli, M. Gasparini, A. Carini, F. Bettarelli, An Adaptive Multichannel Identification System for Room Response Equalization, in: Proc. Int. Conf. Electronics, Computers and Artificial Intelligence, Bucharest, România, 2015.

[20] M. Petraglia, R. Alves, P. Diniz, New Structures for Adaptive Filtering in Subbands with Critical Sampling, IEEE Trans. Signal Process. 48 (12) (2000) 3316–3327.

[21] M. M. Sondhi, D. R. Morgan, J. L. Hall, Stereophonic Acoustic Echo Cancellation - An Overview of the Fundamental Problem, IEEE Signal processing letters 2 (8) (1995) 148–151.

[22] S. Shimauchi, S. Makino, Stereo Projection Echo Canceller with True Echo Path Estimation, in: 1995 International Conference on Acoustics, Speech, and Signal Processing, Vol. 5, IEEE, 1995, pp. 3059–3062.

[23] J. Benesty, M. M. Sondhi, Y. Huang, Springer Handbook of Speech Processing, Springer, 2008.

[24] E. Larsen, R. M. Aarts, Audio Bandwidth Extension, J. Wiley & Sons, 2004.

[25] L. Romoli, S. Cecchi, F. Piazza, A Novel Decorrelation Approach for Multichannel System Identification, in: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, Florence, Italy, 2014, pp. 6652–6656.

[26] J. Lee, E. Song, Y. Park, D. Youn, Effective Bass Enhancement Using Second-Order Adaptive Notch Filter, IEEE Trans. Consum. Electron. 54 (2008) 663–668.

[27] M. Ali, Stereophonic Acoustic Echo Cancellation System Using Time-Varying All-pass Filtering for Signal Decorrelation, in: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 6, Seattle, WA, USA, 1998, pp. 3689–3692.

[28] S. Cecchi, L. Romoli, P. Peretti, F. Piazza, Low-complexity Implementation of a Real-time Decorrelation Algorithm for Stereophonic Acoustic Echo Cancellation, Signal Processing 92 (11) (2012) 2668–2675.

[29] A. V. Oppenheim, R. W. Schafer, J. R. Buck, Discrete-Time Signal Processing, Prentice Hall International Inc., 1999, pp. 274-279.

[30] E. Zwicker, H. Fastl, Psychoacoustics: Facts and Models, Springer-Verlag, 1990.

[31] M. Petraglia, P. Batalheiro, Prototype Filter Design for Subband Adaptive Filtering Structures with Critical Sampling, in: Proc. IEEE International Symposium on Circuits and Systems, Vol. 1, Geneva, Switzerland, 2000, pp. 543–546.

[32] B. Bank, Combined Quasi-Anechoic and In-Room Equalization of Loudspeaker Responses, in: Proc. 134th Audio Engineering Society Convention, 2013.

[33] P. D. Hatziantoniou, J. N. Mourjopoulos, Generalized Fractional-Octave Smoothing of Audio and Acoustic Responses, J. Audio Eng. Soc. 48 (2000) 259–280.

[34] O. Kirkeby, P. A. Nelson, H. Hamada, F. Orduna-Bustamante, Fast Deconvolution of Multichannel Systems using Regularization, IEEE Trans. Speech Audio Process. 6 (2) (1998) 189–194.

[35] A. Lattanzi, F. Bettarelli, S. Cecchi, NU-Tech: the entry tool of the hArtes toolchain for algorithms design, Proc. 124th Audio Engineering Society Convention.

[36] S. Cecchi, L. Palestini, P. Peretti, L. Romoli, F. Piazza, A. Carini, Evaluation of a Multipoint Equalization System based on Impulse Responses Prototype Extraction, in: Proc. 127th Audio Engineering Society Convention, New York, NY, USA, 2009.

[37] S. Cecchi, et al., Audio samples, `http://www2.units.it/ipl/research/AudEq/AppliedAcoustics20/data.htm`, [Online] (2020).

[38] S. Bharitkar, C. Kyriakakis, Immersive Audio Signal Processing, Springer, New York, 2006.

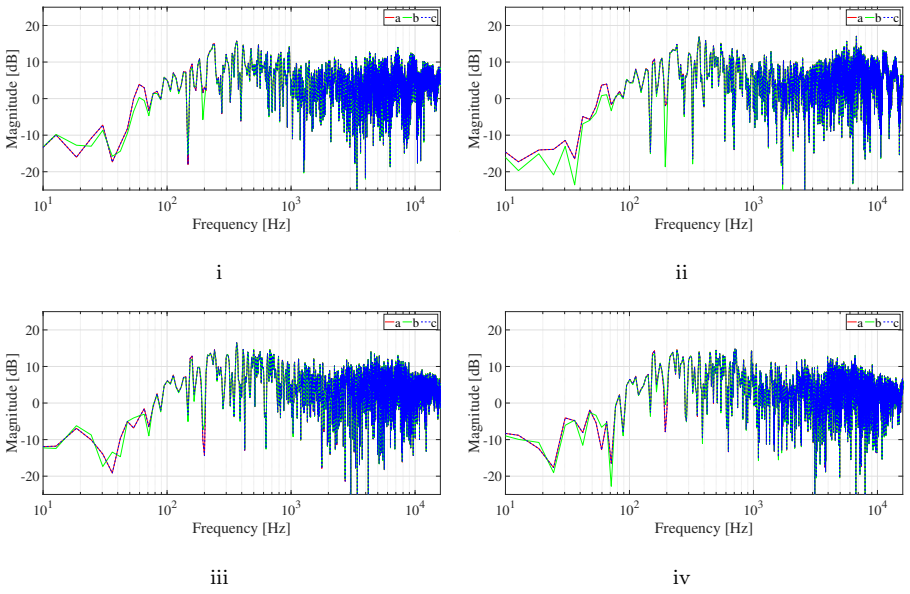


Figure 6: Experiment 1: (a) Real room magnitude responses, (b) identified room magnitude responses for $D = 1$, (c) identified room responses for $D = 256$ considering 4 microphones with reference to the Right channel.

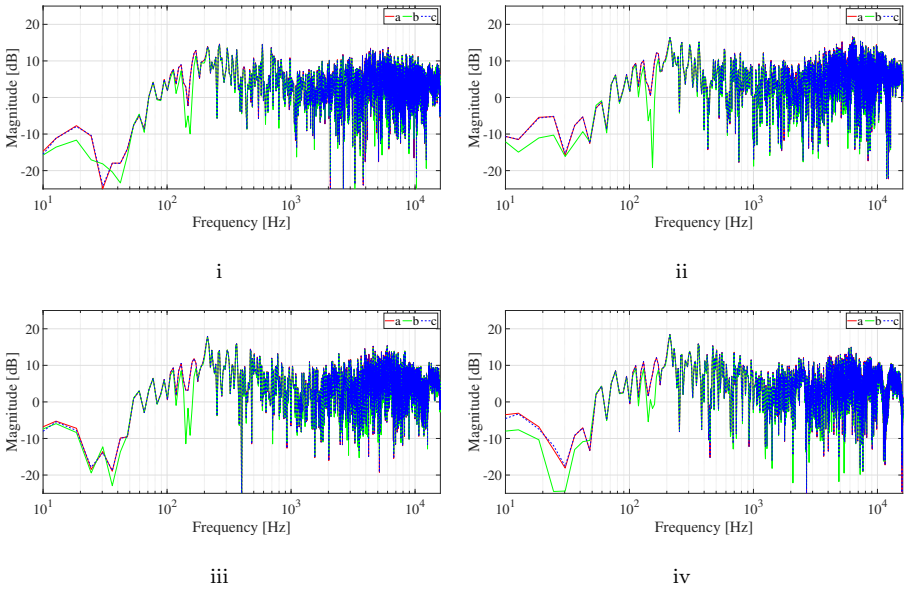


Figure 7: Experiment 2: (a) Real room magnitude responses, (b) identified room magnitude responses for $D = 1$, (c) identified room responses for $D = 256$ considering 4 microphones with reference to Left channel.

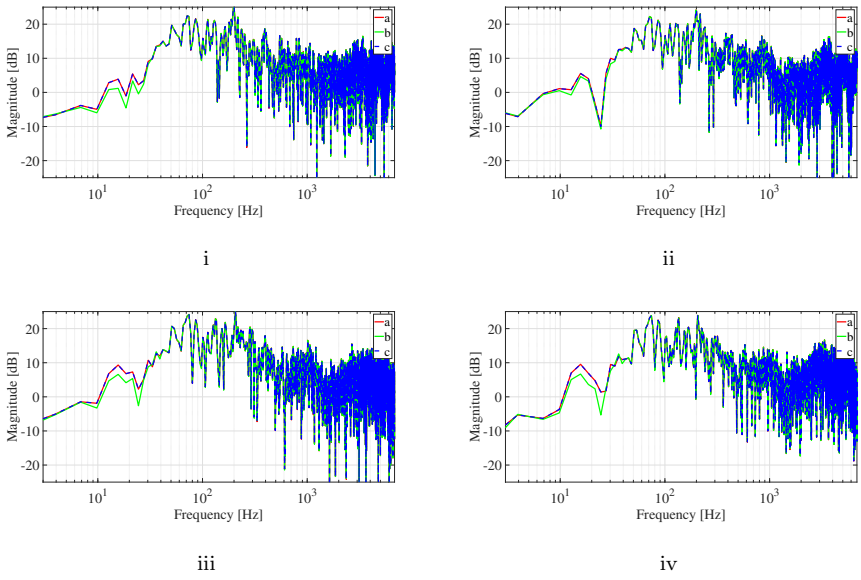


Figure 8: Experiment 3: (a) Real room magnitude responses, (b) identified room magnitude responses for $D = 1$, (c) identified room responses for $D = 256$ considering 4 microphones with reference to the Right channel.

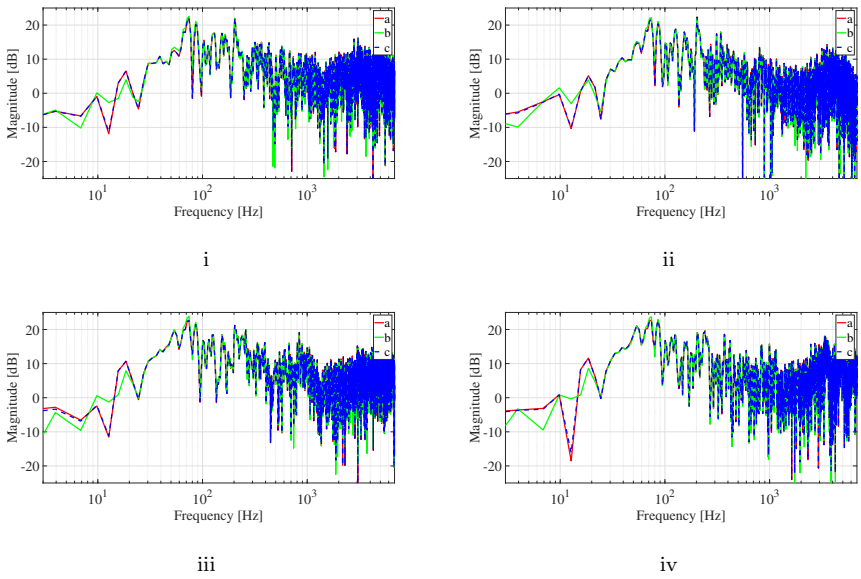
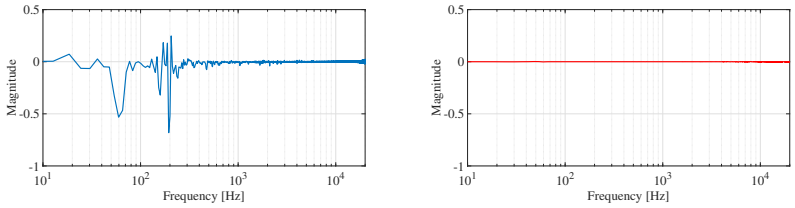
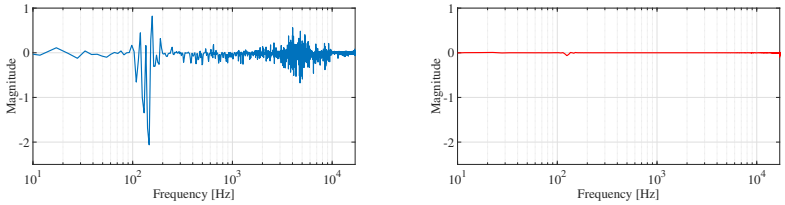


Figure 9: Experiment 4: (a) Real room magnitude responses, (b) identified room magnitude responses for $D = 1$, (c) identified room responses for $D = 256$ considering 4 microphones with reference to Left channel.



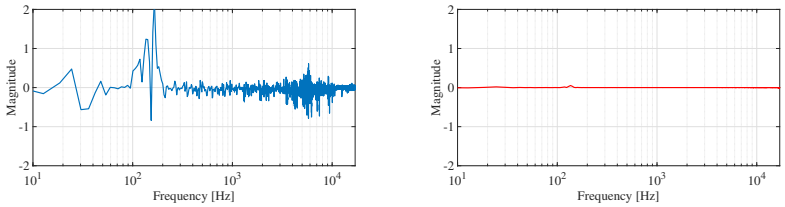
i ii

Figure 10: Experiment 1: (i) difference between the real room magnitude responses and the identified room magnitude responses for (i) $D = 1$ and for (ii) $D = 256$ considering one microphone with reference to the Right channel.



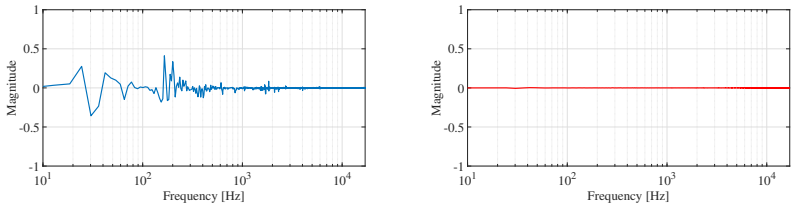
i ii

Figure 11: Experiment 2: difference between the real room magnitude responses and the identified room magnitude responses for (i) $D = 1$ and for (ii) $D = 256$ considering one microphone with reference to Left channel.



i ii

Figure 12: Experiment 3: difference between the real room magnitude responses and the identified room magnitude responses for (i) $D = 1$ and for (ii) $D = 256$ considering one microphone with reference to Left channel.



i ii

Figure 13: Experiment 4: difference between the real room magnitude responses and the identified room magnitude responses for (i) $D = 1$ and for (ii) $D = 256$ considering one microphone with reference to Left channel.

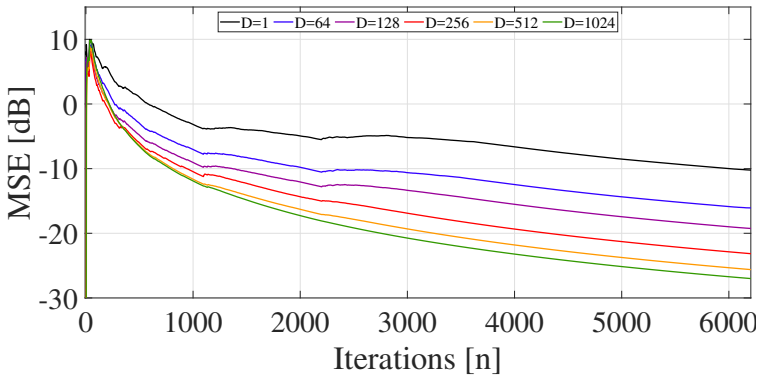
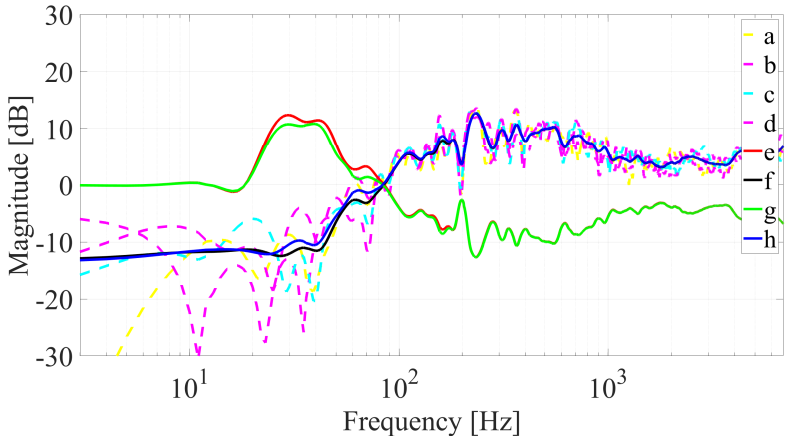
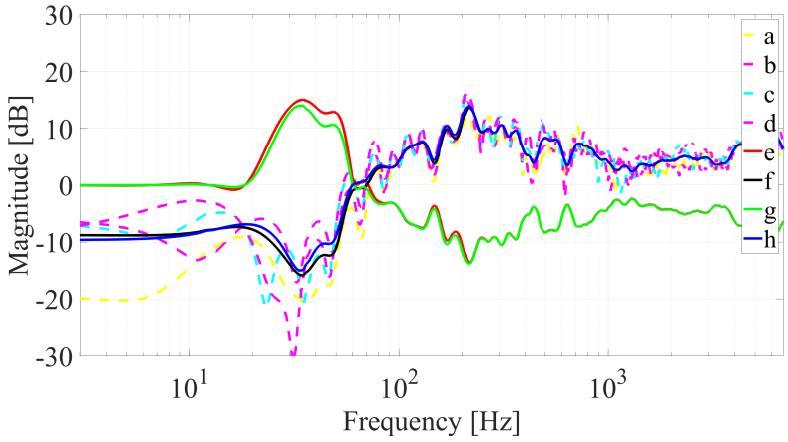


Figure 14: MSE for D=1, 64, 128, 256, 512, 1024, considering white noise as input signal.

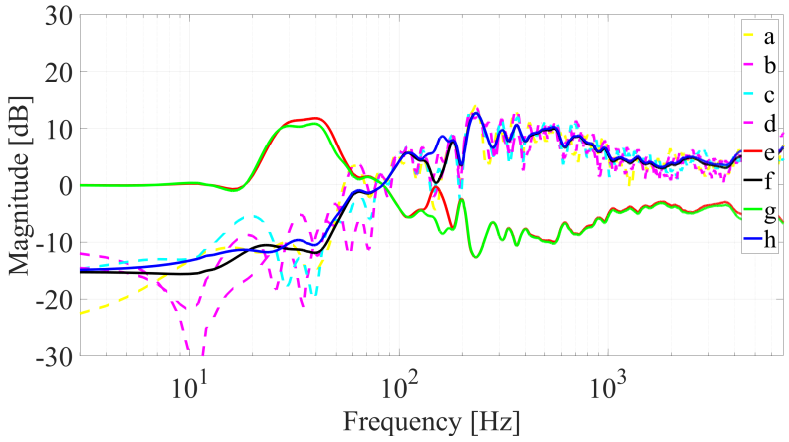


i

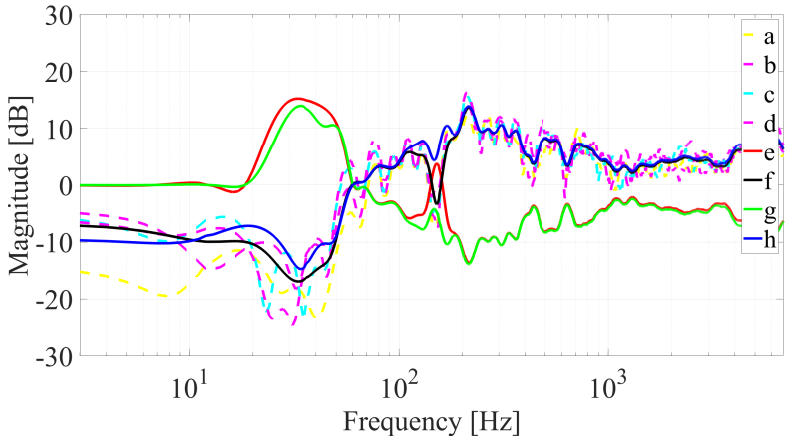


ii

Figure 15: Experiment 1: (a)-(d) Real room magnitude response, (e) equalization curve for single band identification, (f) prototype response considering a single band identification (g) equalization curve for subband identification (h) prototype response considering a subband identification, with respectively (i) Left channel and (ii) Right channel with a smoothing factor of 1/12.

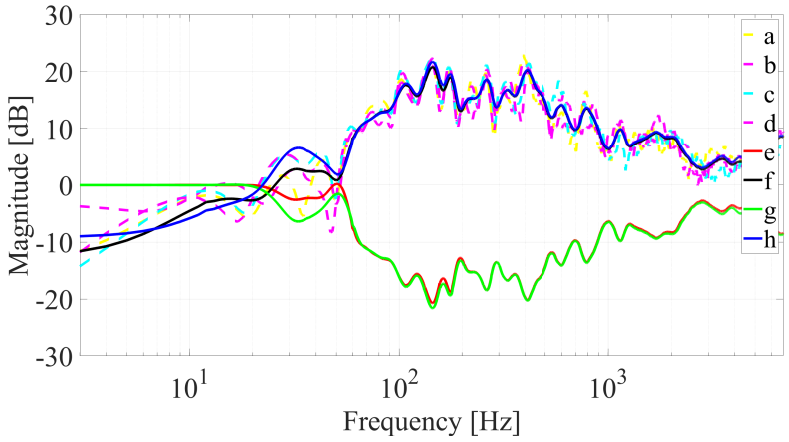


i

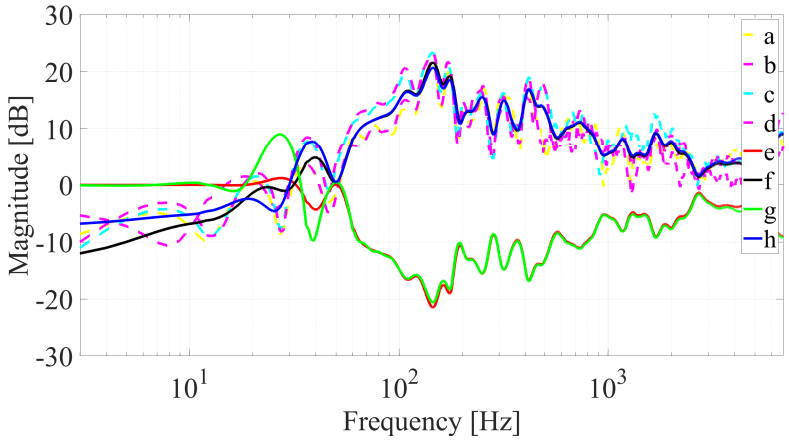


ii

Figure 16: Experiment 2: (a)-(d) Real room magnitude response, (e) equalization curve for single band identification, (f) prototype response considering a single band identification (g) equalization curve for subband identification (h) prototype response considering a subband identification, with respectively (i) Left channel and (ii) Right channel with a smoothing factor of 1/12.

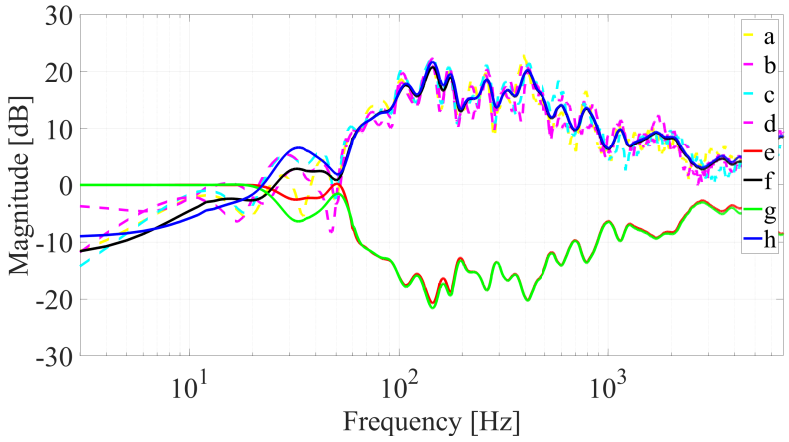


i

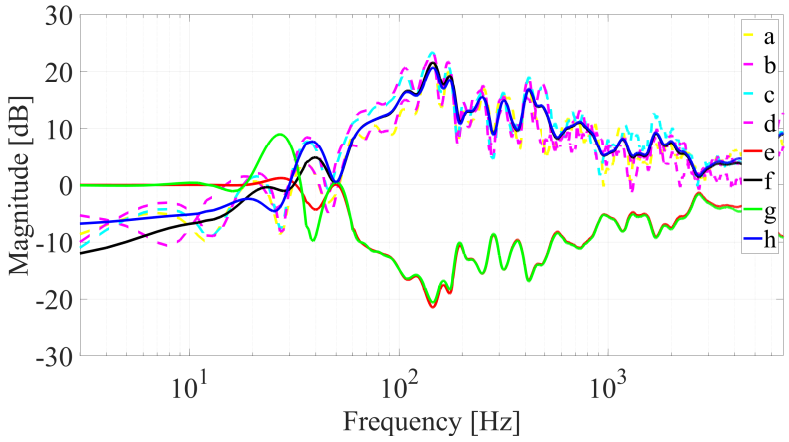


ii

Figure 17: Experiment 3: (a)-(d) Real room magnitude response, (e) equalization curve for single band identification, (f) prototype response considering a single band identification (g) equalization curve for subband identification (h) prototype response considering a subband identification, with respectively (i) Left channel and (ii) Right channel with a smoothing factor of 1/12.



i



ii

Figure 18: Experiment 4: (a)-(d) Real room magnitude response, (e) equalization curve for single band identification, (f) prototype response considering a single band identification (g) equalization curve for subband identification (h) prototype response considering a subband identification, with respectively (i) Left channel and (ii) Right channel with a smoothing factor of 1/12.