

MULTI-CHANNEL LOW-FREQUENCY ROOM EQUALIZATION USING PERCEPTUALLY MOTIVATED CONSTRAINED OPTIMIZATION

Mihailo Kolundžija, Christof Faller, and Martin Vetterli

School of Computer and Communication Sciences
Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland
{mihailo.kolundzija, christof.faller, martin.vetterli}@epfl.ch

ABSTRACT

We consider the problem of multiple-loudspeaker low-frequency room equalization for a wide listening area, where the equalized loudspeaker is helped using the remaining ones. Using a spatial discretization of the listening area, we formulate the problem as a multipoint error minimization between desired and synthesized magnitude frequency responses. The desired response and cost function are formulated with a goal of capturing the room’s spectral power profile, and penalizing strong resonances. Considering physical and psychoacoustical observations, we argue for the use of gain-limited, short, and well-localized equalization filters, with an additional delay for loudspeakers that help the equalized one. We propose a convex optimization framework for computing room equalization filters, where the mentioned filter requirements are incorporated as convex constraints. We verify the effectiveness of our equalization approach through simulations.

Index Terms— Room equalization, MIMO filtering, convex optimization

1. INTRODUCTION

Low-frequency room acoustics is dominated by usually clearly separated resonances or room modes. Since the density of resonances increases with the square of frequency [1], starting from some frequency—called the Schroeder frequency—room modes overlap and combine in a complex, location-dependent way that is best modeled by a statistical theory of room acoustics. The Schroeder frequency is dependent on the room’s geometry. For concert halls, it is on the lower end of audible frequencies; in listening rooms, which we focus on in this paper, it is roughly between 100 Hz and 200 Hz; and it can go up to several hundred Hertz in cars.

The low-frequency room resonances below the Schroeder frequency are thus characteristic of the entire room’s listening area, and they define the room’s bass performance. In music, strong resonances affect the timbre in an audible way, while for speech, the long reverberation tails associated with strong resonances blur the syllables and decrease the intelligibility. Thus, both in music and speech reproduction, it is important to reduce the effect of excessive resonances in order to improve the listening experience.

This work was supported by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS, <http://www.mics.org>), a center supported by the Swiss National Science Foundation under grant number 5005-67322. It was also supported by the Swiss National Science Foundation grant “New Sampling Methods for Processing and Communication”, grant number 200021-121935.

Above the Schroeder frequency, the high spatial variation of the resonant structure of room impulse responses (RIRs) makes the resonance control a highly location-dependent effort. This is the reason why some systems for wide-area room equalization do not correct the room beyond the Schroeder frequency, and why we focus only on equalizing low frequencies.

The first works on room equalization go back to 1960s, and they were in the spirit of controlling room modes. Namely, Boner [2] proposed the use of equalization to attenuate the resonances in sound systems. Groh [3] analyzed the low-frequency modal behavior of a room and performed equalization by finding adequate placement for a loudspeaker within a room. Similar, albeit more systematic approaches to optimizing the placement and number of low-frequency loudspeakers, have been investigated more recently by a number of authors (e.g., see [4, 5]). There have also been a fair amount of works recently with a focus on correcting the low-frequency modal behavior of a room using infinite impulse response (IIR) filters (e.g., see [6, 5, 7]).

In this paper, we present an approach for multiple-loudspeaker low-frequency room correction in an extended listening area, based on convex optimization. Our approach resembles multiple-point RIR equalization approaches (e.g., see [8]). However, it is more general, since it allows one to systematically incorporate physical and psychoacoustical aspects relevant to RIR correction through convex constraints. The psychoacoustical phenomena of particular interest for room equalization are temporal masking and the precedence effect, and they can be incorporated through constraining the time-domain profile of the equalization filters’ impulse responses. Excessive driving of loudspeakers at some frequencies, characteristic to the efforts of correcting deep notches, is prevented by limiting the maximum gain of equalization filters over frequency.

2. PROPOSED ROOM EQUALIZATION

2.1. Problem description

Consider a listening room with a multichannel loudspeaker setup, consisting of L loudspeakers S_1, \dots, S_L . An example setup, which is considered later in Section 3, is shown in Fig. 1. Note that we are focusing on equalizing the response of one loudspeaker, denoted as the *main loudspeaker*, with the help of the remaining, for this task called *auxiliary loudspeakers*. Without loss of generality, we assign index 1 to the main loudspeaker.

As the first step, one needs to measure RIRs of all loudspeakers in N control points C_1, \dots, C_N which cover the listening area where RIRs are being equalized (gray rectangular area in Fig. 1). The placement of control points can be systematic or random, and as few as $N = 4$ control points can capture with high accuracy the 3D

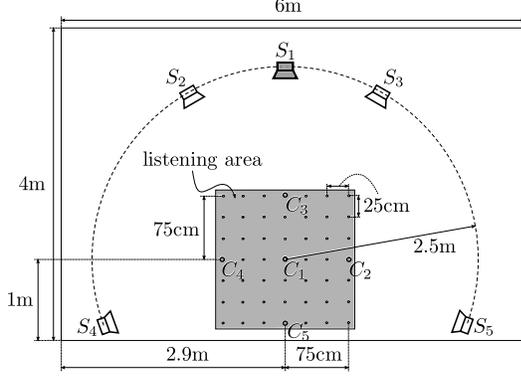


Fig. 1. Equalization of a five-channel loudspeaker setup in a room of dimensions (6m, 4m, 2.5m). In the illustration, the response of loudspeaker S_1 is equalized in the listening area around the central control point C_1 .

sound energy in a room, as reported in [9]. We denote by $G_{ij}(\omega)$ the frequency response of the RIR between loudspeaker j and control point i , and by $\mathbf{G}(\omega)$ the $N \times L$ matrix containing the frequency responses $G_{ij}(\omega)$.

One also needs to decide on the length N_h of equalization filters $h_i[n]$. Note that working with highly downsampled signals allows using relatively short filters, and multirate filtering offers savings in the computational complexity. Thus, N_h can be up to the order of the length of RIRs, which is still short in the downsampled domain corresponding to a highly reduced sampling frequency f'_s . Let the vector $\mathbf{h}_i = [h_i[0] \dots h_i[N_h - 1]]^T$ contain the samples of the equalization filter of loudspeaker i , and the vector $\mathbf{h} = [\mathbf{h}_1^T \dots \mathbf{h}_L^T]^T$ contain the samples of all loudspeaker filters stacked together. Note that vector \mathbf{h} is what we are looking for.

Since our design procedure considers the equalized RIRs in the frequency domain, we discretize the frequency axis into N_f uniformly-spaced normalized frequencies $\omega_0, \dots, \omega_{N_f-1}$, where $\omega_0 = 0$, and $\omega_{N_f-1} = \pi$ corresponds to the Nyquist frequency $f'_s/2$. The frequency spacing needs to be of the order of room resonances' bandwidth, which can go down to around 1 Hz [1].

A vector $\mathbf{H}(\omega_i) = [H_1(\omega_i) \dots H_L(\omega_i)]^T$, containing the frequency responses of the equalization filters, is obtained by the following product:

$$\mathbf{H}(\omega_i) = \mathbf{V}(\omega_i) \mathbf{h}, \quad (1)$$

where

$$\mathbf{V}(\omega_i) = \mathbf{I}_{L \times L} \otimes [1 \ e^{j\omega_i} \dots e^{j(N_h-1)\omega_i}], \quad (2)$$

$\mathbf{I}_{L \times L}$ is an $L \times L$ identity matrix, and \otimes denotes the Kronecker product.

The frequency responses at the normalized frequency ω_i of the equalized RIRs in the control points are given by

$$\mathbf{Y}(\omega_i) = [Y_1(\omega_i) \dots Y_N(\omega_i)]^T = \mathbf{G}(\omega_i) \mathbf{H}(\omega_i). \quad (3)$$

In order to consider the equalized frequency responses in all control points C_1, \dots, C_N and at all frequencies $\omega_0, \dots, \omega_{N_f-1}$, vectors $\mathbf{Y}(\omega_i)$ are stacked into one long vector defined by

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}(\omega_0) \\ \vdots \\ \mathbf{Y}(\omega_{N_f-1}) \end{bmatrix} = \begin{bmatrix} \mathbf{G}^T(\omega_0) \mathbf{V}(\omega_0) \\ \vdots \\ \mathbf{G}^T(\omega_{N_f-1}) \mathbf{V}(\omega_{N_f-1}) \end{bmatrix} \mathbf{h}.$$

Essentially, the goal of equalization is to make the equalized RIRs as close as possible to the desired responses, discussed next, in all control points at all considered frequencies.

2.2. Desired response calculation

One of the main challenges when designing room equalizers is specifying the desired frequency response $D(\omega)$ the equalized system needs to achieve, and there is no wide consensus on this issue. On the other hand, it has long been suggested that an equalization procedure should not undo the effect of a room and make the reproduced sound artificially anechoic, but it should sensibly correct the room's undesired features, usually associated with strong low-frequency resonances.

As briefly mentioned at the beginning of this section, the listening area is sampled with control points in order to capture the essential properties of a sound field developed in the equalized room. At the same time, sampling multiple points allows to avoid position-dependent anomalies, such as deep frequency-response notches associated with nodes of some of the room modes.

It was shown in [9] that the root mean square (RMS) value of the magnitude frequency response of RIR taken over several measurement points gives a stable estimate of the room's spectral power profile. Hence, for defining the desired frequency characteristic, we combine the mentioned spatial power averaging with magnitude frequency response smoothing in fractional octave bands. More specifically, the desired response $D(\omega_i)$ is obtained as follows:

$$D(\omega_i) = \sqrt{\frac{1}{N} \sum_{m=1}^N (\tilde{G}_{m1}(\omega_i))^2}, \quad (4)$$

where $\tilde{G}_{m1}(\omega)$ is fractional-octave (e.g., 1/3-octave) smoothed magnitude characteristic of the RIR between the main loudspeaker and control point m .¹

Since the same desired frequency response is taken for each control point, the vector of desired responses at frequency ω_i is given by $\mathbf{D}(\omega_i) = D(\omega_i) \mathbf{1}_{N \times 1}$, where the column-vector $\mathbf{1}_{N \times 1}$ contains N ones. In order to compare the equalized with desired frequency responses in control points, vectors $\mathbf{D}(\omega_i)$ need to be stacked together into the column vector $\mathbf{D} = [\mathbf{D}^T(\omega_0) \dots \mathbf{D}^T(\omega_{N_f-1})]^T$.

2.3. Choice of a cost function

Before defining the cost function, we first define a resonance detection vector $\mathbf{R} = [R(\omega_0) \dots R(\omega_{N_f})]^T$ by

$$R(\omega_i) = |\max_m (|G_{m1}(\omega_i)|) - D(\omega_i)| + \epsilon, \quad (5)$$

where ϵ is a predefined minimum weight that can be given to a spectral magnitude error. Note that \mathbf{R} is designed to peak at *resonant frequencies*.

For a cost function, we chose to use a weighted magnitude error that penalizes more the errors at resonant frequencies, and it is given by

$$J = \|\mathbf{W} (|\mathbf{Y}| - |\mathbf{D}|)\|_2, \quad (6)$$

where $\mathbf{W} = \text{diag}(\mathbf{R}) \otimes \mathbf{I}_{N \times N}$, and $\text{diag}(\cdot)$ makes a diagonal matrix from a vector, with vector's entries on the main diagonal.

¹Fractional smoothing is thoroughly described in [10].

2.4. Equalization filter constraints

Here we present some psychoacoustical and physical considerations that are used to constrain the computed filters in order to avoid the location-sensitivity characteristic of some room equalization approaches.

2.4.1. Temporal-masking constraints

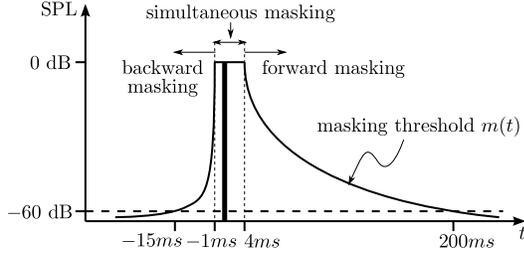


Fig. 2. Illustration of temporal masking for a single pulse.

Temporal masking, illustrated in Fig. 2, is a phenomenon where a sound stimulus renders inaudible sounds which precede (backward masking) and follow it (forward masking) [11]. In [12], Fielder proposed considering temporal masking as a criterion for time-domain distortions of equalization systems. In light of the arguments about the limitations of using long equalization filters or filters that exhibit pre-echos, and considering temporal masking, we can expect that the use of short equalization filters has a good chance of avoiding sensitivity to location changes. In other words, we argue that short and well localized (“spiky”) filters, whose amplitude profiles fit into the temporal masking threshold curve $m(t)$ from Fig. 2, are good solutions for wide-area equalization. Our argument is corroborated by the fact that if a sharp transient is emitted by a loudspeaker, a listener close to the loudspeaker will not hear a distortion thanks to temporal masking.

The temporal-masking constraint is defined as a maximum-amplitude limit to a filter’s impulse response. If $m[n]$ is a sampled version of the temporal masking threshold curve $m(t)$, then the equalization filter $h_i[n]$ is constrained with

$$|h_i[n]| \leq m[n]. \quad (7)$$

2.4.2. Auxiliary loudspeaker filter constraints

In addition to the temporal masking constraint, two additional constraints are considered for the auxiliary loudspeakers:

- To prevent that the sounds from the auxiliary loudspeakers appear as echoes, we use a combination of delay and gain relative to the main loudspeaker’s filter. The delay and gain should be such that the auxiliary loudspeaker signals are below the echo threshold [13].
- To prevent that the signal is perceived away from the main loudspeaker, the delay needs to be at least about 1 ms over the whole listening area, such that the precedence effect [13] is active and sound is perceived at the main loudspeaker.

The described constraints for the auxiliary loudspeakers are realized by the following modification of the temporal-masking constraint (7):

$$|h_i[n]| \leq a m[n - n_0], \quad (8)$$

where a is a positive attenuation factor, and n_0 is a lag corresponding to a delay of around ten milliseconds. Note that both (7) and (8) are convex constraints. Using vectors, (7) and (8) are combined to

$$|\mathbf{h}| \preceq \mathbf{m}, \quad (9)$$

where \preceq denotes a component-wise \leq .

2.4.3. Maximum-gain constraints

An additional insurance against location sensitivity is putting a limit on the equalization filter’s gain over frequencies. Namely, this avoids excessive driving of a loudspeaker in order to correct a deep notch, which is usually highly position-dependent. The gain of equalization filters are limited using the following set of convex constraints:

$$|\mathbf{V}(\omega_i) \mathbf{h}| \preceq \mathbf{H}^{max}, \quad \forall \omega_i, \quad (10)$$

where $\mathbf{H}^{max} = [H_1^{max} \dots H_L^{max}]^T$ is a vector with the maximum gain for each loudspeaker.

2.5. Filter computation procedure

The cost function (6) is not convex, preventing the use of conventional convex optimization tools [14] and making the problem difficult to solve. We propose Algorithm 1, which is a variation of the local optimization algorithm by Kassakian [15]. It iteratively solves a convex program and improves the solution at every step, converging to a local optimum.

Algorithm 1 Solving a constrained, weighted magnitude least squares problem (adapted from [15]).

1. Choose the solution tolerance ϵ
2. Choose the initial solution \mathbf{h}
3. **repeat**
4. $J \leftarrow \|\mathbf{W} (|\mathbf{Y}| - |\mathbf{D}|)\|_2$
5. Compute $\hat{\mathbf{D}}$ such that $\forall \omega_j \in \{\omega_0, \dots, \omega_{N_f-1}\}$

$$\begin{aligned} |\hat{\mathbf{D}}(\omega_j)| &= |\mathbf{D}(\omega_j)| \\ \angle \hat{\mathbf{D}}(\omega_j) &= \angle (\mathbf{G}(\omega_j) \mathbf{H}(\omega_j)) \end{aligned}$$

6. Solve the following convex program

$$\begin{aligned} &\text{minimize} && \|\mathbf{W} (\mathbf{Y} - \hat{\mathbf{D}})\|_2 \\ &\text{subject to} && |\mathbf{h}| \preceq \mathbf{m} \\ &&& |\mathbf{V}(\omega_j) \mathbf{h}| \preceq \mathbf{H}^{max}, \quad j \in \{0, \dots, N_f - 1\} \end{aligned}$$

7. $J' \leftarrow \|\mathbf{W} (|\mathbf{Y}| - |\mathbf{D}|)\|_2$
 8. **until** $|J' - J| < \epsilon$
-

Algorithm 1 is not guaranteed convergence to the global minimum, but it is at least as good as the complex least squares solution, which uses (6) without absolute-value signs.

3. SIMULATIONS

In order to verify the effectiveness of our room equalization strategy, we performed a simulation of a five-channel full-range loudspeaker setup, shown in Fig. 1. We show the equalization of the front loudspeaker S_1 (center channel)² with the help of the remaining four.

²Note that our numbering differs from the usual way channels are numbered in five-channel surround.

Equalization is done up to the frequency $f_{max} = 200$ Hz, which enables filtering in the downsampled domain corresponding to the sampling frequency $f'_S = 400$ Hz. Equalization filters have the length of $N_h = 16$ samples. For auxiliary loudspeakers' filters, we used a delay n_0 that corresponds to 10 ms, with attenuation $a = 0.25$. The resulting equalization filters, both in the time and frequency domain, together with the amplitude constraints, are shown in Fig. 3.

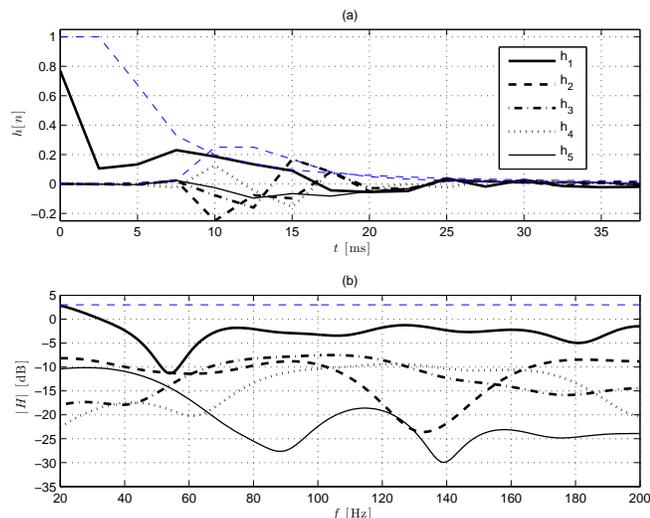


Fig. 3. Time-domain (a) and frequency-domain (b) characteristics of loudspeaker equalization filters $h_i[n]$. Thin dashed lines mark the amplitude constraints for the main and auxiliary loudspeakers (a) and the maximum-gain constraint (b).

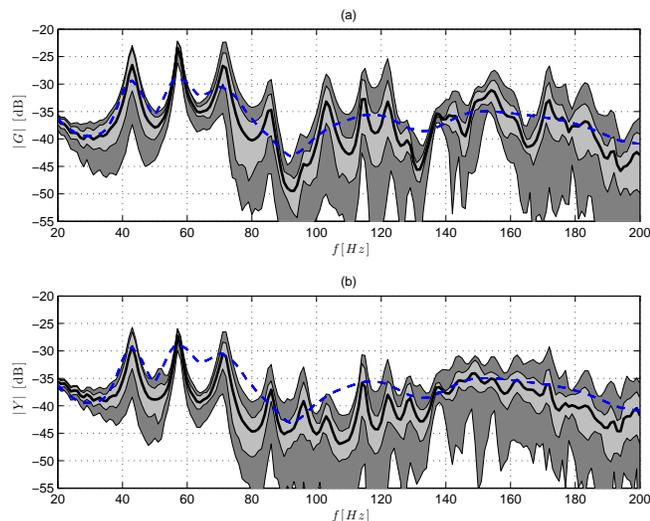


Fig. 4. Mean (solid), 25-75 (light-gray), and 3-97 (dark-gray) percentiles of the magnitude frequency responses on a rectangular grid of control points spaced at 25 cm, shown in Fig. 1, before (a) and after (b) equalizing loudspeaker S_1 . The desired frequency characteristic $D(\omega)$ is shown with a dashed line.

Fig. 4 shows the variations of RIR frequency characteristics on a

rectangular grid of control points, spaced at 25 cm, covering the listening area (see Fig. 1). From Fig. 4, it can be seen that after equalization, strong resonances get attenuated, as desired. Also, above 80 Hz, the average magnitude frequency characteristic of the equalized RIRs inside the listening area gets significantly improved, exhibiting a smoother behavior and approximating the desired magnitude characteristic more closely.

4. CONCLUSION

In this paper, we have presented an approach for low-frequency multiple-loudspeaker RIR equalization based on convex optimization. We have shown a way to formulate typical psychoacoustical and physical RIR equalization design constraints in terms of convex constraints on the equalization filters, allowing finding optimal solutions in a systematic fashion. We have also shown the effectiveness of our approach for equalizing a five-channel loudspeaker system in an extended listening area, using short filters and making sure that the equalization system does not cause undesired audible echoes and localization biases. The proposed approach, using the temporal masking constraint, can also be applied to single-loudspeaker room equalization.

5. REFERENCES

- [1] H. Kuttruff, *Room Acoustics*, Taylor & Francis, 2000.
- [2] C. P. Boner and C.R. Boner, "A procedure for controlling room-ring modes and feedback modes in sound systems with narrow-band filters," *J. Audio Eng. Soc.*, vol. 13, no. 4, pp. 297–299, 1965.
- [3] A. R. Groh, "High-fidelity sound system equalization by analysis of standing waves," *J. Audio Eng. Soc.*, vol. 22, no. 10, pp. 795–799, 1974.
- [4] A. Celestinos, *Low frequency sound field enhancement system for rectangular rooms using multiple loudspeakers*, Ph.D. thesis, Aalborg University, 2006.
- [5] T. Welti and A. Devantier, "Low-frequency optimization using multiple subwoofers," *J. Audio Eng. Soc.*, vol. 54, no. 5, pp. 347–364, 2006.
- [6] A. Mäkivirta, P. Antsalo, M. Karjalainen, and V. Välimäki, "Modal equalization of loudspeaker-room responses at low frequencies," *J. Audio Eng. Soc.*, vol. 51, no. 5, pp. 324–343, 2003.
- [7] R. J. Wilson, M. D. Capp, and J. R. Stuart, "The loudspeaker-room interface-controlling excitation of room modes," in *Proc. of the 23rd AES Conf*, 2003.
- [8] S. J. Elliott and P. A. Nelson, "Multiple-point equalization in a room using adaptive digital filters," *J. Audio Eng. Soc.*, vol. 37, no. 11, pp. 899–907, 1989.
- [9] J. A. Pedersen, "Sampling the energy in a 3-D sound field," *Preprint 130th Conv. Aud. Eng. Soc.*, Oct 2007.
- [10] P. D. Hatziantoniou and J. N. Mourjopoulos, "Generalized fractional-octave smoothing of audio and acoustic responses," *J. Audio Eng. Soc.*, vol. 48, no. 4, pp. 259–280, 2000.
- [11] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, Academic Press, 1989.
- [12] L. D. Fielder, "Analysis of traditional and reverberation-reducing methods of room equalization," *J. Audio Eng. Soc.*, vol. 51, no. 1/2, pp. 3–26, 2003.
- [13] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, 1997.
- [14] S. P. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [15] P. W. Kassakian, *Convex approximation and optimization with applications in magnitude filter design and radiation pattern synthesis*, Ph.D. thesis, University of California, Berkeley, 2006.