Room Impulse Response Shortening/Reshaping With Infinity- and p-Norm Optimization

Alfred Mertins, *Senior Member, IEEE*, Tiemin Mei, and Markus Kallinger, *Member, IEEE*

***Abstract*—The purpose of room impulse response (RIR) shortening and reshaping is usually 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 an alternative, but mathematically equivalent setting, one may aim to postfilter a recorded microphone signal to remove audible echoes. While least-squares methods have mainly been used for the design of shortening/reshaping filters for RIRs until now, we propose to use the infinity- or p-norm as optimization criteria. In our method, design errors will be uniformly distributed over the entire temporal range of the shortened/reshaped global impulse response. In addition, the psychoacoustic property of masking effects is considered during the filter design, which makes it possible to significantly reduce the filter length, compared to standard approaches, without affecting the perceived performance.**

***Index Terms*—Infinity-norm, optimization, p-norm, reshaping, room impulse response (RIR), shortening.**

I. Introduction

FOR the enhancement of speech intelligibility in reverberant rooms and for new applications in audio–visual communications and virtual acoustics, a suitable preprocessing of loudspeaker signals is needed to compensate room reverberation, namely, the listening-room-compensation (LRC) or room reverberation compensation [1]–[3]. Similarly, for improving the quality of far-field microphone recordings, a postfiltering stage may be introduced for the received signals. Both problems are mathematically equivalent, and for the sake of conciseness, we describe our method for the LRC problem.

Room-reverberation compensation is somewhat different from channel equalization. For channel equalization, the aim is to exactly recover the original signal from the received one and thus to invert the channel [4]. Room-reverberation compensation, on the other hand, only needs to compensate the channel so that signals are perceived without reverberation. In other words, it would be sufficient to equalize a room impulse response (RIR) only partially [5], [6], so that all audible echoes are removed and the inaudible ones remain. Such a relaxed requirement may greatly alleviate the pressure of designing a compensation system.

To shorten or reshape a room impulse response, the required filter can be designed in different ways and according to different criteria. For instance, for the least-squares plus postprocessing method in [1], the performance depends closely on the postprocessing filter. Another example is the approach in [6] for the design of minimum-phase inverse filters, which is based on a homomorphic transformation and a relocation of the dominant poles. 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, one should not only consider standard optimality criteria such as least squares, but also take the psychoacoustic properties of the human auditory system into account and aim for prefilters that are optimal in the sense of giving best intelligibility at the lowest implementation cost. There are different known appropriate psychoacoustic criteria [7], [8], [10]. For example, investigations into the properties of the human auditory system have shown that echoes will not be heard when they are lower than a masking limit that is induced by the direct sound [7], [8]. This is known as the so-called temporal masking effect of the human auditory system. While echoes are inaudible when they are just under the masking limit, the human auditory system is very sensitive to those echoes that are beyond the masking limit [7]. The deviation from the masking threshold can therefore be used to define a criterion for room reverberation compensation, as further described in Section III of this paper. While the masking effects of the human auditory system are signal dependent in general [19], [20] and call for a signal-dependent filtering process to achieve ultimate performance, we concentrate on linear, signal-independent filtering of the RIR, where our optimality criterion is based on an average masking curve that has been found a good compromise between masking curves obtained for various signals [7]. Thus, similar to mechanical damping of a room, we essentially shorten the effective room reverberation, and in doing so, we achieve suppression of otherwise audible room reflections for a class of signals. Another well-known criterion, also described in detail in Section III, 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. According to this measure, the intelligibility is guaranteed if the RIR is shortened so

Manuscript received November 07, 2008; revised May 22, 2009. First published June 19, 2009; current version published November 13, 2009. This work was supported by the German Research Foundation under Grant ME1170/3-1. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Stephen J. Elliott.

A. Mertins is with the Institute for Signal Processing, University of Lübeck, Lübeck 23538, Germany (e-mail: mertins@isip.uni-luebeck.de).

T. Mei is with the Institute for Signal Processing, University of Lübeck, Lübeck 23538, Germany, on leave from the School of Information Science and Engineering, Shenyang Ligong University, Shenyang 110168, China (e-mail: mei@isip.uni-luebeck.de; meitiemin@163.com).

M. Kallinger was with the Department of Mathematics and Science, University of Oldenburg, D-26129 Oldenburg, Germany. He is now with the Fraunhofer Institute for Integrated Circuits IIS, 91058 Erlangen, Germany.

Digital Object Identifier 10.1109/TASL.2009.2025789

that the energy is concentrated within 50 ms after the first impulse of the RIR.

A problem associated with listening-room compensation is spatial robustness. In practice, the RIR will change significantly when the position of the source or the receiver is changed. Moreover, for a good performance at the design position, the prefilter usually has a high length, which further decreases spatial robustness. The authors in [10] proposed to use an all-pole RIR model and vector quantization to overcome these problems by designing prefilters that are optimal for clusters of RIRs. However, for the sake of conciseness, we concentrate on the shortening/reshaping of single, given RIRs and leave the robustness issue associated with changing RIRs for later investigations.

The least-squares method is the most widely used approach in optimization—simplicity and linearity are its advantages. However, it has drawbacks too, namely, non-uniformity. For instance, in traditional least-squares equalization for LRC, the squared errors are distributed non-uniformly along the time axis, which typically results in audible late diffuse echoes as reported in [1]. In this paper, we exploit the infinity- and p-norms and combine them with the psychoacoustic properties of the human auditory system to define suitable optimization criteria. Although this will lead to high nonlinearity in the optimization process, it also leads to high uniformity. With this method, we can explicitly control the error distribution along the time course of the global impulse response consisting of RIR and prefilter, so that no late diffuse reverberation is produced, and the echoes are controlled exactly.

The following notations are used in this paper: bold face capital letters are used for matrices, bold face small letters are used for vectors, and regular small letters are used for simple variables and series. Superscripts * and $^{\mathrm{T}}$ are the conjugation and matrix transpose operators, respectively. The expression $\|\cdot\|_p$ is the p-norm operator, and $|\cdot|$ is the absolute value operator—if its variable is a vector, it acts componentwise. The operator $\mathrm{diag}[\cdot]$ produces a diagonal matrix made up of the elements of the input vector. $\mathrm{Max}[\cdot]$ gives out the maximum component of a vector variable, and $\mathrm{sign}[\cdot]$ produces a sign vector of its input vector variable. The asterisk $*$ denotes the convolution operator.

The paper is organized as follows. The problem is described in Section II, then the new approaches based on infinity and p-norms are developed in Section III. Simulations are presented in Section IV. Section V gives some conclusions and closes the paper.

II. 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:

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

with $\mathbf{C}$ being an 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 $g(n)$ not only attenuate faster than the impulse response of the room but also allows it to satisfy certain psychoacoustic conditions so that there will be no audible echoes for a large class of signals.

For filter shortening and reshaping, we use two window functions $w_d(n)$ and $w_u(n)$ to derive a desired part $g_d(n) = w_d(n)g(n)$ and an unwanted part $g_u(n) = w_u(n)g(n)$ from the global impulse response $g(n)$. For shortening, the windows $w_d(n)$ and $w_u(n)$ show no overlap, while they may have significant overlap in the case of reshaping. 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 $g_u(n)$ should be minimized while the energy of $g_d(n)$ is kept constant.

A conventional approach is to optimize $h(n)$ under the least-squares error criterion (i.e., the 2-norm), that is [1], [11]

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

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

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

where

$$\mathbf{B} = \mathbf{C}^{\mathrm{T}}\mathrm{diag}[\mathbf{w}_d]^{\mathrm{T}}\mathrm{diag}[\mathbf{w}_d]\mathbf{C}$$

and

$$\mathbf{A} = \mathbf{C}^{\mathrm{T}}\mathrm{diag}[\mathbf{w}_u]^{\mathrm{T}}\mathrm{diag}[\mathbf{w}_u]\mathbf{C}.$$

In [18], 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, as can be seen from Figs. 1–3, a prefilter $\mathbf{h}_{\mathrm{opt}}$ that is optimal in the least-squares sense (2) will usually make great distortion in the frequency domain, and, in addition, the time-course of $|g(n)|$ will typically cause obvious late diffuse echoes. Although some measures have been taken to overcome such drawbacks [1], further improvement is needed in practice.

As an alternative to least-squares techniques, the infinity- and p-norm criteria are often used for robust estimation and robust control-system design [17], as they better allow one to influence the error behavior in detail. Therefore, in the next section, we combine the infinity- and p-norm criteria with the properties of the human auditory system in order to explicitly control the perceived quality during the prefilter design.

III. 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 $w_u(n)$

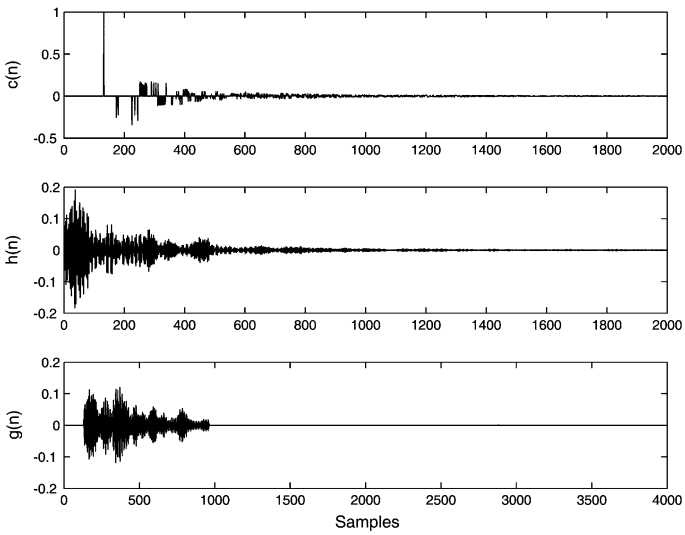

Fig. 1. Least-squares filter shortening. The original filter $c(n)$ (top), the shortening filter $h(n)$ (middle), and the global impulse response $g(n)$ (bottom).

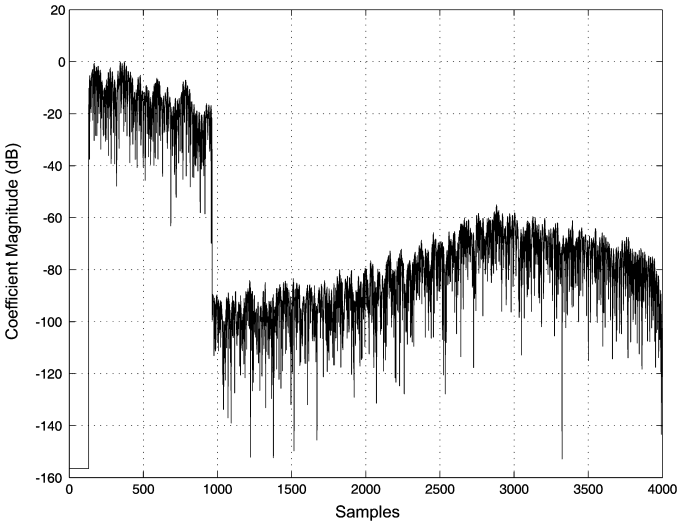

Fig. 2. Least-squares filter shortening. The decay of the global impulse response $g(n)$ designed with the D50 measure.

will be helpful for improving this solution, but the more important point is that we should look for some best-suited optimization criteria.

The idea in this paper is motivated by the equiripple filter design [12]. 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 error is minimized in frequency domain, i.e., the so-called minimax or Chebyshev criterion. In our case, however, the measure of optimality is somewhat different from the equiripple criterion in filter design. For the optimization of prefilters, we would like to minimize the norm of the unwanted part $g_u(n)$ while keeping the norm of the desired part $g_d(n)$ as large as possible, where as norm we either define the infinity- or the p-norm. With properly designed windows it will then be possible to force the shortened or reshaped global impulse response to approximate a desired decaying behavior.

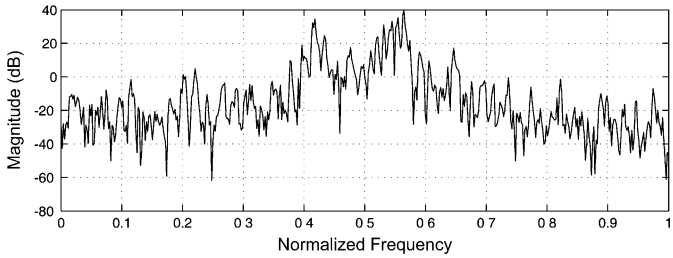

Fig. 3. Least-squares filter shortening: The magnitude frequency response of the shortened global system designed via the D50 measure.

A. Algorithm Based on Infinity-Norm

Using the fact that the logarithm is a monotonic function, we define the following optimization problem:

$$\mathrm{MIN}_{\mathbf{h}} : f(\mathbf{h}) = \log\left(\frac{f_u(\mathbf{h})}{f_d(\mathbf{h})}\right) = \log\left(\frac{\mathrm{Max}[|\mathbf{g}_u|]}{\mathrm{Max}[|\mathbf{g}_d|]}\right) \tag{4}$$

where $f_d(\mathbf{h}) = ||\mathbf{g}_d||_\infty = \mathrm{Max}[|\mathbf{g}_d|]$ with $\mathbf{g}_d = \mathrm{diag}[\mathbf{w}_d]\mathbf{Ch}$ is the infinity-norm of the desired part, and $f_u(\mathbf{h}) = ||\mathbf{g}_u||_\infty = \mathrm{Max}[|\mathbf{g}_u|]$ with $\mathbf{g}_u = \mathrm{diag}[\mathbf{w}_u]\mathbf{Ch}$ is the infinity-norm of the unwanted part.

The minimization of $f(\mathbf{h})$ will result in the tradeoff between the minimization of $f_u(\mathbf{h})$ and, at the same time, the maximization of $f_d(\mathbf{h})$. Interestingly, both of these criteria have quite intuitive interpretations: the maximization of $f_d(\mathbf{h})$ will lead to the most-possibly flat frequency-domain characteristic of the global impulse response, because, when one tap of $|g_d(n)|$ is maximized, then the other samples of $|g_d(n)|$ will become small, and $g_d(n)$ is dominated by a single tap. On the other hand, the minimization of $f_u(\mathbf{h})$ will result in the most-uniform distribution of the errors along the time-course of $g(n)$, because all of the samples of the unwanted part $|g_u(n)|$ will converge to almost the same value.

The gradient-based learning rule for the optimality criterion defined in (4) is given by

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

where μ is a small step size.

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:

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

and

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

where $\mathbf{C}_{I_u}$ and $\mathbf{C}_{I_d}$ are the I_uth and I_dth rows of matrix $\mathbf{C}$, respectively. Taking into account that $f_d(\mathbf{h}) = |g_d(I_d)|$ and $f_u(\mathbf{h}) = |g_u(I_u)|$, and furthermore that $g_d(n) = w_d(n)g(n)$ and $g_u(n) = w_u(n)g(n)$, the learning rule can be eventually written as follows:

$$\mathbf{h}^{l+1} = \mathbf{h}^l - \mu\left(\frac{1}{g\left(I_u^l\right)}\mathbf{C}_{I_u^l}^{\mathrm{T}} - \frac{1}{g\left(I_d^l\right)}\mathbf{C}_{I_d^l}^{\mathrm{T}}\right). \tag{8}$$

TABLE I
LEARNING RULE BASED ON INFINITY-NORM CRITERION

Step 1: Set iteration index $l = 0$. Select a learning rate μ. Initialize the prefilter $\mathbf{h}^l = [0.01, 0, ..., 0]^{\mathrm{T}}$.
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: Update $\mathbf{h}$: $\mathbf{h}^{l+1} = \mathbf{h}^l - \mu\left(\frac{1}{g(I_u^l)}\mathbf{C}_{I_u^l}^{\mathrm{T}} - \frac{1}{g(I_d^l)}\mathbf{C}_{I_d^l}^{\mathrm{T}}\right)$.
Step 4: Set $l := l + 1$ and go to Step 2.

TABLE II
LEARNING RULE BASED ON p-NORM CRITERION

Step 1: Set $l = 0$. Select a learning rate μ and the FFT block size: $L_0 = 2^\alpha \geq L_g$; compute $C(k) = \mathrm{FFT}[c(n), L_0]$; initialize the prefilter $\mathbf{h}^l = [0.01, 0, ..., 0]^{\mathrm{T}}$.
Step 2: Compute $H^l(k) = \mathrm{FFT}[h^l(n), L_0]$;
$\mathbf{g}_L = \mathrm{IFFT}[C(k)H^l(k)]$, $\mathbf{g}_0 = [g_L(0), g_L(1), ..., g_L(L_g - 1)]^{\mathrm{T}}$;
$\mathbf{g}_u = \mathrm{diag}[\mathbf{w}_u]\mathbf{g}_0$, $\mathbf{g}_d = \mathrm{diag}[\mathbf{w}_d]\mathbf{g}_0$.
Step 3: $\phi_{f_u} = \sum_{k=0}^{L_g-1} |g_u(k)|^{p_u}$, $\phi_{f_d} = \sum_{k=0}^{L_g-1} |g_d(k)|^{p_d}$.
Step 4: $\mathbf{b}_u = \mathrm{diag}\,[\mathrm{sign}[\mathbf{g}_u]]\,\mathrm{diag}[\mathbf{w}_u]|\mathbf{g}_u|^{(p_u-1)}$, $\mathbf{b}_d = \mathrm{diag}\,[\mathrm{sign}[\mathbf{g}_d]]\,\mathrm{diag}[\mathbf{w}_d]|\mathbf{g}_d|^{(p_d-1)}$; $B_u(k) = \mathrm{FFT}[\mathbf{b}_u, L_0]$, $B_d(k) = \mathrm{FFT}[\mathbf{b}_d, L_0]$.
Step 5: $\mathbf{a}_u = \mathrm{IFFT}[C^*(k)B_u(k)]$, $\mathbf{a}_d = \mathrm{IFFT}[C^*(k)B_d(k)]$;
$\mathbf{a}_{u_0} = [a_u(0), a_u(1), ..., a_u(L_h - 1)]^{\mathrm{T}}$,
$\mathbf{a}_{d_0} = [a_d(0), a_d(1), ..., a_d(L_h - 1)]^{\mathrm{T}}$.
Step 6: Update $\mathbf{h}$: $\mathbf{h}^{l+1} = \mathbf{h}^l - \mu\left(\frac{1}{\phi_{f_u}}\mathbf{a}_{u_0} - \frac{1}{\phi_{f_d}}\mathbf{a}_{d_0}\right)$.
Step 7: Go to Step 2.

If the maxima of $|g_d(n)|$ and $|g_u(n)|$ occur multiple times at different positions, then (8) should be modified into the following multiple-maxima learning rule:

$$\mathbf{h}^{l+1} = \mathbf{h}^l - \mu\left(\frac{1}{N_u^l}\sum_{i=1}^{N_u^l}\frac{1}{g\left(I_u^l(i)\right)}\mathbf{C}_{I_u^l(i)}^{\mathrm{T}} - \frac{1}{N_d^l}\sum_{i=1}^{N_d^l}\frac{1}{g\left(I_d^l(i)\right)}\mathbf{C}_{I_d^l(i)}^{\mathrm{T}}\right) \quad (9)$$

where $|g_u(I_u^l(1))| = |g_u(I_u^l(2))| = \cdots = |g_u(I_u^l(N_u))|$ and $|g_d(I_d^l(1))| = |g_d(I_d^l(2))| = \cdots = |g_d(I_d^l(N_d))|$ are the maxima of $|g_u(i)|$ and $|g_d(i)|$, respectively.

The implementation of algorithm (8) is presented in Table I. For the implementation of algorithm (9), Steps 2 and 3 in Table I need to be extended accordingly.

The design of the window functions will be addressed in Section III-D, but it should already be mentioned at this point that 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 inverse of the window function $w_u(n)$. So we can easily and exactly control the attenuation behavior of $g(n)$, and this enables us to remove audible reverberation and echoes by exploiting the auditory-masking property during the prefilter design procedure.

B. Algorithm Based on P-Norm

For more flexibility in implementation than the infinity-norm based algorithm in III-A, we exploit the p-norm for defining the objective function. The corresponding optimization problem is given by

$$\mathrm{MIN}_{\mathbf{h}} : f(\mathbf{h}) = \log\left(\frac{f_u(\mathbf{h})}{f_d(\mathbf{h})}\right) \quad (10)$$

with

$$f_d(\mathbf{h}) = \|\mathbf{g}_d\|_{p_d} = \left(\sum_{k=0}^{L_g-1}|g_d(k)|^{p_d}\right)^{\frac{1}{p_d}}$$

and

$$f_u(\mathbf{h}) = \|\mathbf{g}_u\|_{p_u} = \left(\sum_{k=0}^{L_g-1}|g_u(k)|^{p_u}\right)^{\frac{1}{p_u}}$$

where p_u and p_d are integers. The learning rule reads

$$\mathbf{h}^{l+1} = \mathbf{h}^l - \mu\nabla_{\mathbf{h}} f(\mathbf{h}^l). \quad (11)$$

First, the gradients $\nabla_{\mathbf{h}} f_d(\mathbf{h})$ and $\nabla_{\mathbf{h}} f_u(\mathbf{h})$ are calculated as

$$\nabla_{\mathbf{h}} f_d(\mathbf{h}) = \left(\sum_{k=0}^{L_g-1}|g_d(k)|^{p_d}\right)^{\frac{1}{p_d}-1}\mathbf{C}^{\mathrm{T}}\mathbf{b}_d \quad (12)$$

and

$$\nabla_{\mathbf{h}} f_u(\mathbf{h}) = \left(\sum_{k=0}^{L_g-1}|g_u(k)|^{p_u}\right)^{\frac{1}{p_u}-1}\mathbf{C}^{\mathrm{T}}\mathbf{b}_u \quad (13)$$

with

$$\mathbf{b}_d = \mathrm{diag}\,[\mathrm{sign}[\mathbf{g}_d]]\,\mathrm{diag}[\mathbf{w}_d]|\mathbf{g}_d|^{(p_d-1)},$$
$$\mathbf{b}_u = \mathrm{diag}[\mathrm{sign}[\mathbf{g}_u]]\mathrm{diag}[\mathbf{w}_u]|\mathbf{g}_u|^{(p_u-1)}.$$

The gradient of $f(\mathbf{h})$ to be used in (11) reads

$$\nabla_{\mathbf{h}} f(\mathbf{h}) = \frac{1}{f_u(h)}\nabla_{\mathbf{h}} f_u(\mathbf{h}) - \frac{1}{f_d(h)}\nabla_{\mathbf{h}} f_d(\mathbf{h}) = \frac{1}{\phi_{f_u}}\mathbf{C}^{\mathrm{T}}\mathbf{b}_u - \frac{1}{\phi_{f_d}}\mathbf{C}^{\mathrm{T}}\mathbf{b}_d \quad (14)$$

where $\phi_{f_u} = \sum_{k=0}^{L_g-1}|g_u(k)|^{p_u}$ and $\phi_{f_d} = \sum_{k=0}^{L_g-1}|g_d(k)|^{p_d}$.

In (14), the computational burden is located at the factors $\mathbf{C}^{\mathrm{T}}\mathbf{b}_u$ and $\mathbf{C}^{\mathrm{T}}\mathbf{b}_d$, but fortunately, because of the special structure of the convolution matrix $\mathbf{C}$, they can be determined in the frequency domain with the fast Fourier transform (FFT) and the inverse fast Fourier transform (IFFT). This is shown in the following: let $C(\cdot) = \mathrm{FFT}[c(\cdot), L_0]$ and $B_d(\cdot) = \mathrm{FFT}[b_d(\cdot), L_0]$, where $L_0 \geq L_g$ is the FFT size. We then compute $\mathbf{a}_d = \mathrm{IFFT}[C^*(\cdot)B_d(\cdot)]$, which is the desired result:

$$\mathbf{C}^{\mathrm{T}}\mathbf{b}_d = [a_d(0), a_d(1), \ldots, a_d(L_h - 1)]^{\mathrm{T}}. \quad (15)$$

In the same way, $\mathbf{C}^{\mathrm{T}}\mathbf{b}_u$ will be computed.

A summary of the p-norm based algorithm is presented in Table II.

For $p_u = p_d = 2$, the p-norm based algorithm degenerates to the least mean squares algorithm like that in [1]. However, in our case, p_u and p_d are set bigger than 2. In our experiments,

we used p_u and p_d equal to 10 and 20 (shortening) and p_u and p_d equal to 20 and 10 (reshaping), respectively.

The greater the parameters p_u and p_d are, the closer the p-norm gets to the infinity-norm, but the behavior of the two learning rules (8) and (11) still differs significantly in terms of computational efficiency and convergence speed. Definitely, the infinity-norm algorithm is computationally more efficient than the p-norm algorithm: however, it converges slow during later iterations. For the infinity-norm algorithm (8), the formula for the gradient of the objective function depends only on the first maximum of $|g_d(n)|$ and the first maximum of $|g_u(n)|$ although it is possible that the maxima of $|g_d(n)|$ and $|g_u(n)|$ occur multiple times. Therefore, the updated $h(n)$ will make the first maxima of $|g_d(n)|$ and $|g_u(n)|$ change in the desired directions, but it is not clear what will happen to the other maxima of $|g_d(n)|$ and $|g_u(n)|$. Taking this factor into account, the modified version (9) will accelerate the converging speed, but, although the multiplicity of the maxima of $g_d(i)$ will converge to $N_d = 1$, the number N_u will become bigger and bigger as the iteration moves on, which will result in increasingly heavy computational burden.

For the p-norm algorithm, the behavior will be different. Although it is not computationally as efficient as the infinity-norm algorithm (8) for a single maximum, it is, due to the use of the FFT, computationally more efficient than the infinity-norm algorithm (9) for multiple equal maxima, so that, overall, it converges faster than the infinity-norm algorithm.

C. A Combined p- and Infinity-Norm Criterion

A good compromise between the p- and infinity-norm criteria is given when the infinity norm criterion is considered for the desired part, and the p-norm criterion is used for the unwanted part. We obtain the following learning rule:

$$\mathbf{h}^{l+1} = \mathbf{h}^{l} - \mu\left(\frac{1}{\phi_{f_u}}\mathbf{C}^{\mathrm{T}}\mathbf{b}_u - \frac{1}{g(I_d)}\mathbf{C}_{I_d}^{\mathrm{T}}\right) \tag{16}$$

where ϕ_{f_u} and $\mathbf{b}_u$ are the same as those defined in (14). The advantage of (16) is that it will save almost half of the computations of the p-norm algorithm while keeping the convergence speed. Because the maximum of the desired part is maximized, the position I_d of the maximum can be kept fixed during optimization, and the computation of the gradient of the desired part takes almost no time.

D. Design of the Window Functions

The energy-decay property of the global system's impulse response and its frequency response clearly depend on the selected window functions, so that the window design plays an important role in the entire shortening/reshaping filter design process. Importantly, the global impulse response should decay in such a way that there will be no audible echoes, which means that the reverberation should be masked by the direct sound through the forward-masking effect of the human auditory system. Similarly, the frequency-domain characteristics of the overall system should not change the perceived timbre in a significant way.

The forward-masking effects of the human auditory system in real-world acoustic environments depend on both the signals under consideration and the room characteristics as described by the room impulse response [20], [19]. Taking the results of various experiments [13]–[16] into account, the [7] and [8] give a rough criterion for the assessment of the average audibility of echoes: first, the loudness of sound components is determined by convolution with a temporal integration function; the backward-masking (pre-masking) limit is set to 15 ms; the forward masking (post masking) then acts like simultaneous masking for the first 4 ms after the initial direct sound impulse and then falls off by 35 dB/decade. Clearly, the masking threshold computed according to the above rules does not hold exactly for every signal at hand, but it shows a compromise of masking thresholds determined for various signals ranging from bursts of pulses and white noise [13] over shaped sinewave bursts [16] to clicks [14].

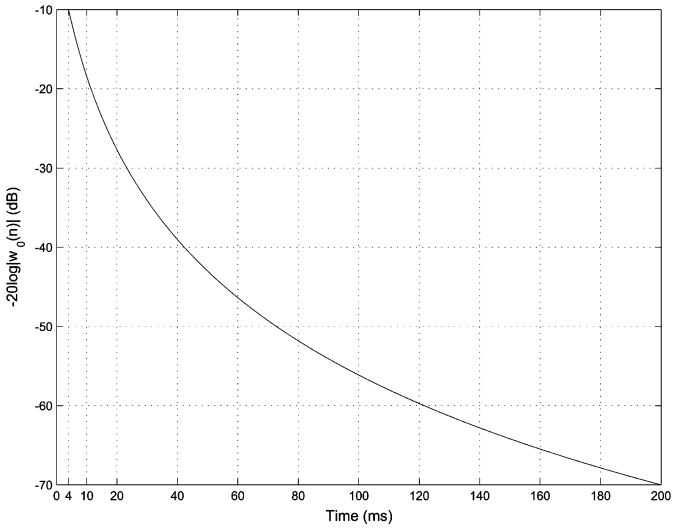

Fig. 4. Logarithmic reciprocal of window function $w_0(n)$. It approximately falls off with -35 dB/decade which represents the compromise masking limit of the human auditory system from [7].

Based on the above mentioned compromise rules, we define the two window functions as follows:

$$\mathbf{w}_u = [\underbrace{0, 0, \ldots, 0}_{N_1+N_2}, \underbrace{\mathbf{w}_0^{\mathrm{T}}}_{N_3}]^{\mathrm{T}} \tag{17}$$

$$\mathbf{w}_d = [\underbrace{0, 0, \ldots, 0}_{N_1}, \underbrace{1, 1, \ldots, 1}_{N_2}, \underbrace{0, 0, \ldots, 0}_{N_3}]^{\mathrm{T}} \tag{18}$$

where $N_1 = t_0 f_s$, $N_2 = 0.004 f_s$, and $N_3 = L_g - N_1 - N_2$ with f_s being the sampling frequency and t_0 being the time taken by the direct sound. The window $\mathbf{w}_0$ is defined as

$$w_0(n) = 10^{\frac{3}{\log(N_0/(N_1+N_2))}\log\left(\frac{n}{N_1+N_2}\right)+0.5} \tag{19}$$

with $N_0 = (0.2 + t_0) f_s$ and time index n ranging from $N_1 + N_2 + 1$ to $L_g - 1$.

The function $w_0(n)$ has the property that its reciprocal approximately falls off with -35 dB/decade, so that it represents the compromise masking limit of the human auditory system as determined in [7]. This is shown in Fig. 4. Its decay is -10 dB at 4 ms and then it decays exponentially to -70 dB at 200 ms.

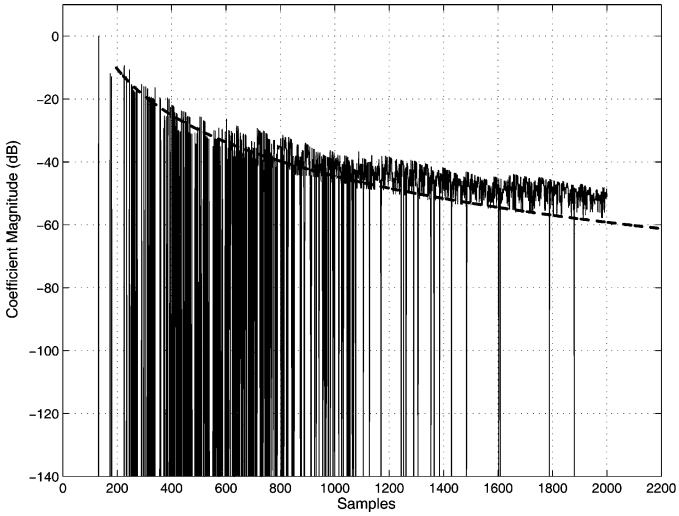

Fig. 5. Decay of the original RIR $c(n)$. In addition, the dashed curve is the logarithmic reciprocal of the masking limit window function $w_0(n)$. When $c(n)$ exceeds the average masking limit it means that for many signals echoes will be heard.

If we use the D50 measure for intelligibility of speech [1] instead of the previously described masking threshold, we define the two windows as

$$\mathbf{w}_u = [\underbrace{0,0,\ldots,0}_{N_1+N_2},\underbrace{\mathbf{w}_0^{\mathrm{T}}}_{N_3}]^{\mathrm{T}} \tag{20}$$

$$\mathbf{w}_d = [\underbrace{0,0,\ldots,0}_{N_1},\underbrace{1,1,\ldots,1}_{N_2},\underbrace{0,0,\ldots,0}_{N_3}]^{\mathrm{T}} \tag{21}$$

where N_1 is the same as in (18). For the other parameters, we have $N_2 = 0.05\, f_s$ and $N_3 = L_g - N_1 - N_2$. The window $\mathbf{w}_0$ is defined as

$$w_0(n) = 1 + \frac{a-1}{N_3 - 1} n \tag{22}$$

where $n = 0, 1, 2, \ldots, N_3 - 1$ and usually $a > 1$, 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 filter design.

IV. Simulations

A simulated room impulse response [21] with $L_c = 2000$ taps at a sampling frequency of $f_s = 16$ kHz was used in a first set of experiments. The time- and frequency-domain characteristics of $c(n)$ are shown in Figs. 5 and 6, respectively. The impulse response $c(n)$ itself is depicted in the top panel of Fig. 7. The room impulse response shows that the room has medium reverberation with a reverberation time of approximately 200 ms.

For removing the echoes, we insert an elaborately designed prefilter between the sound source and the playback loudspeaker, where the prefilter performs either shortening of the global impulse response to satisfy the D50 measure, or it tries to make the global impulse response attenuate quick enough to stay under the desired decay curve which represents the compromise masking limit as described above. Design

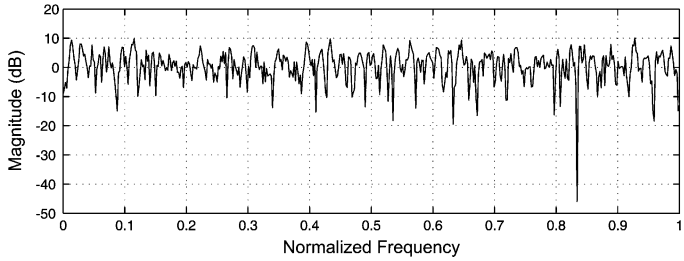

Fig. 6. Magnitude frequency response of the RIR $c(n)$.

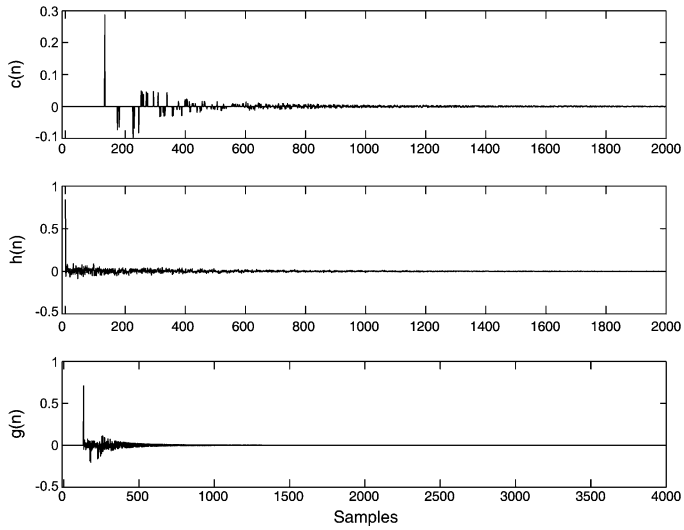

Fig. 7. Original filter $c(n)$ (top), the reshaping filter $h(n)$ (middle), and the global impulse response $g(n)$ (bottom). Parameters: masking-limit measure, $p_u = 20, p_d = 10$, learning rate $\mu = 10^{-6}$, length of prefilter $L_h = 2000$.

experiments have been carried out based on all of the three algorithms. The simulations showed that the infinity-norm based algorithm does not converge as fast as the algorithms based on p-norm. Both p-norm based algorithms required the same number of iterations for convergence, but the one of Section III-C is computationally more efficient than the one in Section III-B. The obtained final results after convergence were almost identical for all of the three algorithms. Therefore, an explicit comparison is not carried out here, and the following results are all given for the p-norm based algorithm.

A. Masking-Limit Measure

The basic parameters were selected as follows: $p_u = 20, p_d = 10$, learning rate $\mu = 10^{-6}$, length of prefilter $L_h = 2000$, number of iterations: 3×10^5. The windows defined in (17) and (18) were used in this experiment.

The RIR $c(n)$, the optimal prefilter $h(n)$, and the global impulse response are depicted in Fig. 7. The decay characteristic of the global impulse response $g(n)$ is shown in Fig. 8. We can observe that it is exactly controlled by the windows. The dashed curve in Fig. 8 is the logarithmic reciprocal of the function $w_0(n)$ representing the compromise masking limit. As one can see, the envelope of the logarithm of the global impulse response is almost perfectly controlled by the function $w_0(n)$. Fig. 8 shows that the reshaped RIR $g(n)$ with a prefilter of length

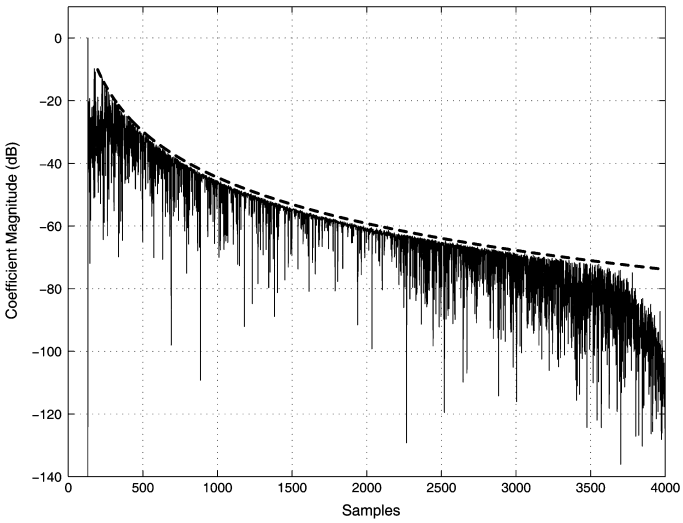

Fig. 8. Decay of the global impulse response $g(n)$. The dashed curve is the logarithmic reciprocal of the window function $w_0(n)$.

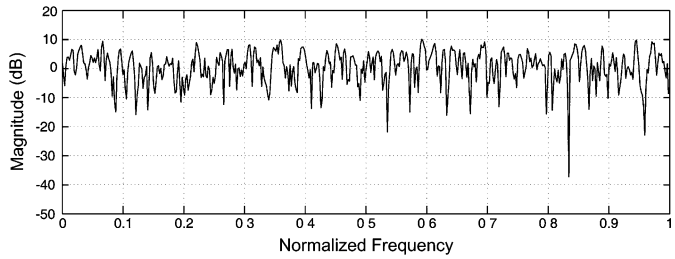

Fig. 9. Magnitude frequency response of the reshaped global system $g(n)$ designed with the masking-limit measure.

2000 samples is just under the average masking limit. Results for other filter lengths will be discussed next.

Not only the time-domain but also the frequency-domain characteristics are important in filter reshaping, as serious frequency-domain distortion should be avoided. For the considered design example, the frequency response of the global system is depicted in Fig. 9. The comparison of Figs. 6 and 9 shows that the frequency response of the reshaped system has some necessary deviation from the original system. However, it is not so serious that it damages the quality of the output signals. Numerical results regarding the frequency-domain deviations will be described in more detail in Section IV-C.

Finally, we have studied the relationship between the length of the prefilter impulse response and the attenuation degree of the global impulse response. Results are depicted in Fig. 10. They show that the decay behavior of the global impulse response $g(n)$ is controlled by the window $w_0(n)$, but the attenuation degree of $g(n)$ is controlled by the length of the prefilter $h(n)$. For a short prefilter, the decay will be above the average masking limit by a constant amount. With increasing prefilter length, the attenuation degree of $g(n)$ is increased. In addition, the prefilter impulse response has to be long enough to force the global impulse response to stay under the temporal masking limit. The deviations between the impulse-response decay and the desired one are reported in Table III.

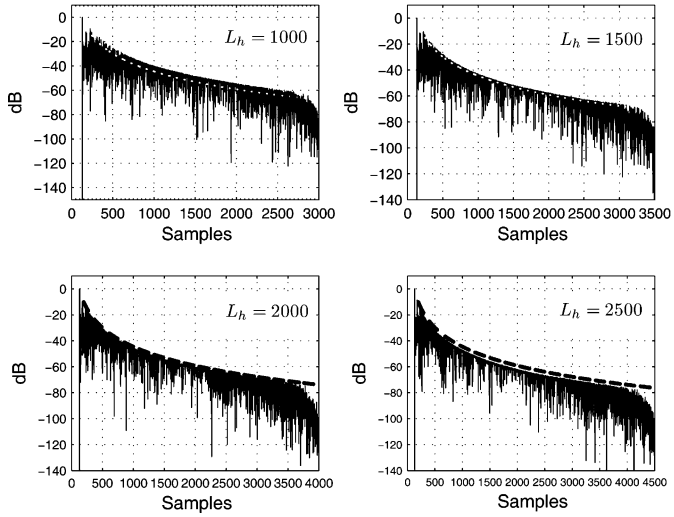

Fig. 10. Relationship between the prefilter length and the decay of the global impulse response. The dashed curve is the logarithmic reciprocal of function $w_0(n)$.

TABLE III
RELATIONSHIP BETWEEN THE LENGTH OF THE PREFILTER (L_h) AND THE TIME-COURSE DEVIATION BETWEEN THE SHAPED RIR AND THE DESIRED DECAY CURVE

L_h	2500	2000	1500	1000
Δ (dB)	-1.86	-0.37	2.22	5.30

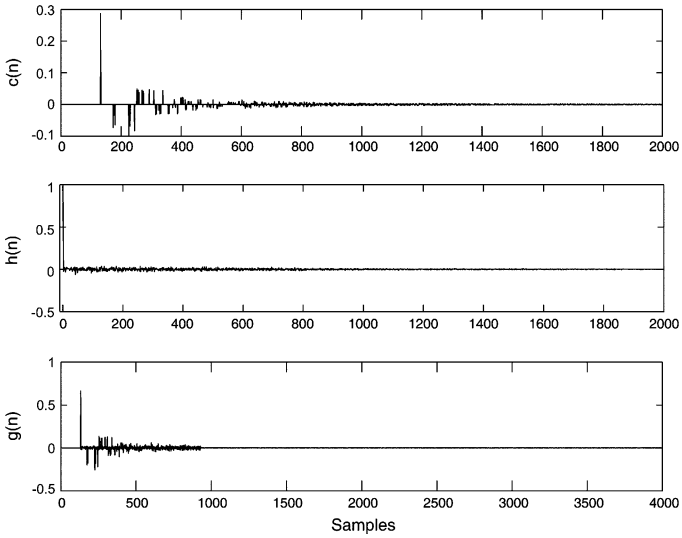

Fig. 11. Original filter $c(n)$ (top), the shortening filter $h(n)$ (middle), and the global impulse response $g(n)$ (bottom). Parameters: D50 measure, $a = 2.0, p_u = 10, p_d = 20$, learning rate $\mu = 2.5 \times 10^{-11}$, length of prefilter $L_h = 2000$.

B. D50 Measure

In this experiment, the basic parameters were selected as follows: $a = 2.0, p_u = 10, p_d = 20$, learning rate $\mu = 2.5 \times 10^{-11}$, length of prefilter $L_h = 2000$, number of iterations: 852000. The windows defined in (20) and (21) were used.

Results are shown in Figs. 11–13. In Fig. 11, we find that the desired part of the global impulse response $g(n)$ seems to be simply the truncated version of the RIR $c(n)$. The comparison of the corresponding frequency responses in Figs. 6 and 13 shows

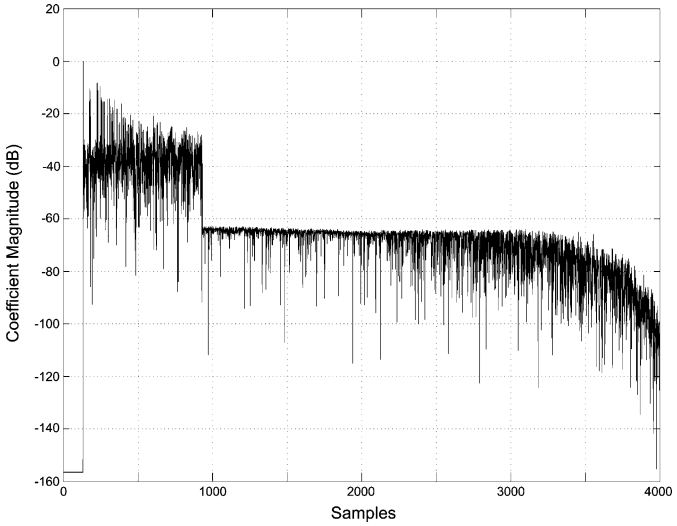

Fig. 12. Decay of the global impulse response $g(n)$ designed with the D50 measure.

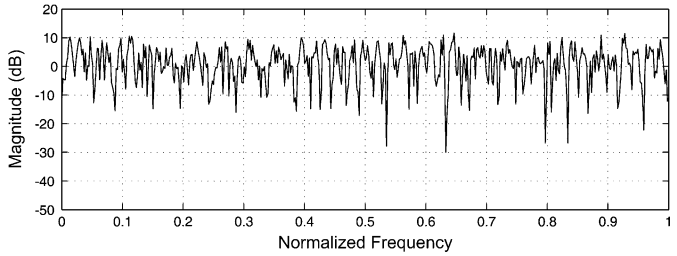

Fig. 13. Magnitude frequency response of the shortened global system designed via the D50 measure.

TABLE IV
RELATIONSHIP BETWEEN THE LENGTH OF THE PREFILTER (L_h) AND THE ATTENUATION OF THE UNWANTED PART (A_u) FOR $t_d = 50$ ms

L_h	4000	3500	3000	2500	2000
A_u (dB)	82.0	80.5	74.5	67.7	62.2

TABLE V
RELATIONSHIP BETWEEN THE LENGTH OF THE DESIRED PART (t_d) AND THE ATTENUATION OF THE UNWANTED PART (A_u) FOR A PREFILTER LENGTH OF $L_h = 3500$

t_d (ms)	50	40	30	20	10
A_u (dB)	80.5	76.5	72.0	66.6	55.4

TABLE VI
FREQUENCY-DOMAIN DEVIATIONS $\|C_{ps}(\cdot) - G_{ps}(\cdot)\|_2$ BETWEEN THE SHORTENED/RESHAPED RIR AND THE ORIGINAL RIR

	D50, p-norm	Masking, p-norm	D50, 2-norm
$\|C_{ps}(\cdot) - G_{ps}(\cdot)\|_2$	0.37	1.29	13.79

that this truncation-like property prevents $g(n)$ from producing serious distortion in the frequency domain. The unwanted part of the global impulse response is attenuated more than 62 dB. If the shortening filter is set to be of a length of $L_h = 3500$, then the unwanted part will be attenuated to more than 80 dB (see Table IV), so this part cannot be heard from the output signal.

TABLE VII
FREQUENCY-DOMAIN DEVIATIONS $\|C_{ps}(\cdot) - G_{ps}(\cdot)\|_2$ BETWEEN THE SHORTENED RIRS AND THE ORIGINAL ONES FOR FOUR DISTANCES BETWEEN MICROPHONES AND SPEAKERS IN THE LABORATORY (LAB.) AND THE LECTURE THEATER (LT)

Room	Method	Distance			
		1 m	2 m	3 m	4 m
Lab.	D50 measure	0.53	0.35	0.27	1.65
Lab.	masking window	1.92	2.54	2.43	1.55
Lab.	D50 measure, 2-norm	12.76	8.77	9.70	12.12
Lab.	masking window, 2-norm	7.88	7.43	7.31	10.03
LT	D50 measure	1.16	0.58	2.42	0.88
LT	masking window	1.66	2.44	2.05	1.81
LT	D50 measure, 2-norm	11.14	8.90	8.62	8.60
LT	masking window, 2-norm	8.68	7.66	7.52	9.10

TABLE VIII
FREQUENCY-DOMAIN DEVIATIONS BETWEEN PERFECTLY FLAT FREQUENCY RESPONSES AND THE ONES OF ORIGINAL AS WELL AS SHORTENED RIRS FOR FOUR DISTANCES BETWEEN MICROPHONES AND SPEAKERS IN THE LABORATORY (LAB.) AND THE LECTURE THEATER (LT)

Room	Method	Distance			
		1 m	2 m	3 m	4 m
Lab.	original	3.10	2.94	2.61	2.69
Lab.	D50 measure	2.87	2.87	2.64	3.97
Lab.	masking window	2.00	1.14	1.16	1.94
Lab.	D50 measure, 2-norm	14.15	8.23	10.10	12.53
Lab.	masking window, 2-norm	8.40	7.81	7.81	11.21
LT	original	3.60	3.58	2.96	2.82
LT	D50 measure	3.96	3.45	4.81	3.30
LT	masking window	3.47	2.99	2.58	3.55
LT	D50 measure, 2-norm	12.51	9.76	10.82	9.41
LT	masking window, 2-norm	10.90	8.89	8.57	10.54

Informal listening tests showed that the echoes are effectively suppressed.

If the length of the desired part is changed, the attenuation in the unwanted part will change accordingly. Results for different durations of the desired part are reported in Table V, where the basic parameters were kept as mentioned at the beginning of this subsection. Informal listening tests showed that no echoes are heard if 20 ms $\leq t_d \leq$ 50 ms. On the other hand, if the length of the desired part is fixed and the attenuation of the unwanted part is not sufficient, then we have to increase the length of the prefilter $h(n)$. The relationship between the length of the prefilter and the attenuation of the unwanted part for a desired part of length $t_d = 50$ ms is given in Table IV.

C. Frequency-Domain Changes by Shortening/Reshaping

The frequency-domain characteristics of a RIR are inevitably changed by the shortening/reshaping processes. Therefore, it would be the best if we could achieve reverberation suppression while keeping the perceivable changes of the frequency-domain characteristics as small as possible. In order to provide a numerical measure that reflects the deviation between frequency responses from a perceptual point of view, we follow the approach in [22]. Thus, for a given impulse response, we first compute the squared frequency response, interpolate it uniformly on a logarithmic scale, smooth it with 0.2-octave resolution, and then represent the result in dB. In order to ensure that our measure is independent of a gain factor and the frequency resolution used during computation, we compensate for

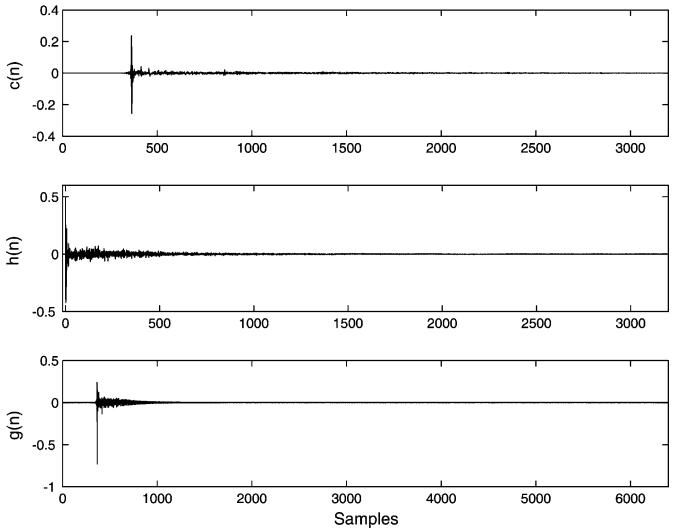

Fig. 14. Realistic example (masking limit): the room impulse response $c(n)$ (top) was measured in a regular laboratory room, the reshaping filter $h(n)$ (middle) designed according to c(n), and the global impulse response $g(n) = c(n) * h(n)$ (bottom).

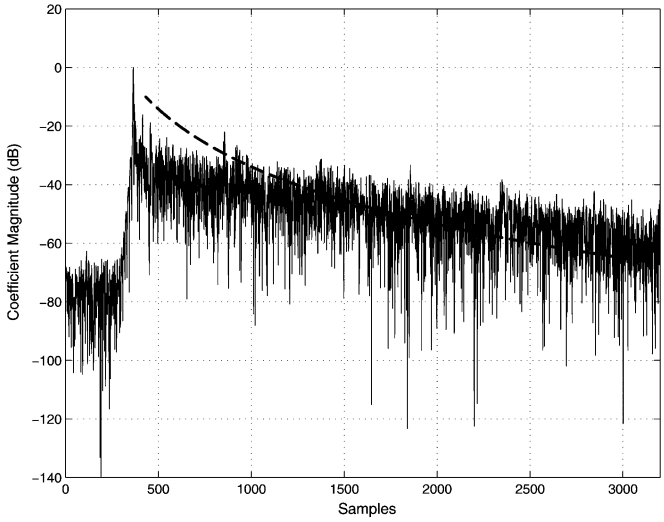

Fig. 15. Decay of the measured room impulse response $c(n)$ in Fig. 14. The dashed curve is the logarithmic reciprocal of the masking limit window function $w_0(n)$.

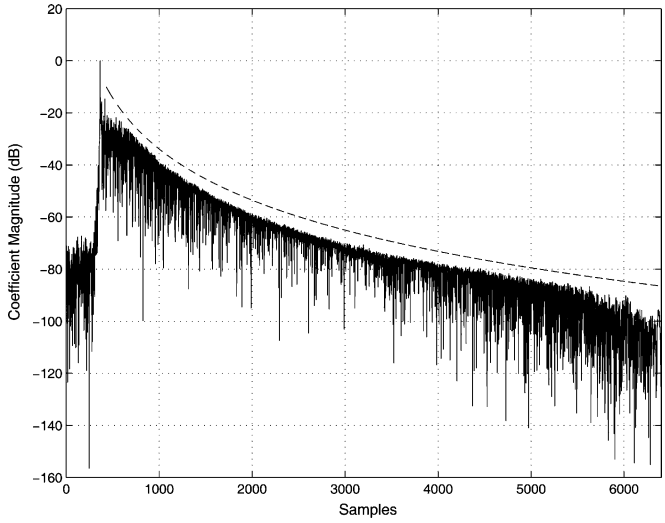

Fig. 16. Decay of the global impulse response $g(n)$ in Fig. 14. The dashed curve is the logarithmic reciprocal of the masking limit window function $w_0(n)$.

the mean and the number of computed frequency points. For impulse responses $c(n)$ and $g(n)$, we denote the obtained transfer functions by $C_{ps}(\log(\omega))$ and $G_{ps}(\log(\omega))$, respectively, where the index $_{ps}$ stands for perceptual smoothing. The root mean squared error between $C_{ps}(\log(\omega))$ and $G_{ps}(\log(\omega))$, i.e., $\|C_{ps}(\cdot) - G_{ps}(\cdot)\|_2$ then gives us an objective measure that expresses the perceptual spectral difference between the two filters.

The overall impulse responses $g(n)$ in Figs. 1, 7, and 11 are used for a comparison with the original RIR $c(n)$. The measures $\|C_{ps}(\cdot) - G_{ps}(\cdot)\|_2$ are presented in Table VI. One can see that the RIR shortened under the D50 measure with the p-norm algorithm has the least deviation from the original one. This is consistent with listening tests: speech signals from such a shortened RIR sound most similar to the original signals. When using the masking-window method, dereverberation can be achieved with a shorter reshaping filter than under the D50 measure, but this is at the cost at a slightly larger deviation in the frequency domain. The 2-norm algorithm under the D50 measure yields by far the worst result and also shows the smallest reverberation suppression.

Of course, the frequency-domain measures can be improved with the postprocessing method of [1] for all of the three methods. In particular, as our proposed methods do not introduce significant frequency-domain errors, the postfilter method from [1], which was originally designed to alleviate the problems created by the least-squares method, can also be used for preprocessing RIRs that contain high-Q resonances before the reshaping takes place on the basis of the adjusted RIR. Simulations confirmed that this is a feasible and highly effective approach.

D. Measurements in a Real Room

Experiments were done in a regular laboratory room of size 6.9 m $\times$ 5.0 m $\times$ 3.0 m and a lecture theater of size 11 m $\times$ 6.9 m $\times$ 3.0 m. The reverberation times of the laboratory room and the lecture theater were $\mathrm{RT60} = 318$ ms and $\mathrm{RT60} = 450$ ms (estimated from RIR inverse integration), respectively. One loudspeaker and four microphones were used for signal playing and recording. The microphones were placed in distances of 1, 2, 3, and 4 m from the loudspeaker. First, the RIRs (length 3200 for the laboratory room and length 4000 for the lecture theater) were measured with pseudo-white noise as the sources. Then shortening/reshaping filters of length 3200 and 4000, respectively, were calculated for these RIRs. The RIR to the microphone placed 1 m from the loudspeaker and the corresponding reshaping results are demonstrated in Figs. 14–16. In Fig. 15, we see that the tail of the original RIR clearly exceeds the average masking curve, whereas the reshaped response in Fig. 16 stays under this limit.

In another setup, original speech signals were played by the loudspeaker, and the reverberated room signals were recorded

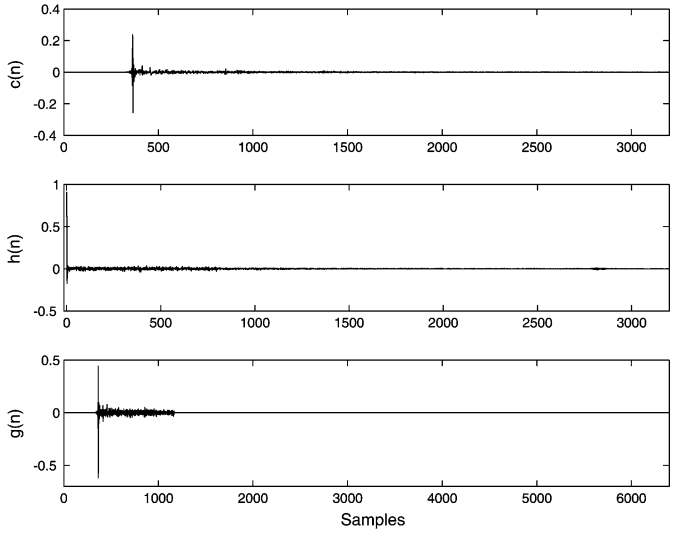

Fig. 17. Realistic example (D50 measure): the room impulse response $c(n)$ (top, the same as that in Fig. 14), the shortening filter $h(n)$ (middle) designed according to c(n), and the global impulse response $g(n) = c(n) * h(n)$ (bottom).

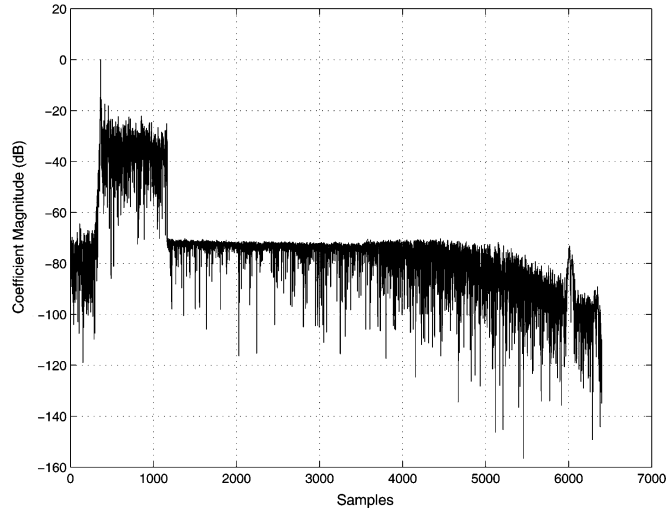

Fig. 18. Decay of the global impulse response $g(n)$ in Fig. 17.

simultaneously with the four microphones. Informal listening tests showed that reverberations of room signals were obviously heard, but that the reverberations were almost inaudible after postprocessing with the designed reshaping filters. Moreover, we preprocessed the source signals with the reshaping filters before playback. In this case, the reverberations were clearly suppressed in the room-recorded signals at the corresponding microphone positions.

Results for the D50 measure are demonstrated in Figs. 17 and 18. Listening tests confirmed that also this method works well for the realistic situations.

The frequency-domain similarity measures for shortening/reshaping with different methods and for different microphone distances are presented in Table VII. One can see that these real-world results show a similar behavior as the simulated ones. Again, the shortened RIR under the D50 measure is the most similar one to the original RIR in frequency-domain. Moreover, one can see that the frequency-domain deviations are quite independent of the room dimensions and the distances between the microphones and loudspeakers. In addition, Table VIII shows frequency-domain deviations between perfectly flat frequency responses and the ones of the original as well as the shortened RIRs for the four distances between microphones and speakers. These results show that the frequency-domain imperfections of responses shortened with our proposed method are in the same range as the original ones. In most cases, the masking-window method can even improve the behavior in comparison to the original one, which can be explained by the fact that this method tries to obtain one dominant filter tap.

V. Conclusion

The approaches for RIR shortening/reshaping proposed in this paper are motivated by the idea of equiripple filter design. Infinity- and p-norm criteria have been used instead of the traditionally applied least squares criterion. For a good perceived quality with relatively short prefilters, masking effects of the human auditory system and the D50 measure were exploited during the prefilter design. Experiments prove that the proposed methods are feasible for RIR shortening and reshaping and are superior to least-squares approaches.

References

[1] M. Kallinger and A. Mertins, "Room impulse response shortening by channel shortening concepts," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, 2005, pp. 898–902.

[2] M. Kallinger and A. Mertins, "Impulse response shortening for acoustic listening room compensation," in *Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC)*, 2005, pp. 197–200.

[3] A. Azzali, A. Bellini, A. Farina, and E. Ugolotti, "Design and implementation of psychoacoustics equalizer for infotainment," DSP Application Day, Politecnico di Milano, Milan, Italy, 2002 [Online]. Available: http://www.ramsete.com/Public/Papers/list_pub_auro.htm

[4] B. D. Radlović and R. A. Kennedy, "Nonminimum-phase equalization and its subjective importance in room acoustics," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 6, pp. 728–736, Nov. 2000.

[5] M. Karjalainen and T. Paatero, "Equalization of loudspeaker and room responses using Kautz filter: Direct least squares design," *EURASIP J. Adv. Signal Process.*, vol. 2007, pp. 1–13, 2007.

[6] A. Maamar, I. Kale, A. Krukowski, and B. Daoud, "Partial equalization of non-minimum-phase impulse responses," *EURASIP J. Appl. Signal Process.*, vol. 2006, pp. 1–8, 2006.

[7] L. D. Fielder, "Practical limits for room equalization," in *Proc. AES 111th Conv.*, 2001, pp. 1–19.

[8] 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.

[9] *ISO Norm 3382: Acoustics-Measurement of the Reverberation Time of Rooms With Reference to Other Acoustical Parameters*, SO Norm 3382, Int. Org. Standardization (ISO).

[10] J. N. Mourjopoulos, "Digital equalization of room acoustics," *J. Audio Eng. Soc.*, vol. 42, no. 11, pp. 884–900, 1994.

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

[12] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Upper Saddle River, NJ: Prentice Hall, 1999.

[13] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*. New York: Springer-Verlag, 1999, pp. 82–84.

[14] D. H. Raab, "Forward and backward masking between acoustic clicks," *J. Acoust. Soc. Amer.*, vol. 33, pp. 137–139, 1961.

[15] S. H. Olive and F. E. Toole, "The detection of reflections in typical rooms," *J. Audio Eng. Soc.*, vol. 37, pp. 539–553, 1989.

[16] W. Jesteadt, S. P. Bacon, and J. R. Lehmen, "Forward masking as a function of frequency, masker level and signal delay," *J. Acoust. Soc. Amer.*, vol. 71, pp. 950–962, 1982.

[17] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
[18] R. K. Martin, M. Ding, B. L. Evans, and C. R. Johnson Jr., "Efficient channel shortening equalizer design," *EURASIP J. Appl. Signal Process.*, vol. 13, pp. 1279–1290, 2003.
[19] T. Dau, B. Kollmeier, and A. Kohlrausch, "Modelling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," *J. Acoust. Soc. Amer.*, vol. 102, pp. 2892–2905, 1997.
[20] J. Buchholz, J. Mourjopoulos, and J. Blauert, "Room masking: Understanding and modelling the masking of room reflections," in *Proc. Audio Eng. Soc. 110th Conv.*, Amsterdam, The Netherlands, May 2001, Paper 5312.
[21] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustic," *J. Acoust. Soc. Amer.*, vol. 65, pp. 943–950, 1979.
[22] J. Huopaniemi, N. Zacharov, and M. Karjalainen, "Objective and subjective evaluation of head-related transfer function filter design," *J. Audio Eng. Soc.*, vol. 47, no. 4, pp. 218–239, Apr. 1999.

Alfred Mertins (M'96–SM'03) received the Dipl.-Ing. degree from the University of Paderborn, Paderborn, Germany, in 1984, the Dr.-Ing. degree in electrical engineering and the Dr.-Ing.habil. degree in telecommunications from the Hamburg University of Technology, Hamburg, Germany, in 1991 and 1994, respectively.

From 1986 to 1991, he was a Research Assistant at the Hamburg University of Technology, and from 1991 to 1995 he was a Senior Scientist at the Microelectronics Applications Center Hamburg. From 1996 to 1997, he was with the University of Kiel, Kiel, Germany, and from 1997 to 1998 with the University of Western Australia, Perth. In 1998, he joined the University of Wollongong, where he was at last an Associate Professor of Electrical Engineering. From 2003 to 2006, he was a Professor in the Faculty of Mathematics and Science at the University of Oldenburg, Germany. In November 2006, he joined the University of Lübeck, Lübeck, Germany, where he is a Professor and Director of the Institute for Signal Processing. His research interests include speech, audio, and image processing, wavelets and filter banks, pattern recognition, and digital communications.

Tiemin Mei received the B.S. degree in physics from Sun Yat-sen University, Guangzhou, China, in 1986, the M.S. degree in biophysics from China Medical University, Shenyang, in 1991, and the Ph.D. degree from Dalian University of Technology, Dalian, China, in 2006.

He has been with the Institute for Signal Processing, University of Lübeck, Lübeck, Germany, since 2007. He was a Visiting Fellow with the School of Electrical Computer and Telecommunications Engineering, University of Wollongong, Wollongong, Australia, from 2004 to 2005. He has also been a member of academic staff of the School of Information Science and Engineering, Shenyang Ligong University, Shenyang, since 1996. His current research interests include stochastic signal and speech processing.

Markus Kallinger (M'06) received the Dipl.-Ing. degree in electrical engineering from the University of Ulm, Ulm, Germany, in 1999 and the Dr.-Ing. degree in electrical engineering from the University of Bremen, Bremen, Germany, in 2006.

From 1999 to 2004, he was a Research Assistant with the Department of Communications Engineering, University of Bremen. From 2004 to 2006, he was a Research Fellow at the Faculty of Mathematics and Science, University of Oldenburg, Oldenburg, Germany. Since 2007, he has been with the Audio Department, Fraunhofer IIS, Erlangen, Germany. His research interests include speech and audio processing, adaptive filters, and digital audio effects.