ROOM IMPULSE RESPONSE SHORTENING WITH INFINITY-NORM OPTIMIZATION

Tiemin Mei, Alfred Mertins, and Markus Kallinger

Institute for Signal Processing,
University of Lübeck, 23538 Germany

ABSTRACT

The purpose of room impulse response (RIR) shortening is to improve the intelligibility of the received signal by prefiltering the source signal before it is played with a loudspeaker in a closed room. In this paper, we propose to use the infinity-norm as optimization criterion for the design of shortening filters of RIRs. Similar to the equiripple filter design method, design errors will be uniformly distributed over the unwanted temporal range of the shortened global impulse response. The D50 measure is exploited during the design of the shortening filter, which makes it possible to significantly reduce the length of the prefilter without affecting the perceived performance.

***Index Terms*—** room impulse response, shortening, infinity-norm, optimization.

1. INTRODUCTION

For the enhancement of speech intelligibility in reverberant rooms or closed spaces such as cars and for new applications in audio-visual communication and virtual acoustics, a suitable pre-processing of loudspeaker signals is needed to reduce room reverberation, namely, the listening-room-compensation (LRC) or room-reverberation compensation [1][2]. Room-reverberation compensation is something different from channel equalization. For channel equalization, the original signal is recovered completely from the received signal which is deformed by the channel [3]. On the other hand, room-reverberation compensation will compensate the received signal so that it is perceived without reverberation, in other words, there are not any echoes being heard, which means it is a kind of partial equalization [4][5].

Room impulse responses (RIRs) are usually nonminimum phase and of very high order. Strong late echoes will possibly be heard and deteriorate the intelligibility. According to the special psychoacoustic perceptual properties of the human auditory system, a complete equalization of room reverberation is not necessary [9]. A simpler and more efficient way than complete equalization is to equalize partially the RIR, which implies that the equivalent room impulse response is either reshaped or shortened, so that there remains no audible strong and long-delayed echo. This will greatly alleviate the pressure of designing such a compensation system.

To shorten or reshape the global impulse response, the prefilter can be designed in different ways and according to different criteria. For instance, for the least-squares plus post-processing approach in [1], the performance depends closely on the post-processing filter. Another example is the homomorphic-based minimum-phase inverse filter design method plus dominant-poles relocation [5]. In this method, the dominant poles of the inverse filter of the minimum-phase part of the RIR, which are the ones that are closest to the unit circle in the complex plane, are moved closer to the origin so as to quicken the decay of the resulting inverse filter. Therefore the global impulse response will decay faster, which leads to a partial rather than a complete equalization.

For room-reverberation compensation, we should not only consider the simple shortening/reshaping process according to some standard optimality criteria such as least squares, but also take the psychoacoustic properties of the human auditory system into account. In particular, one should aim to obtain an optimal prefilter in the sense of intelligibility at the lowest implementation cost. There are different appropriate psychoacoustic criteria [6][7][8]. One of them is the D50 measure for intelligibility of speech [9], which is defined as the ratio of the energy within 50 ms after the first peak of the room impulse response versus the complete impulse response's energy. The intelligibility is guaranteed if the RIR is shortened so that the energy is concentrated in 50ms after the first impulse of the RIR.

The least-squares method is the most widely used approach in optimization - simplicity and linearity are its advantages. But it has drawbacks too, namely, non-uniformity. For instance, in traditional least-square equalization for LRC, the squared errors are distributed non-uniformly along the time axis, which typically results in additional late reverberation of the global system with audible echoes as reported in [1].

It is well known that the infinity-norm criterion is often used for robust estimation and robust control system design [10]. In this paper we combined the infinity-norm with the D50 measure to define suitable optimization criteria. Al-

This work has been supported by the German Research Foundation under Grant No. ME1170/3-1. Email: mertins@isip.uni-luebeck.de. M. Kallinger is now with the Fraunhofer Institute for Integrated Circuits IIS, Erlangen

though this will lead to high nonlinearity in the optimization process, this also leads to high uniformity. So we can control the errors distributed uniformly, no late reverberation is produced, and the echoes are controlled exactly.

2. PROBLEM STATEMENT

Let $c(n)$ denote the impulse response of a room, and let L_c be the length of $c(n)$. Moreover, let $h(n)$ denote the impulse response of a prefilter with length L_h. The global impulse response of this prefilter-loudspeaker-room system is as follows, where we have subsumed the loudspeaker response as a part of the room impulse response:

$$g(n) = h(n) * c(n) = \mathbf{Ch}, \tag{1}$$

where $\mathbf{C}$ is a L_g-by-L_h convolution matrix made up of sequence $c(n)$. The length of $g(n)$ is $L_g = L_c + L_h - 1$. Our aim is to design a prefilter that makes the global impulse response $g(n)$ decay faster than the impulse response of the room and increase the D50 measure.

In filter shortening, we use two window function $w_d(n)$ and $w_u(n)$ to divide the global impulse response $g(n)$ into two parts: the desired part $g_d(n) = w_d(n)g(n)$ and the unwanted part $g_u(n) = w_u(n)g(n)$. Our goal is to minimize some function of $|g_u(n)|$ while maximizing (or keeping constant) another function of $|g_d(n)|$ with respect to the prefilter $h(n)$ without significantly affecting the magnitude frequency response of the global system. For quadratic functions, and when not taking the frequency responses into account, this means that the energy of $g(n)$ should be concentrated in the desired part $g_d(n)$.

A conventional approach is to optimize $h(n)$ under the least-squares error criterion, that is [1][12],

$$\begin{cases} \text{Min}_{\mathbf{h}}: & f(\mathbf{h}) = \mathbf{g}_u^{\mathrm{H}}\mathbf{g}_u \\ \text{S. t.}: & \mathbf{g}_d^{\mathrm{H}}\mathbf{g}_d = \text{constant} \end{cases} \tag{2}$$

This least-squares problem is equivalent to the following generalized eigenvalue decomposition,

$$\mathbf{A}\mathbf{h}_{\text{opt}} = \lambda_{\min}\mathbf{B}\mathbf{h}_{\text{opt}}, \tag{3}$$

with

$$\begin{aligned} \mathbf{B} &= \mathbf{C}^{\mathrm{H}}\text{diag}[\mathbf{w}_d]^{\mathrm{H}}\text{diag}[\mathbf{w}_d]\mathbf{C} \\ \mathbf{A} &= \mathbf{C}^{\mathrm{H}}\text{diag}[\mathbf{w}_u]^{\mathrm{H}}\text{diag}[\mathbf{w}_u]\mathbf{C}. \end{aligned}$$

In [13], the window $w_d(n)$ is defined as a rectangular window, and $w_u(n)$ the complement of $w_d(n)$. The position of window $w_d(n)$ is optimized at the same time so as to get the optimally shortened global impulse response $g(n)$.

Unfortunately, such a prefilter $\mathbf{h}_{\text{opt}}$ that is optimal in the least-squares sense will make great distortion to the global impulse response $g(n)$ in frequency domain, and, in addition, the global time-domain impulse response $g(n)$ often shows increased late taps which will cause obvious echoes in its output [1]. Although some measures have been taken to overcome such drawbacks [1], further improvement is needed in practice.

3. APPROACH DEVELOPMENT

For an optimal prefilter, we expect a quickly and monotonously decaying characteristic of the global impulse response $g(n)$ so that there will be no noticeable echoes. In other words, we want to control the attenuation characteristics of the global impulse response. Properly selected windows $w_d(n)$ and $w_u(n)$ will be helpful for improving this solution, but the more important point is that we should look for some best-suited optimality criteria rather than fix our attention on least-squares optimization.

The idea in this paper is motivated by the equiripple filter design [11]. In equiripple filter design, the purpose is to design a finite impulse response (FIR) filter which approximates a given filter so that the maximum of the weighted approximation errors is minimized in frequency domain, i.e., the so-called minimax or Chebyshev criterion. In our case, however, the criterion is somewhat different from the equiripple criterion in filter design. For the optimization of prefilters, we would like to minimize the maximum absolute value (i.e., the infinity-norm) of the unwanted part $g_u(n)$ given by

$$f_u(\mathbf{h}) = \|\mathbf{g}_u\|_\infty = \text{Max}\left[|\mathbf{g}_u|\right],$$

where $\mathbf{g}_u = \text{diag}[\mathbf{w}_u]\mathbf{Ch}$. At the same time, we would like to maximize the infinity-norm of the desired part $g_d(n)$:

$$f_d(\mathbf{h}) = \|\mathbf{g}_d\|_\infty = \text{Max}\left[|\mathbf{g}_d|\right]$$

with $\mathbf{g}_d = \text{diag}[\mathbf{w}_d]\mathbf{Ch}$. Maximizing $f_d(\mathbf{h})$ while keeping $f_u(\mathbf{h})$ constant will lead to the most-possibly flat frequency-domain characteristic of the global impulse response, because the other samples of $|g_d(n)|$ will become as small as possible when the largest tap, i.e. $\text{Max}\left[|\mathbf{g}_d|\right]$, is maximized. On the other hand, the minimization of $f_u(\mathbf{h})$ with constant $f_d(\mathbf{h})$ will result in the most-uniform distribution of errors, because for most of the samples of the unwanted part $|g_u(n)|$ will converge to the same value.

Since the logarithm is a monotonic function, we construct the following optimization problem, which expresses the above mentioned requirements:

$$\text{Min}_{\mathbf{h}}\, f(\mathbf{h}) = \log\left(\frac{f_u(\mathbf{h})}{f_d(\mathbf{h})}\right) = \log\left(\frac{\text{Max}\left[|\mathbf{g}_u|\right]}{\text{Max}\left[|\mathbf{g}_d|\right]}\right). \tag{4}$$

The gradient-based learning rule is

$$\mathbf{h}^{l+1} = \mathbf{h}^l - \mu\left(\frac{1}{f_u(\mathbf{h}^l)}\nabla_{\mathbf{h}} f_u(\mathbf{h}^l) - \frac{1}{f_d(\mathbf{h}^l)}\nabla_{\mathbf{h}} f_d(\mathbf{h}^l)\right). \tag{5}$$

Now suppose that $|g_d(n)|$ and $|g_u(n)|$ have distinct maxima at positions I_d and I_u, respectively. Then, with $f_d(\mathbf{h}) = |g_d(I_d)|$ and $f_u(\mathbf{h}) = |g_u(I_u)|$ for given $h(n)$, the corresponding gradients of $f_u(\mathbf{h})$ and $f_d(\mathbf{h})$ are as follows,

$$\nabla_{\mathbf{h}} f_u(\mathbf{h}) = \text{sign}\left[g_u(I_u)\right] w_u(I_u)\mathbf{C}_{I_u}^{\mathrm{T}} \tag{6}$$

and

$$\nabla_{\mathbf{h}} f_d(\mathbf{h}) = \text{sign}\left[g_d(I_d)\right] w_d(I_d)\mathbf{C}_{I_d}^{\mathrm{T}}, \tag{7}$$

Table 1. Learning rule based on infinity-norm criterion

Step 1: Set iteration index $l = 0$. Select a learning rate μ. Initialize the prefilter $\mathbf{h}^l \neq \mathbf{0}$.
Step 2: Compute: $\mathbf{g}^l = \mathbf{C}\mathbf{h}^l$, $\mathbf{g}_u^l = \mathrm{diag}(\mathbf{w}_u)\mathbf{g}^l$, $\mathbf{g}_d^l = \mathrm{diag}(\mathbf{w}_d)\mathbf{g}^l$; determine the positions of the maxima of $|\mathbf{g}_u^l|$ and $|\mathbf{g}_d^l|$, i.e., $f_u(\mathbf{h}^l) = \max(|\mathbf{g}_u^l|) = |g_u^l(I_u^l)|$ and $f_d(\mathbf{h}^l) = \max(|\mathbf{g}_d^l|) = |g_d^l(I_d^l)|$;
Step 3: Compute the gradients of $f_u(\mathbf{h}^l) = |g_u^l(I_u^l)|$ and $f_d(\mathbf{h}^l) = |g_d^l(I_d^l)|$ with respect of $\mathbf{h}^l$:
$\nabla_{\mathbf{h}^l} f_u(\mathbf{h}^l) = \mathrm{sign}\left[g_u(I_u^l)\right] w_u(I_u^l)\mathbf{C}_{I_u^l}^{\mathrm{T}}$,
$\nabla_{\mathbf{h}^l} f_d(\mathbf{h}^l) = \mathrm{sign}\left[g_d(I_d^l)\right] w_d(I_d^l)\mathbf{C}_{I_d^l}^{\mathrm{T}}$.
Step 4: Update $\mathbf{h}$:
$\mathbf{h}^{l+1} = \mathbf{h}^l - \mu\left(\frac{1}{f_u(\mathbf{h}^l)}\nabla_{\mathbf{h}^l} f_u(\mathbf{h}^l) - \frac{1}{f_d(\mathbf{h}^l)}\nabla_{\mathbf{h}^l} f_d(\mathbf{h}^l)\right)$.
Step 5: Set $l := l + 1$ and go to Step 2.

where $\mathbf{C}_{I_u}$ and $\mathbf{C}_{I_d}$ are the I_uth and I_dth rows of matrix $\mathbf{C}$, respectively. So the learning rule is given as follows,

$$\mathbf{h}^{l+1} = \mathbf{h}^l - \mu\left(\frac{1}{|g_u(I_u^l)|}\mathrm{sign}\left[g_u(I_u^l)\right] w_u(I_u^l)\mathbf{C}_{I_u^l}^{\mathrm{T}} - \frac{1}{|g_d(I_d^l)|}\mathrm{sign}\left[g_d(I_d^l)\right] w_d(I_d^l)\mathbf{C}_{I_d^l}^{\mathrm{T}}\right). \quad (8)$$

The implementation of algorithm (8) is shown in Table 1.

One of the advantages of the infinity-norm based algorithm is that the envelope of the unwanted part of the global impulse response $g(n)$ is exactly determined by the window function $w_u(n)$. So we can easily and exactly control the attenuating property of $g(n)$, and this enables us to remove audible reverberation and echoes by exploiting the D50 measure for intelligibility of speech during the prefilter design procedure.

To express the D50 measure, we define the two windows as follows [1]:

$$\mathbf{w}_u = [\underbrace{0, 0, ..., 0}_{N_1+N_2}, \underbrace{\mathbf{w}_0^{\mathrm{T}}}_{N_3}]^{\mathrm{T}} \quad (9)$$

$$\mathbf{w}_d = [\underbrace{0, 0, ..., 0}_{N_1}, \underbrace{1, 1, ..., 1}_{N_2}, \underbrace{0, 0, ..., 0}_{N_3}]^{\mathrm{T}} \quad (10)$$

where $N_1 = t_0 f_s$, $N_2 = t_d f_s$, and $N_3 = L_g - N_1 - N_2$ with f_s being the sampling rate, t_0 the time taken by the direct sound, and $t_d = 50$ms.

The window $\mathbf{w}_0$ is defined as

$$w_0(n) = 1 + \frac{a-1}{N_3 - 1}n, \quad (11)$$

where $n = 0, 1, 2, ..., N_3 - 1$ and usually $a \geq 1$, for a quick and uniform attenuation of $g_u(n)$. If we change N_2 and define $\mathbf{w}_0$ differently, we will get different windows for different purposes in shortening filter design.

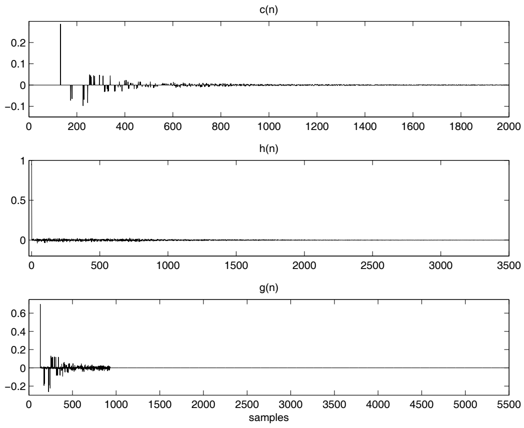

Fig. 1. The original filter $c(n)$ (top), the prefilter $h(n)$ (middle), and the global impulse response $g(n)$ (bottom).

Table 2. The relationship between the length of the desired part (t_d) and the attenuation of the unwanted part ($\min A_u$) for a prefilter length of $L_h = 3500$.

t_d(ms)	50	40	30	20
$\min A_u$(dB)	77.4	74.5	70.7	66.0

4. SIMULATIONS

A simulated room impulse response $c(n)$ with $L_c = 2000$ taps at a sampling frequency of $f_s = 16$kHz is used in the experiments. Informal listening tests show that strong echoes will be heard in such a room.

The basic parameters were selected as follows: $a = 1.2$, dynamic learning rate $\mu(i) = 10^{-11} - (10^{-11} - 10^{-13})i/L$ ($L = 2852000$ is the total iteration number), length of prefilter $L_h = 3500$. The windows defined in (9) and (10) were used in this experiment.

Results are shown in Figs. 1-3. In Fig. 1 we find that the desired part of the global impulse response $g(n)$ seems to be simply the truncated version of the RIR $c(n)$. This can also be seen in the frequency domain by comparing Fig. 3 and Fig. 4. This property prevents $g(n)$ from producing serious distortion in the frequency domain. The unwanted part of the global impulse response $g(n)$ is attenuated more than 77.4dB, so this part can not be heard from the output signal. Informal listening tests showed that the echoes are effectively suppressed.

The relationship between the length of the desired part ($t_d = N_2/f_s$) and the minimal attenuation of the unwanted part ($\min A_u$) is shown in Table 2. Listening tests showed that no echoes are heard if $30\mathrm{ms} \leq t_d \leq 50\mathrm{ms}$.

5. CONCLUSIONS

The approach for RIR shortening proposed in this paper is motivated by the idea of equiripple filter design. An infinity-

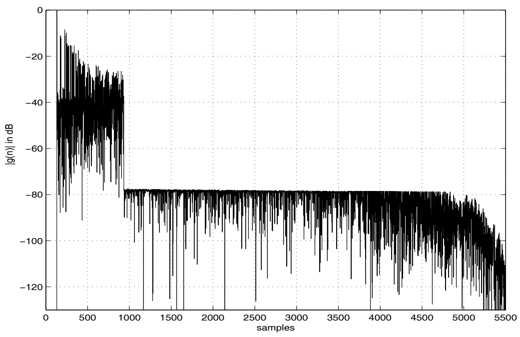

Fig. 2. The decay of the global impulse response $g(n)$.

norm criterion has been used instead of the traditionally applied least squares criterion. For a good perceived quality with relatively short prefilters, the D50 measure of speech intelligibility was exploited during the prefilter design. Experiments prove that the proposed method is feasible for RIR shortening and is superior to least-squares approaches. Further work will be directed toward shaping the overall impulse response with respect to psychoacoustic criteria other than the D50 measure.

6. REFERENCES

[1] M. Kallinger and A. Mertins, "Room impulse response shortening by channel shortening concepts," Proc. Asilomar Conference on Signals Systems and Computers, pp. 898-902, 2005.

[2] M. Kallinger and A. Mertins, "Impulse response shortening for acoustic listening room compensation," Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC), pp. 197-200, 2005.

[3] B. D. Radlović and R. A. Kennedy, "Nonminimum-phase equalization and its subjective importance in room acoustics," IEEE Trans. on Speech and Audio Processing, vol. 8, no. 6, pp. 728-736, 1998.

[4] M. Karjalainen and T. Paatero, "Equalization of loudspeaker and room responses using Kautz filter: direct least squares design," EURASIP Journal on Advances in Signal Processing, vol. 2007, pp.1-13, 2007.

[5] A. Maamar, I. Kale, A. Krukowski, and B. Daoud, "Partial equalization of non-minimum-phase impulse responses," EURASIP Journal on Applied Signal Processing, vol. 2006, pp. 1-8, 2006.

[6] L. D. Fielder, "Practical limits for room equalization," Proceedings on AES 111th Convention, pp.1-19, 2001.

[7] 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.

[8] J. N. Mourjopoulos, "Digital equalization of room

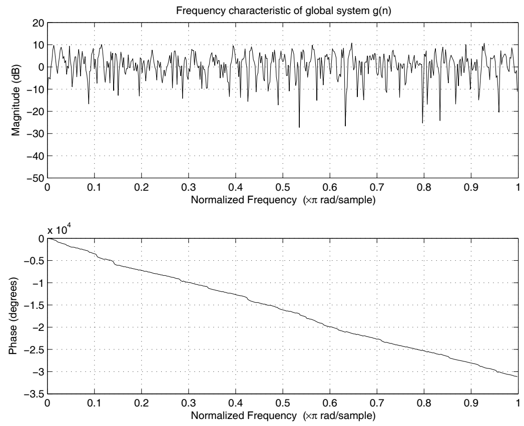

Fig. 3. The frequency characteristic of the shortened global system $g(n)$.

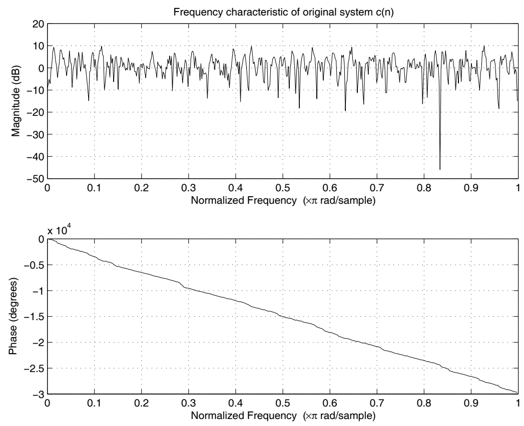

Fig. 4. The frequency characteristic of the original RIR $c(n)$.

acoustics," Journal of Audio Engineering Socity, vol. 42, no. 11, pp. 884-900, 1994.

[9] International Organization for Standardization (ISO), "ISO norm 3382: acoustics-measurement of the reverberation time of rooms with reference to other acoustical parameters".

[10] K. Zhou, J. C. Doyle, and K. Glover, Robust and Optimal Control, Prentice Hall Inc., New Jersey 1996.

[11] A. V. Oppenheim and R. W. Schafer, Discrete-Time Signal Processing, Prentice Hall International, Inc., 1999.

[12] G. Arslan, B. L. Evans, and S. Kiaei, "Equalization for discrete multitone transceivers to maximize bit rate," IEEE Trans. Signal Processing, vol. 49, no.12, pp. 3123-3135, 2001.

[13] R. K. Martin, M. Ding, B. L. Evans, and C. R. Johnson Jr., "Efficient channel shortening equalizer design," EURASIP J. on Applied Signal Processing, vol. 13, pp. 1279-1290, 2003.