ROOM IMPULSE RESPONSE RESHAPING/SHORTENING BASED ON LEAST MEAN SQUARES OPTIMIZATION WITH INFINITY NORM CONSTRAINT

Tiemin Mei, Alfred Mertins, and Markus Kallinger

Institute for Signal Processing,
University of Lübeck, 23538 Germany

ABSTRACT

The purpose of room impulse response (RIR) reshaping or shortening is to accelerate the attenuation of the original RIR so that the reverberation effect will be weakened and the intelligibility of speech played in the associated room will be improved. The unwanted energy of the RIR, which is captured with the help of a window function defined according to the average masking effect of the auditory system, is minimized with the temporal constraint of keeping the infinity norm of the global impulse response constant. Compared with some well known approaches, this new method demonstrates excellent performance in terms of the effectiveness of reshaping/shortening the impulse response while closely retaining the frequency response of the room.

Index Terms— room impulse response, reshaping, shortening, least mean squares, infinity norm, optimization

1. INTRODUCTION

It is well known that reverberation has different effects on music and speech played in a room. If the reverberation time of a room is too short, it will be listening "too dry" for music playing, but in contrary, if speech is played in the room, a short reverberation time is preferred for good intelligibility.

For the enhancement of speech intelligibility in reverberant rooms and for new applications in audio-visual communications and virtual acoustics, a suitable pre-processing of loudspeaker signals is needed to compensate room reverberation, namely, the listening-room-compensation or room reverberation compensation [1][2][3]. Similarly, for improving the quality of far-field microphone recordings, a post-filtering stage may be introduced for the received signals. Both problems are mathematically equivalent.

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 in such a way that signals are perceived without reverberation. In other words, it would be sufficient to partially equalize a RIR [5][6], so that all audible echoes are removed and the inaudible ones remain. Such a relatively relaxed requirement is not only possible according to the temporal masking effect of the human auditory system but also will greatly alleviate the pressure of designing a compensation system.

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 which is induced by the direct sound [7][8]. This is known as the temporal masking effect. While the masking effects of the human auditory system are signal dependent in general [9][10], 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]. We will show that a window function constructed from such an average masking limit can be used to control the attenuation behavior of the reshaped impulse response.

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:

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

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.

In this paper, for filter reshaping/shortening, we use windows $w_u(n)$ and $w_d(n)$ to derive an unwanted part $g_u(n) = w_u(n)g(n)$ and a wanted part $g_d(n) = w_d(n)g(n)$ from the global impulse response $g(n)$. Our goal is to minimize some

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

978-1-4244-3298-1/09/$25.00 ©2009 IEEE

function of $|g_u(n)|$ while keeping the infinity norm of the global impulse response $g(n)$ or $g_d(n)$ constant with respect to the prefilter $h(n)$, which should enable a reshaping of $g(n)$ without significantly affecting the magnitude frequency response of it.

3. APPROACH DEVELOPMENT

The purpose of filter reshaping/shortening is to accelerate the decreasing of the original impulse response $c(n)$ or shorten its effective duration, so that the reverberation time will be shortened and the intelligibility of speech will be enhanced. From this point of view, it is straightforward that we prefer under a given condition to keep the global impulse response as close as possible to a delayed unite impulse $\delta(n-d)$.

It is intuitive if we minimize under some measure the unwanted part of the global impulse response $g(n)$ while keeping the maximal absolute value of the desired part of $g(n)$. This is just exactly what we do in the inverse filter design, but here we have extended this idea to the reshaping/shortening problem. Accordingly, we define the following optimization problem in terms of least mean squares and infinity norm,

$$\left\{ \begin{array}{l} \mathrm{MIN}_{\mathbf{h}}: \ f(\mathbf{h}) = \mathbf{g}_u^{\mathrm{T}}\mathbf{g}_u = \mathbf{h}^{\mathrm{T}}\mathbf{A}\mathbf{h} \\ \mathrm{S.\,T.}: \ \|g_d(n)\|_\infty = |g(l_{max})| = 1 \end{array} \right. , \tag{2}$$

where $|g(l_{max})|$ is the maximal absolute value of $g(n)$ and $\mathbf{A} = \mathbf{C}^{\mathrm{T}}\mathrm{diag}[\mathbf{w}_u^2]\mathbf{C}$.

Before we get the optimal $h(n)$, the problem is how to determine the maximal absolute value position l_{max} of $g(n)$. From the point of view of reducing the global impulse response delay, the maximal absolute value of $g(n)$ should appear as early as possible, so it is a natural way to suppose that the maximal absolute value of $g(n)$, i.e., $|g(l_{max})|$, will appear after optimization at the same position as that of $c(n)$, i.e., $l_{max} = l_{maxc}$, where l_{maxc} represents the maximal absolute value position of $c(n)$.

For the simplicity of expressions, we suppose that $c(n)$ is normalized so that the maximal value of $c(n)$ is also its maximal absolute value and has the value one, i.e., $|c(l_{maxc})| = c(l_{maxc}) = 1$. Moreover, we suppose without the loss of generality that $|g(l_{max})| = g(l_{max}) = 1$. So the optimization problem is simplified as:

$$\left\{ \begin{array}{l} \mathrm{MIN}_{\mathbf{h}}: \ f(\mathbf{h}) = \mathbf{h}^{\mathrm{T}}\mathbf{A}\mathbf{h} \\ \mathrm{S.\,T.}: \ \sum_{k=0}^{k=l_{maxc}} h(k)c(l_{maxc}-k) = 1 \end{array} \right. \tag{3}$$

or in matrix form:

$$\left\{ \begin{array}{l} \mathrm{MIN}_{\mathbf{h}}: \ f(\mathbf{h}) = \mathbf{h}^{\mathrm{T}}\mathbf{A}\mathbf{h} \\ \mathrm{S.\,T.}: \ \mathbf{c}_{maxc}^{\mathrm{T}}\mathbf{h} = 1 \end{array} \right. , \tag{4}$$

where $\mathbf{c}_{maxc} = [1, c(l_{maxc}-1), ..., c(0), 0, ..., 0]^{\mathrm{T}}$ is the transpose of the l_{maxc}th row of the convolution matrix $\mathbf{C}$.

Applying the Lagrange multiplier method to (4), we have

$$\begin{bmatrix} \mathbf{A} & \mathbf{c}_{maxc} \\ \mathbf{c}_{maxc}^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{h} \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \tag{5}$$

where λ is Lagrange multiplier. Because of the special form of the coefficient matrix and the right-hand-side vector of the linear equation system (5), $\mathbf{h}$ is obtained as

$$\mathbf{h} = \frac{\mathbf{A}^{-1}\mathbf{c}_{maxc}}{\mathbf{c}_{maxc}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{c}_{maxc}}. \tag{6}$$

In practice, in most cases the direct sound is the strongest and arrives at microphone earliest, which implies that the vector $\mathbf{c}_{maxc}$ has the special form $\mathbf{c}_{maxc} = [1, 0, ..., 0]^{\mathrm{T}}$, this enables us to further simplify the solving for $\mathbf{h}$. Firstly, if $\mathbf{c}_{maxc} = [1, 0, ..., 0]^{\mathrm{T}}$, it is easy to see that in (5) $h(0) = 1$; secondly, let

$$\mathbf{B} = \begin{bmatrix} a_{22}, \cdots, a_{2L_h} \\ \vdots, \ \ddots, \ \vdots \\ a_{L_h 2}, \cdots, a_{L_h L_h} \end{bmatrix} \tag{7}$$

and $\mathbf{q} = [a_{21}, a_{31}, ..., a_{L_h 1}]^{\mathrm{T}}$, where a_{ij} $(i, j = 1, 2, ..., L_h)$ are the elements of matrix $\mathbf{A}$. Then the components of $\mathbf{h}$, except $h(0)$, are the solution of the following equation:

$$\mathbf{B}\mathbf{h}_b = \mathbf{q}, \tag{8}$$

where $\mathbf{h}_b = [h(1), h(2), ..., h(L_h-1)]^{\mathrm{T}}$ and $\mathbf{h} = [1, \mathbf{h}_b^{\mathrm{T}}]^{\mathrm{T}}$.

In addition to the above methods, we also propose the following constrained optimization problem for reshaping/shortening filter design with a constraint that is different from the one in (2):

$$\left\{ \begin{array}{l} \mathrm{MIN}_{\mathbf{h}}: \ f(\mathbf{h}) = \mathbf{h}^{\mathrm{T}}\mathbf{A}\mathbf{h} \\ \mathrm{S.\,T.}: \ h(0) = 1 \end{array} \right. . \tag{9}$$

The solution of this problem is exactly the same as that of (4) for the special case $\mathbf{c}_{maxc} = [1, 0, ..., 0]^{\mathrm{T}}$, i.e., $\mathbf{h} = [1, \mathbf{h}_b^{\mathrm{T}}]^{\mathrm{T}}$ and $\mathbf{h}_b$ is the solution of (8), whatever the true $\mathbf{c}_{maxc}$ of problem (9) is.

4. A SUMMARY OF TWO RELATED ALGORITHMS BY OTHER AUTHORS

For comparison purposes, two related algorithms are summarized here. Firstly, a well-known approach for filter shortening is to optimize $h(n)$ under the least-squares error criterion, i.e. [11][12],

$$\left\{ \begin{array}{l} \mathrm{MIN}_{\mathbf{h}}: f(\mathbf{h}) = \mathbf{h}^{\mathrm{T}}\mathbf{A}\mathbf{h} \\ \mathrm{S.\,T.}: \mathbf{g}_d^{\mathrm{T}}\mathbf{g}_d = \mathbf{h}^{\mathrm{T}}\mathbf{B}\mathbf{h} = 1 \end{array} \right. , \tag{10}$$

where $\mathbf{A}$ is the same as that in (3) and $\mathbf{B} = \mathbf{C}^{\mathrm{T}}\mathrm{diag}[\mathbf{w}_d^2]\mathbf{C}$. The approach is definitely suitable for filter reshaping/shortening through well-defined windows $w_u(n)$ and $w_d(n)$.

This problem is equivalent to the following generalized eigenvalue decomposition and the optimal $\mathbf{h}$ is the eigenvector corresponding to the smallest generalized eigenvalue:

$$\mathbf{A}\mathbf{h} = \lambda_{\min}\mathbf{B}\mathbf{h}. \tag{11}$$

In [12], 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, a prefilter $\mathbf{h}$ that is optimal in the least-squares sense (10) will usually make great distortion in the frequency domain, moreover, the temporal shape 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.

Secondly, the following optimization problem was proposed in [12]:

$$\begin{cases} \mathrm{MIN}_{\mathbf{h}} : f(\mathbf{h}) = \mathbf{h}^{\mathrm{T}}\mathbf{A}\mathbf{h} \\ \mathrm{S.\,T.} : \mathbf{h}^{\mathrm{T}}\mathbf{h} = 1 \end{cases}. \tag{12}$$

The optimal solution is the eigenvector corresponding to the smallest eigenvalue of the following eigenvalue problem:

$$\mathbf{A}\mathbf{h} = \lambda_{\min}\mathbf{h}. \tag{13}$$

5. SIMULATIONS

A simulated RIR has been used in the experiments, where basic parameters were selected as $f_s = 16\mathrm{kHz}$, $L_c = 2000$, and $L_h = 1500$. Different windows have been designed for the filter reshaping and shortening tasks, respectively.

For reshaping, we define the window $w_u(n)$ proportional to the inverse of the average temporal masking curve [13],

$$w_u(n) = 10^{\frac{3}{\log(N/N_0)}\log(\frac{n}{N_0})+0.5}$$

where $N_0 = 196$, $N = 3332$, and $N_0 \leq n \leq L_g$ for our case.

For shortening, the window $w_u(n)$ is defined in such a way that it will allow us to suppress the late diffuse wave of the global impulse response. So we propose the following shortening window:

$$w_u(n) = 3 * \sin^2(\omega(n - n_0)/N) + 2,$$

where ω and n_0 are adjustable to obtain a suitable window $w_u(n)$ for any given RIR. In our examples, the selected parameters were $\omega = 0.99\pi$, $n_0 = 200$, $N = L_g - l_{maxc} - l_d$, where $l_d = 800$ is the effective interval of the shortened filter and $l_{maxc} = 132$, $0 \leq n \leq N$.

Algorithms (2) and (9) give similar results when the vector $\mathbf{c}_{maxc}$ has the property $\mathbf{c}_{maxc} \approx [1, 0, ..., 0]^{\mathrm{T}}$, so the following comparisons are focused on (2), (10) and (12). The same windows were used for all of the three algorithms.

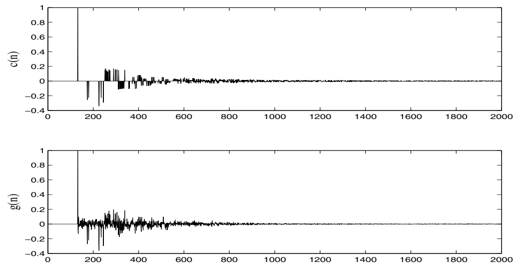

Fig. 1. The original (Top) and reshaped (Bottom, for comparison purpose, only the first 2000 samples are presented) room impulse responses with algorithm (2).

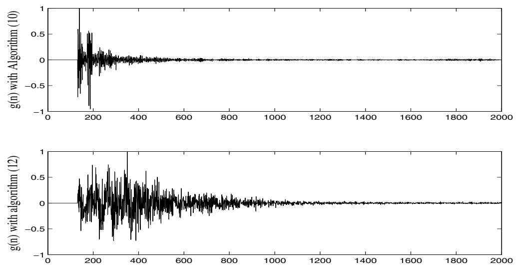

Fig. 2. The reshaped impulse responses with algorithms (10) (Top) and (12) (Bottom).

5.1. Reshaping performance comparison

For room impulse response reshaping, informal listening tests show that algorithm (2) outperforms algorithms (10) and (12). Almost no echoes are heard from the output signal with our new approach, moreover, the room characteristic sounds quite natural. On the contrary, both of the output signals with approaches (10) and (12) are even worse than those before reshaping. We can refer this to both of the time-domain and frequency-domain characteristics of the reshaped impulse responses given with different methods. The time-domain characteristics are shown in Figs. 1 and 2, and the frequency domain effects are depicted in Figs. 3 to 6 for the different methods. These figures show that algorithm (2) keeps the reshaped global impulse response as close as possible to the original room impulse response in both time and frequency domains while accelerating the decay of the global impulse response, so the reverberation effects can be obviously improved. However, algorithms (10) and (12) can not effectively accelerate the decay of the global impulse response, and furthermore, they also strongly distort the frequency-domain characteristic of the global impulse response. For further illustration, the time-domain attenuation characteristics of the original and the reshaped impulse responses are also shown as plots of $20\log_{10}(|g(n)|)$ in Figs. 7 to 10.

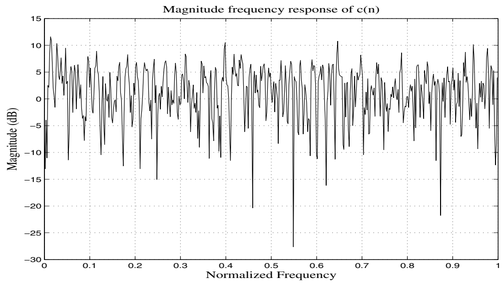

Fig. 3. The magnitude frequency response of the original room impulse response.

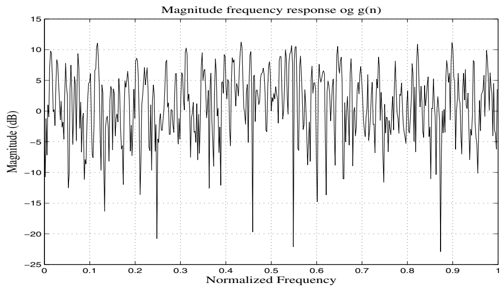

Fig. 4. The magnitude frequency response of the reshaped room impulse response with algorithm (2).

5.2. Shortening performance comparison

For impulse response shortening, the newly proposed algorithm (2) shows also a better performance in both time and frequency domains than the other two approaches. This can be seen clearly in Figs. 11 to 17. Although the least-squares measures always result in late diffuse echoes, as a result of the carefully designed window $w_u(n)$, the unwanted parts of the global impulse responses are obviously improved for all of the three approaches. After shortening with our new method, the global impulse response $g(n)$ seems as if it were the truncated version of the original response $c(n)$, plus some noise. Most of its energy is concentrated in one dominating tap. However, for the approaches (10) and (12) the energies of the $g(n)$'s are almost uniformly distributed over the desired parts of them. This can be clearly seen in Fig. 11 and Figs. 12, 13, and 14. In addition, our new approach causes less frequency-domain distortion to the global system than the algorithms (10) and (12), see Figs. 15, 16, 17 and 3.

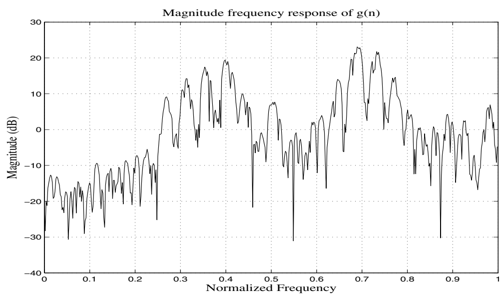

Fig. 5. The magnitude frequency response of the reshaped room impulse response with algorithm (10).

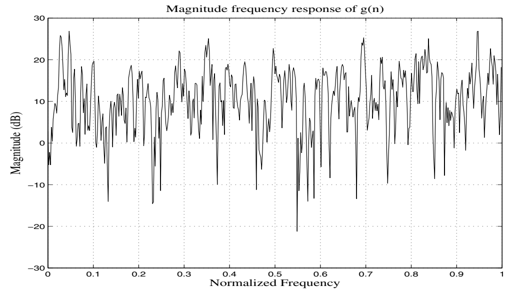

Fig. 6. The magnitude frequency response of the reshaped room impulse response with algorithm (12).

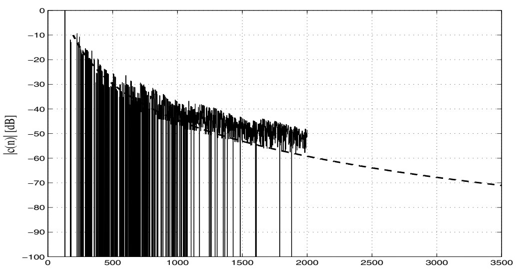

Fig. 7. Magnitude of the original room impulse response in dB. The dashed curve represents the average masking limit.

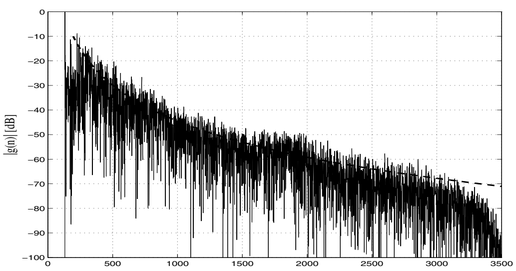

Fig. 8. Magnitude of the reshaped impulse response with algorithm (2). The dashed curve represents the average masking limit.

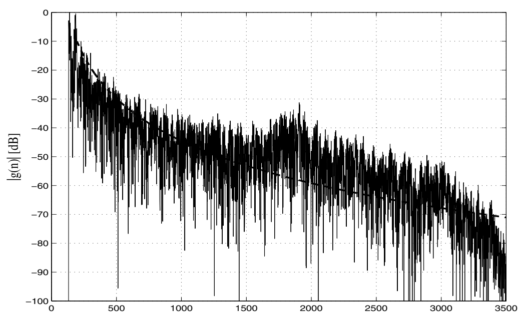

Fig. 9. Magnitude of the reshaped impulse response with algorithm (10).

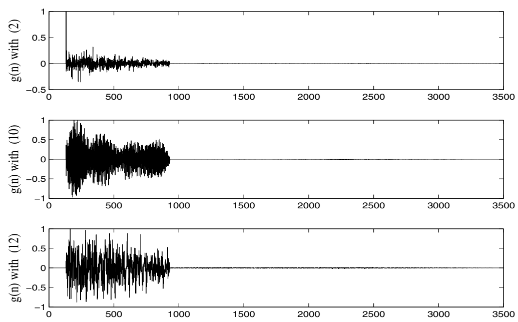

Fig. 11. The three shortened room impulse responses with algorithms (2), (10) and (12): From Top to Bottom.

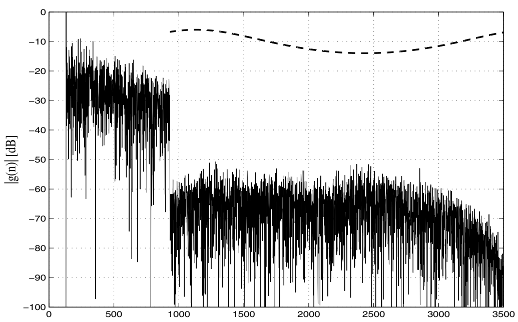

Fig. 12. Magnitude of the shortened impulse response with algorithm (2), the dashed curve is the attenuation of the reciprocal of window $w_u(n)$ in dB.

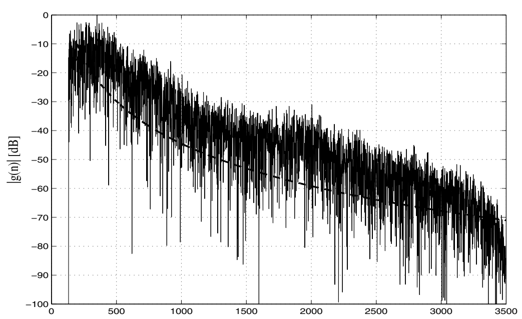

Fig. 10. Magnitude of the reshaped impulse response with algorithm (12).

6. CONCLUSIONS

In this paper, for room impulse response reshaping and shortening, we have presented an analytical solution of the least squares minimization problem under infinity norm constraint. The optimal reshaping/shortening filter is solved under some simplification of the infinity-norm constraint. Simulations prove that this method possesses some attractive characteristics and outperforms some known approaches.

7. 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] A. Azzali, A. Bellini, A. Farina, and E. Ugolotti, "Design and implementation of psychoacoustics equalizer for infotainment," DSP Application Day, Politecnico di Milano, 2002

[4] 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.

[5] 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.

[6] 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.

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

[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] T. Dau, B. Kollmeier, and A. Kohlrausch, "Modelling auditory processing of amplitude modulation. I. Detec-

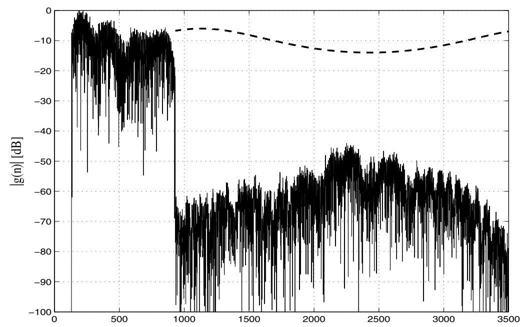

Fig. 13. Magnitude of the shortened impulse response with algorithm (10), the dashed curve is the attenuation of the reciprocal of window $w_u(n)$ in dB.

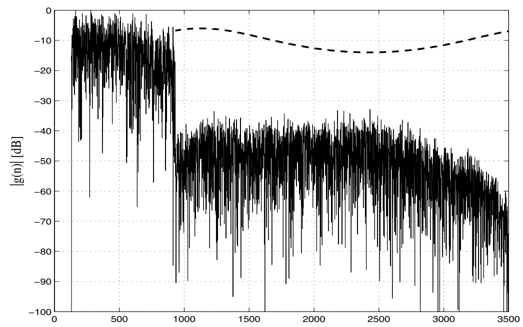

Fig. 14. Magnitude of the shortened impulse response with algorithm (12), the dashed curve is the attenuation of the reciprocal of window $w_u(n)$ in dB.

tion and masking with narrow-band carriers," J. Acoust. Soc. Am., vol. 102, pp. 2892-2905, 1997.

[10] J. Buchholz, J. Mourjopoulos, and J. Blauert, "Room Masking: Understanding and Modelling the Masking of Room Reflections", Audio Engineering Society 110th Convention, paper 5312, Amsterdam, May 2001.

[11] 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.

[12] 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.

[13] A. Mertins, T. Mei, and M. Kallinger, "Room Impulse Response Shortening/Reshaping with Infinity- or p-Norm Optimization," submitted to IEEE Trans. on Audio, Speech and Language Processing.

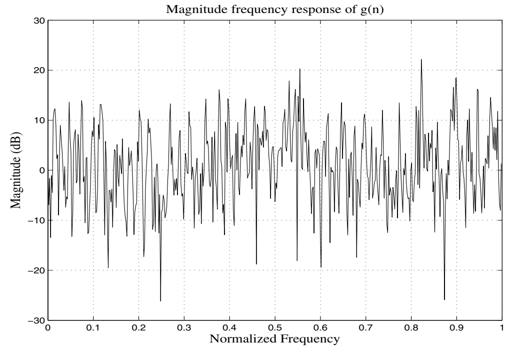

Fig. 15. The magnitude frequency response of the shortened room impulse response with algorithm (2).

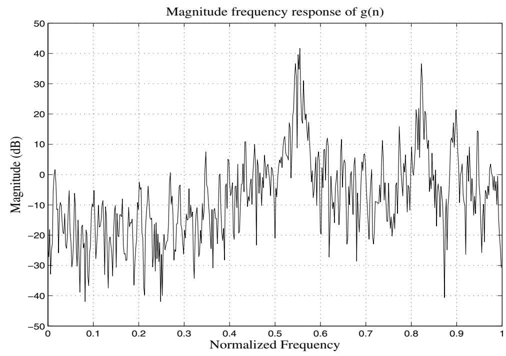

Fig. 16. The magnitude frequency response of the shortening room impulse response with algorithm (10).

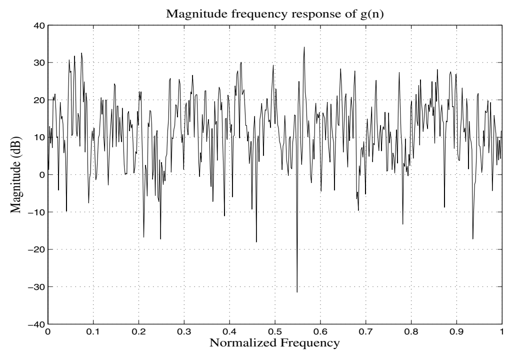

Fig. 17. The magnitude frequency response of the shortened room impulse response with algorithm (12).