Equalization in an Acoustic Reverberant Environment: Robustness Results

Biljana D. Radlović, Robert C. Williamson, and Rodney A. Kennedy[1]

Abstract

This paper investigates the robustness of sound equalization using a room response inverse filter with respect to changing or uncertain source or microphone positions. It is shown that due to the variations of the transfer function from point to point in a room, even small changes in the source or microphone position of just a few tenths of the acoustic wavelength can cause large degradations in the equalized room response. The robustness problem is especially acute at high frequencies, which are known to carry some important attributes of the speech signal. The spatial extent of equalization, derived from the statistical-average properties of sound transmission in rooms, is illustrated by computer simulations which corroborate the theoretical results presented in the paper.

EDICS Number: SA *2.2 Room Acoustics and Acoustic System Modeling*

Corresponding author:

Dr Rodney A. Kennedy
Telecommunications Engineering Group
Research School of Information Sciences and Engineering
Institute of Advanced Studies
Australian National University
Canberra ACT 0200, Australia
Phone: +61 2 6279 8800
Fax: +61 2 6279 8688
E-mail: rodney.kennedy@anu.edu.au

[1]B.D. Radlović and R.A. Kennedy are with the Telecommunications Engineering Group, Institute of Advanced Studies, RSISE, Australian National University, Canberra, Australia. R.C. Williamson is with the Department of Engineering, FEIT, Australian National University, Canberra, Australia.

I. Introduction

For many applications in room acoustics, when a microphone cannot be located close to the source, equalization filters are used to compensate distortion due to the frequency response of the room. Inverse filtering is especially important in conference telephony, where the speech signal is distorted by acoustic reverberation so it is less intelligible to the listeners. Parameters of the inverse filter are usually chosen in such a way that the difference between the filter output and a desired signal is minimized in some respect [1].

Most of the studies so far have been devoted to the case when the sound source and receiver locations are known exactly, which covers a small portion of the problems of interest in room acoustics. In a situation which occurs, for instance, with echo cancellation in a hands-free telephone set [2], the acoustic impulse response can change considerably due to the movement of persons, and the problem of removing unwanted reverberation becomes much more difficult.

These difficulties arise from the extremely irregular sound transmission path between the source and microphone in a room; both the peaks and dips of the acoustic frequency response change considerably throughout the room volume, with differences between response maxima and minima that extend beyond the fluctuation range of 10 dB [3]. For this reason, the performance of an inverse filter will be strongly influenced by even small changes in either the source or microphone position. These problems are further accentuated by the necessity to use inverse filters with a large number of taps, due to the long-delayed signals (echoes) in highly reverberant rooms [4].

Robustness of sound equalization in relation to movement of the sound source or receiver in an enclosed space has been addressed in several papers. Mourjopoulos [5], working in the time domain, evaluated variations in the impulse response functions measured in different rooms as the energy ratio between the early (direct) and late (reverberant) part of these functions. Results in [5] demonstrated that the inverse filtering operation, in general, increases the distortion present in the signals when a response recorded at a different position in the same enclosure is employed for dereverberation. In [6], Mourjopoulos proposed a vector quantization technique that selects an optimum inverse filter from a spatial equalization library whenever the receiver moves. Large sets of possible room transfer functions were grouped together and represented or equalized by a smaller number of equalizers. The main disadvantage of this approach is its complexity, because this technique requires prior measurements of the transfer function over a wide range of different source and microphone positions. Haneda *et al.* [7] proposed a method of modeling a room transfer function which is less sensitive to measurements taken for different source and receiver positions inside a room. The assumption underlying the model is that acoustical poles "do not change even if the source and receiver positions change or people move". A brief inspection of the work presented in [3], [8], [9] suggests that this assumption may be false. In [10], Asano *et al.* proposed a method for providing a wider equalization area based on constraining the partial derivatives of the transfer functions at the receiving points to be zero. This scheme, however, has not been demonstrated to be effective under reverberant conditions. Kerkhof and Kitzen [2] showed that the main difficulties in tracking the changes in an acoustic impulse response with an adaptive filter, when a person walks through a reverberant room, are due to the large bandwidth of speech and long reverberation time.

The purpose of this paper is to investigate the effect of source or microphone position changes on the performance of a fixed room response equalizer. In reverberant rooms, where the sound pressure may vary drastically, seemingly at random, from point to point and from one excitation frequency to another, such an effect is difficult to describe in precise mathematical terms. A rigorous analysis, using complex standing-wave patterns to describe the acoustic behavior of the room, is computationally exhaustive and leads to results applicable only in a particular case. Besides, its accuracy would be largely affected by possible errors in measuring the source or microphone positions with respect to the wall surfaces of the room. For these reasons, the study presented here utilizes statistical methods which have the useful feature that they are easy to implement and give more general and simpler results.

II. Analysis Approach

In this section, we provide a few entry points to the problem of robust dereverberation of the sound picked up by a microphone in an enclosed space.

The sound pressure at the microphone can be considered as being built up of a sound wave commencing at the source, plus many plane waves, due to multiple reflections of the original sound wave from the walls, travelling in different directions and encountering the walls of the room at various angles of incidence. In the time domain, these reflections are perceived as echoes and reverberation, that is, delayed, attenuated versions of the original source signal. In the frequency domain, depending upon the phase relationships between the sound waves at a given point, reflections produce peaks and dips in the frequency response of the room: the so-called "comb-filter" effect that may corrupt the quality of the sound.

It has proved possible to reduce the effect of reflections from the wall surfaces by using an inverse filter designed to compensate for uneveness in the room transfer function at the microphone position. However, such methods are usually less effective if the microphone is moved to some other point in a room where, due to the interference of waves reflected from the walls, it becomes impossible to know the spectrum of sound merely by knowing the spectrum at the reference point.

The pattern of cancellations and reinforcements, which result in the sound pressure being very different at different points in a room, can be precisely described by the theory of *standing waves*, or normal modes of vibration, made up of a number of plane waves that interfere after being reflected from the wall surfaces. At higher frequencies, however, due to a large number of standing waves excited by the source, the complexity of modal analysis increases to the point where exact mathematical expressions are no longer computationally simple or suitable for predicting the sound behavior from one point to another. For the purposes of this paper we will turn to another, not too rigorous but conceptually simple, model based on the statistical properties of reverberant fields in rooms.

The crucial assumption of statistical room acoustics is that the distribution of amplitudes and phases of individual plane waves, which sum up to produce sound pressure at some point in a room, is so close to random that the sound field is fairly uniformly distributed throughout the room volume. This theory closely describes the room acoustic behavior if the following conditions are met [11]:

1. The linear dimensions of the room must be large relative to the wavelength. This condition is easily satisfied in almost all rooms, for the frequencies of interest.

2. The average spacing of the resonance frequencies must be smaller than one-third of their bandwidth. In a room having a volume V (in m^3) and reverberation time ${T_{60}}^2$ (in sec), this criterion can be met for all frequencies that exceed the Schroeder large room frequency given by

$$f_S = 2000 \left(\frac{T_{60}}{V} \right)^{1/2} \mathrm{H}z\,. \tag{1}$$

For example, in an ordinary living room with a reverberation time of 1 s, statistical theory would be relevant above about 200 Hz.

3. Both source and microphone are in the interior of a room, at least a half-wavelength away from the walls, where the wavelength $\lambda = c/f$ and c is the velocity of sound, generally specified as 344 m/s at 21°C.

Under the above conditions, the frequency response between the source and receiver can be treated as a random function, the properties of which are determined by the room volume, reverberation time, and magnitude of the sound pressure [3], [8], [12]. The same will also be true for its inverse, a fact we exploit in the following sections.

III. Main Results

A. Robustness Analysis: Single-Channel Case

Consider the nonminimum-phase linear time-invariant acoustic system consisting of a source and a microphone placed some distance apart in a reverberant room. In order to determine the sensitivity of a single-channel equalizer to changes in the receiver (microphone) position, we start from the following simplifying assumption:

Let G_f be the complex steady-state frequency response between the sound source and the receiver, and H_f the frequency response of an inverse filter designed to equalize room response at the microphone point. We idealize our problem by assuming that the transmission path between the source and receiver is perfectly equalized, where *perfect* means equalizing both the amplitude and the phase of the frequency response.

Moving the microphone will distort the frequency response of the equalized transmission path. A quantitative measure of this degradation can be based on the difference between

[2] The reverberation time T_{60} is the length of time for the sound intensity level in a room to drop by 60 dB after the source is shut off.

the two transfer functions from the source to the reference and the displacement point, respectively.

Definition 1: Let $\widetilde{G}_f$ be the frequency response between the source and the receiver placed some distance away from the equalization point, and H_f an exact inverse of G_f. The *mean squared error* at frequency f due to the displacement of the receiving point is defined by

$$W_f = E\{|\widetilde{G}_f H_f - 1|^2\} \tag{2}$$

where $E\{\cdot\}$ represents the expected value operator. The expectation is taken with respect both to the distribution of source locations (assumed uniform throughout the room volume but at least a half-wavelength away from the walls) and to the distribution of microphone positions (assumed uniformly distributed on the sphere of radius r centered at the reference location). We note that W_f goes to zero at the receiver reference point, where $\widetilde{G}_f$ becomes equal to G_f.

The definition given above can be generalized: if multiplied by the squared amplitude of the source signal at frequency f, $|Q_f|^2$, equation (2) estimates the power of an error signal which is the difference between the equalizer output signal and desired (source) signal. From now on we assume that $Q_f = 1$ for all frequencies of interest, for simplicity.

Theorem 1: Let R denote the distance from the source to the reference location, r the displacement from the equalization point, $k = 2\pi/\lambda = 2\pi f/c$ the wavenumber, and γ the ratio of the direct to the reverberant sound energy density at the reference point. Then, the *mean squared error* at frequency f is

$$W_f \cong \frac{\gamma \frac{R}{2r} \ln |\frac{R+r}{R-r}| + 1}{\gamma + 1} - 2\frac{\sin(kr)}{kr} + 1 \tag{3}$$

with γ given by

$$\gamma = \frac{0.01V}{\pi R^2 (1 - \bar{\alpha}) T_{60}} \tag{4}$$

where R is in meters, V is the room volume (in m^3), T_{60} (in sec) is as defined earlier, and $\bar{\alpha}$ represents the average absorption coefficient (the fraction of incident acoustic power absorbed by the room surfaces). □

Proof: See the Appendix.

We make two observations regarding this result:

1. If the displacement from the equalization point is small compared to the source-to-microphone distance, the first term in (3) approaches 1, meaning that direct field component has negligible effect on the error signal.
2. The average error power depends merely on the *ratio* of the room volume and reverberation time. This property generalizes the results derived herein to rooms of different shapes, volumes, and reverberation times.

B. Equalization in a Diffuse-Field Region

Let us move on to the region of the room where the predominant contribution to the equilibrium sound pressure is the reverberant sound field, with only a negligible contribution by the direct sound field component. It is in this region that sound equalization is of particular importance because of reduced speech intelligibility due to multiple reflections from the walls. For nondirectional sound source, the minimum source-to-microphone distance required to "reach" the region where the sound field can be considered diffuse with respect to the direction of sound incidence is defined as $R_{\min} = 200/f_{\mathrm{S}}$ (in meters, for f_{S} given in Hz) [11].

In (3), the effect of the source-to-microphone distance on the error signal is expressed through the factor γ. At distances larger than $R_{\min}$ (but still at least a half-wavelength from the walls), we find that there is a fairly simple relationship between the error energy and the microphone displacement.

Corollary 1: Let r denote the displacement from the reference location and k the wave number. Then, in a diffuse-field region

$$W_f \cong 2 - 2\frac{\sin(kr)}{kr}. \tag{5}$$

□

In the case of broadband signals, the total reverberant energy in a given frequency range between f_{L} and f_{U} can be calculated by applying Parseval's theorem

$$W_{\mathrm{total}} = \int_{f_{\mathrm{L}}}^{f_{\mathrm{U}}} \left[2 - 2\frac{\sin(2\pi fr/c)}{(2\pi fr/c)}\right] df. \tag{6}$$

We note the following properties of the obtained solution:

1. The greatest amount of distortion can be expected for high frequencies, where the term $\sin(2\pi fr/c)/(2\pi fr/c)$ falls off rapidly with increasing distance from the equalization point.

2. The error energy is independent of the room size and reverberation time as long as these meet the criteria laid down earlier.

3. By reciprocity, the same result would be obtained with the microphone being fixed and source being moved from one point to another.

C. Variabilities of Inverse-Filter Parameters

So far we have been treating the room transfer function as though it had an ideal inverse. Actually, perfect reverberation removal in single-channel systems is not possible because the room transfer functions tend to have non-minimum phase components, and thus the exact inverse filter would consequently be either unstable or noncausal [13] and therefore unrealizable in practice. However, an approximate solution can be found which is based on the least squares error criterion (LSE method), with appropriate modelling delay [1].

In this section, we consider the residual reverberant power that results away from the equalization point, if the phase and magnitude response of an exact inverse are perturbed by a small random variable.

Let us denote an inexact inverse by A_f. We choose the following simple model:

$$A_f = [G_f(1+\varepsilon)e^{j\kappa}]^{-1}. \tag{7}$$

where ε and κ are independent random variables which model the mismatch in magnitude and phase due to the measurement and computational errors, and G_f is the room frequency response at the equalization point, as defined earlier. This particular form of the inverse filter characteristics A_f is chosen so that it is the exact inverse of G_f in the limiting case of $\varepsilon = \kappa = 0$.

Again we define the mean square value of the error signal at frequency f as

$$W_f = E\{|\widetilde{G}_f A_f - 1|^2\} \tag{8}$$

where now the expectation is taken also with respect to the distributions of ε and κ, assumed Gaussian with zero mean. If variances σ_ε^2, σ_κ^2 are small, we can get an approximate formula for the reverberant power at a single frequency, following the same reasoning as in the proof of (3).

Theorem 2: Let σ_ε^2, σ_κ^2 denote the variances of the magnitude and phase, respectively, of the measured frequency response at the equalization point. Then

$$\begin{aligned} W_f \;&\cong\; \frac{\gamma \frac{R}{2r} \ln \left|\frac{R+r}{R-r}\right| + 1}{\gamma + 1}(1+\sigma_\varepsilon^2) \\ &-\; 2\frac{\sin(kr)}{kr}(1-\sigma_\kappa^2/2) + 1. \end{aligned} \tag{9}$$

□

Proof: The proof is given in the Appendix.

The expression above shows the sensitivity of the spatial extent of equalization to variations in the magnitude and phase of an exact inverse filter characteristic. Based on (9), using an inexact inverse for reverberation removal will result in higher error levels compared to those of an exact inverse, at any distance from the equalization point. Hence, it is particularly desirable to have a good approximation at higher frequencies, due to increased sensitivity to position changes.

We remark that this result does not apply in the following situations:

– If an inverse filter is designed so as to achieve an approximate equalization for different source-receiver configurations over some region in a room, providing a more robust inverse over that region [6].

– If an approximate inverse is found as a solution of an LSE method in which the room transfer matrix is modified so as to have a 2-norm condition number (ratio of the largest to the smallest singular value of the transfer matrix) less than that of the unmodified room transfer matrix, resulting in a system less vulnerable to location changes [10], [14].

IV. Examples and Comparison

In order to validate the theoretical results derived in the preceding section, simulations have been performed for several rooms of different volumes and reverberation times, over a wide range of source-microphone positions. In this section, after a brief account of the method used for the simulations, we present an example related to the robustness of a single-channel equalization system when the source signal in a room is a pure tone. Then, for the purpose of comparison, we show some results concerning the robustness of a two-channel equalization system in the presence of a broad-band acoustic source, with microphones placed far apart.

A. Simulation Method

The source-to-microphone transfer function of the room can be simulated using the so-called image method [15]. The basic concept is the following: Suppose that the acoustic source sends out a short pulse of sound at the time $t = 0$; the first signal detected by the microphone is the direct sound, followed by the arrival of a sequence of reflected signals from the walls. These reflections may be considered as being due to image sources obtained by repeatedly mirroring the original source in all the walls of the room. Consequently, the signal received by the microphone can be represented as a superposition of the sound signals originating from the mirror-image sources, plus that of the original sound source.

In carrying out calculations, we use the following formula for the complex steady-state frequency response between the source and receiver:

$$G_f = \sum_{i=0}^{m} \beta^{n_i} \frac{e^{j(2\pi f/c)R_i}}{4\pi R_i} \tag{10}$$

where R_i is the distance between the microphone and the ith image source ($i = 0$ indexes the original source), β is the wall reflection coefficient, n_i denotes the number of reflections that the sound ray (assigned to the ith image) undergoes along its path from the source to the receiver, and m denotes the total number of images within a radius given by the speed of sound times the reverberation time. This means, we include only those images contributing to the sound pressure that normally result from the sound waves reflecting back and forth from the walls during the length of time a sound persists after it has left the source (for more details, see [15]).

B. Single-Channel Example

In the present example, a rectangular room with a volume of 128 m^3 is considered. We chose the room dimensions (length, width and height 6.4, 5, and 4 m, respectively) to satisfy the ratio (1 : 1.25 : 1.6) recommended in [16] for the best distribution of normal room modes. We assume that the reverberation time is frequency invariant and that all walls of the room have the same reflection coefficient β (the average absorption coefficient $\bar{\alpha} = 1 - \beta^2$).

First we calculated the frequency response at some fixed distance from the source placed in the interior of the room. Then, moving the *microphone* in an arbitrary direction, we calculated the error at several points along a straight path of one wavelength. A total of N simulations with different source-microphone positions were made, with both the source and microphone being displaced randomly between the runs. The average power of the error signal at frequency f was estimated by

$$\overline{W}_f = \frac{1}{N} \sum_{n=1}^{N} \left| \frac{\widetilde{G}_f^{(n)}}{G_f^{(n)}} - 1 \right|^2 \tag{11}$$

where $G_f^{(n)}$ denotes the complex room response at the reference location (associated with the nth simulation), $\widetilde{G}_f^{(n)}$ is the room response at some distance r from the reference point, and $N = 200$.

The simulations were then performed again in a similar way, with other values for excitation frequency f, reflection coefficient β, and distance R between the source and equalization point. The sensitivity to deviations in receiver position, as a result of one individual simulation, is illustrated in Fig. 1. In Fig. 2 and Fig. 3, the solid line represents the averaged trend of the error signal for many source and microphone locations; such an average is in very good agreement with what one calculates from (3) (dashed line). The results show that the zone of equalization, where more than 10 dB of reverberation reduction is obtained, is a sphere of a diameter of about $\lambda/7$. In this connection, it is interesting to note that it is approximately equal to the analogously defined zone of quiet generated by cancelling the pressure at a point in a diffuse sound field [17]−[19].

It is clear that single-channel equalization does not provide robust reverberation reduction and may even corrupt the overall system response away from the equalization point. As previously mentioned, the same conclusion holds if the source position changes with respect to the reference point. We are led then to ask whether multiple-channel systems involving more than one microphone are more robust with respect to source-location changes: some of the simulation results for two-channel equalization systems (an example is given in the following section) indicate that the degree of change in the equalized signal would actually be the same.

C. Two-Channel Example

In this example, we consider a rectangular room of volume 128 m^3, reverberation time 0.45 s, and reflection coefficient 0.84. Computer simulations are performed in the frequency range between 140 Hz and 2700 Hz. To exclude the possibility that the measured room responses are correlated, which would make the system more sensitive to the movement of the source point, we set the distance between equalization points to be larger than half the wavelength at the lowest frequency of interest. Again all observation points were in the interior region of the room, more than 1 m away from the walls. In the present investigation we confined our attention to the region above the Schroeder frequency (see Section II)

$$f_{\rm S} = 2000\left(\frac{0.45}{128}\right)^{1/2} = 119\text{Hz}. \tag{12}$$

First we calculated the room impulse response between the sound source and both equalization points from the sampled frequency response, using the inverse fast Fourier transform, with the sampling frequency twice the highest frequency in the range. An FIR filter of length 2048 taps was used to model the acoustics of the room. Working in the time domain and using results of the multiple input/output inverse filter theory (MINT) presented in [20], we found the exact inverse filter coefficients from the following system of equations

$$h_1(l) \circledast g_1(l) + h_2(l) \circledast g_2(l) = \delta(l) \tag{13}$$

for $l = 0, 1, \ldots, L-1$, $(L = 4096)$, where coefficient L denotes the length of the discrete linear convolution of two impulse response functions and l is the sample index. In the last expression, $g_1(l)$ and $g_2(l)$ are the room impulse responses, $h_1(l)$ and $h_2(l)$ are the inverse filter coefficients, and $\delta(l)$ is a delta function.

After, moving the *source* in an arbitrary direction, we calculated the room impulse responses at the two microphones, $\widetilde{g}_1(l)$ and $\widetilde{g}_2(l)$, for several source locations away from its reference point. The energy of the error signal for each source position was calculated from

$$\Psi_{\text{total}} = \sum_{l=0}^{L-1} [\delta(l) - h_1(l) \circledast \tilde{g}_1(l) - h_2(l) \circledast \tilde{g}_2(l)]^2 \tag{14}$$

which is a time-domain error function analogous to that given by (2) if taken over the entire frequency range. The average error energy, for 30 different spatial arrangements of the source and both equalization points, was calculated by:

$$\overline{\Psi}_{\text{total}} = \frac{1}{N} \sum_{n=1}^{N} \Psi_{\text{total}}^{(n)} \tag{15}$$

where $N = 30$ and $\Psi_{\text{total}}^{(n)}$ represents the error energy in the nth simulation, defined by (14).

Fig. 4 shows the average reverberant (error) energy as a function of the source displacement r, obtained from the simulation results. The same plot shows the error energy calculated using Parseval's theorem

$$W_{\text{total}} = \int_{f_{\text{L}}}^{f_{\text{U}}} \left[\frac{\gamma \frac{R}{2r} \ln |\frac{R+r}{R-r}| + 1}{\gamma + 1} - 2 \frac{\sin(2\pi f r / c)}{2\pi f r / c} + 1 \right] df \tag{16}$$

where f_{L} and f_{U} are the lower and upper bandwidth frequencies, and R is as defined earlier.

The simulation results reveal the following:

- The spatial extent of equalization in a reverberant sound field depends on the frequency spectrum of the source signal in the room. In the present example, the 10 dB zone of equalization produced in the vicinity of the source point is a sphere of a diameter of about 1.4 cm. This means that the zone of equalization is *slightly* reduced compared to that of a narrowband equalizer operating at f_{U} ($\lambda/7 \cong 1.8$ cm at $f_{\text{U}} = 2700$ Hz). The reason for this is that the error level drops with increasing λ and is significantly below -10 dB at the lowest frequency of interest.
- The formula (3) derived for a single-channel equalization system proves to be a fairly good approximation even in a two-channel case. We can take this result as an indication that increasing the number of channels used to get an *exact* inverse does not reduce sensitivity to the source position changes, at least in the case (as in the example) where the exact filters are of minimal length.

V. Conclusions

Standard equalization techniques succeed in removing unwanted reverberation from the signal received by a microphone, but only when the microphone and source have fixed positions. An open problem remained as to the robustness of such equalization if there are uncertainities or variations in the positions of the source and microphone. As shown in this paper, position changes on the order of one-tenth of the acoustic wavelength can cause significant degradation in the signal at the output of such a fixed equalization system. In this connection, it was found that as long as diffuse-field conditions are met, the room size, geometry, and reverberation time have no significant effect on the spatial extent of the zone of equalization where more than 10 dB of reverberation reduction is obtained. Outside of such a region, equalization is ineffective and may actually have performance worse than having no equalizer at all. These results put into question the value of attempting equalization in all but environments with very short reverberation times.

Although the changes in equalized response have been quantified in this paper in terms of mean-square deviation, it is difficult to draw any clear conclusion about the subjective effects, i.e., the distortion levels that can be allowed without audible degradation. Finally, the results of this paper should be taken with some caution at low frequencies, where the diffuse-field model may no longer be a good approximation.

Appendix

Proof of Theorem 1

We use the following relation between the frequency domain Green's function (frequency response function), G_f, and the complex sound pressure at some point in a room, P_f [21]:

$$P_f = -jk\rho c S_f G_f \tag{17}$$

where factors ρ and S_f, which will be defined later in this Appendix, are independent of the receiver location. Thus, we may write equation (2) in the form

$$W_f = E\{|\frac{\widetilde{P}_f}{P_f} - 1|^2\}. \tag{18}$$

This expression involves the assumption that the room response at the reference location is perfectly equalized. We can expand (18) to give

$$W_f = E\left\{ \frac{\widetilde{P}_f \widetilde{P}_f^*}{P_f P_f^*} - \frac{\widetilde{P}_f}{P_f} - \frac{\widetilde{P}_f^*}{P_f^*} + 1 \right\} \tag{19}$$

where the symbol * represents complex conjugate.

To calculate the expectation of each term in the sum in (19), we use the Taylor expansion [22], according to which if g is a function of *random* variables with mean values $E\{x_i\} = \bar{x}_i$, $i = 1, \ldots, n$, then $g(x_1, x_2, \ldots, x_n)$, which we write as $g(x)$ for brevity, can be expressed in the form: $g(x) = g(\bar{x}) + \sum_{i=1}^{n} g_i'(\bar{x})(x_i - \bar{x}_i) + \hat{g}(x)$, where $\hat{g}$ is a function of order 2, i.e., all its partial derivatives up to the first order vanish at $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n)$. Thus, to a $0th$ order of approximation, $E\{g(x)\} = g(\bar{x})$.

Let us first multiply and divide the second and the third term in (19) by P_f^* and P_f, respectively. With the definition given above, we find

$$W_f \cong \frac{E\{\widetilde{P}_f \widetilde{P}_f^*\}}{E\{P_f P_f^*\}} - \frac{E\{\widetilde{P}_f P_f^*\}}{E\{P_f P_f^*\}} - \frac{E\{\widetilde{P}_f^* P_f\}}{E\{P_f^* P_f\}} + 1. \tag{20}$$

The total sound pressure at frequency f, at some point away from the source, may be expressed as

$$P_f = P_{f\mathrm{d}} + P_{f\mathrm{r}} \tag{21}$$

where $P_{f\mathrm{d}}$ and $P_{f\mathrm{r}}$ are the direct and reverberant sound pressure components, respectively. Under the same conditions as in the Section II, the direct and reverberant sound pressure are uncorrelated at the point of observation and, therefore, all cross terms in the sum in (20) will vanish, leaving only the following factors

$$\begin{aligned} W_f \;\cong\; & \frac{E\{\widetilde{P}_{f\mathrm{d}} \widetilde{P}_{f\mathrm{d}}^* + \widetilde{P}_{f\mathrm{r}} \widetilde{P}_{f\mathrm{r}}^*\}}{E\{P_{f\mathrm{d}} P_{f\mathrm{d}}^* + P_{f\mathrm{r}} P_{f\mathrm{r}}^*\}} - \frac{E\{\widetilde{P}_{f\mathrm{d}} P_{f\mathrm{d}}^* + \widetilde{P}_{f\mathrm{r}} P_{f\mathrm{r}}^*\}}{E\{P_{f\mathrm{d}} P_{f\mathrm{d}}^* + P_{f\mathrm{r}} P_{f\mathrm{r}}^*\}} \\ & - \frac{E\{P_{f\mathrm{d}} \widetilde{P}_{f\mathrm{d}}^* + P_{f\mathrm{r}} \widetilde{P}_{f\mathrm{r}}^*\}}{E\{P_{f\mathrm{d}} P_{f\mathrm{d}}^* + P_{f\mathrm{r}} P_{f\mathrm{r}}^*\}} + 1. \end{aligned} \tag{22}$$

At this point we will define the ratio of the direct to the reverberant sound energy density [23], in the form equivalent to that given by (4):

$$\gamma \equiv \frac{E_\mathrm{d}}{E_\mathrm{r}} = \frac{S\bar{\alpha}}{16\pi R^2(1-\bar{\alpha})} \tag{23}$$

where R denotes the distance from the source to the observation point, $\bar{\alpha}$ the average absorption coefficient, and S the wall area of the room. In (23), absorption in the air is neglected for simplicity.

The reverberant-field mean-square pressure can be defined as

$$E\{P_{f\mathrm{r}}P^*_{f\mathrm{r}}\} = E\{\widetilde{P}_{f\mathrm{r}}\widetilde{P}^*_{f\mathrm{r}}\} = \frac{4\rho c\Pi(1-\bar{\alpha})}{S\bar{\alpha}} \tag{24}$$

where Π is the power of the acoustic source, and ρ is the density of air in the room.

With (23) and (24), returning to (22), we find

$$\begin{aligned} W_f \;\cong\; & \frac{E\{\widetilde{P}_{f\mathrm{d}}\widetilde{P}^*_{f\mathrm{d}}\}\dfrac{S\bar{\alpha}}{4\rho c\Pi(1-\bar{\alpha})}+1}{\gamma+1} \\ & - \frac{E\{\widetilde{P}_{f\mathrm{d}}P^*_{f\mathrm{d}}\}\dfrac{S\bar{\alpha}}{4\rho c\Pi(1-\bar{\alpha})}+\dfrac{E\{\widetilde{P}_{f\mathrm{r}}P^*_{f\mathrm{r}}\}}{E\{P_{f\mathrm{r}}P^*_{f\mathrm{r}}\}}}{\gamma+1} \\ & - \frac{E\{P_{f\mathrm{d}}\widetilde{P}^*_{f\mathrm{d}}\}\dfrac{S\bar{\alpha}}{4\rho c\Pi(1-\bar{\alpha})}+\dfrac{E\{P_{f\mathrm{r}}\widetilde{P}^*_{f\mathrm{r}}\}}{E\{P_{f\mathrm{r}}P^*_{f\mathrm{r}}\}}}{\gamma+1} + 1. \end{aligned} \tag{25}$$

Next we calculate $E\{\widetilde{P}_{f\mathrm{d}}\widetilde{P}^*_{f\mathrm{d}}\}$ starting from the following equations: The free-space Green's function is defined as

$$g_f = \frac{1}{4\pi R}e^{jkR}. \tag{26}$$

At a distance R from the source point, for a given source strength S_f, the direct sound pressure is of the form [21]

$$P_{f\mathrm{d}} = -jk\rho c S_f g_f. \tag{27}$$

Some distance r away from the equalization point, at an angle θ to the reference direction, the sound pressure is given by

$$\widetilde{P}_{f\mathrm{d}} = -jk\rho c S_f \widetilde{g}_f. \tag{28}$$

We may write thus

$$E\{\widetilde{P}_{f\mathrm{d}}\widetilde{P}^*_{f\mathrm{d}}\} = (k\rho c)^2|S_f|^2E\{\widetilde{g}_f\widetilde{g}^*_f\}. \tag{29}$$

One further substitution is possible in equation (29) [21]

$$E\{\widetilde{P}_{f\mathrm{d}}\widetilde{P}^*_{f\mathrm{d}}\} = 4\Pi\rho c\pi E\{\widetilde{g}_f\widetilde{g}^*_f\}. \tag{30}$$

The function $\widetilde{g}_f$ in (30) can be defined by using the cosine law

$$\widetilde{g}_f = \frac{1}{4\pi(R^2+r^2-2Rr\cos\theta)^{1/2}} e^{jk(R^2+r^2-2Rr\cos\theta)^{1/2}}. \tag{31}$$

Because all directions of the microphone displacement are assumed to be equally probable over the solid angle, $\cos\theta$ is distributed uniformly over the interval -1 to $+1$, and the expectation of $\widetilde{g}_f\widetilde{g}^*_f$ can be found as:

$$\begin{aligned} E\{\widetilde{g}_f\widetilde{g}^*_f\} &= \frac{1}{2}\int_{-1}^{1} \frac{d(\cos\theta)}{(4\pi)^2(R^2+r^2-2Rr\cos\theta)} \\ &= \frac{1}{(4\pi)^2 2Rr}\ln\left|\frac{R+r}{R-r}\right|. \end{aligned} \tag{32}$$

With (30) and (32), one obtains:

$$E\{\widetilde{P}_{f\mathrm{d}}\widetilde{P}^*_{f\mathrm{d}}\} = \frac{\Pi\rho c}{4\pi R^2}\frac{R}{2r}\ln\left|\frac{R+r}{R-r}\right|. \tag{33}$$

The limiting form for this solution (when r approaches zero) agrees with the basic equation for the direct-field mean-square pressure at a distance R from the omnidirectional sound source.

Having determined the first term in (25), we proceed with a similar mathematical analysis to determine $E\{\widetilde{P}_{f\mathrm{d}}P^*_{f\mathrm{d}}\}$. In this case, we find

$$\begin{aligned} E\{\widetilde{g}_f g^*_f\} &= \frac{1}{2}\int_{-1}^{1} \frac{e^{jk(\sqrt{R^2+r^2-2Rr\cos\theta}-R)}d(\cos\theta)}{(4\pi)^2 R\sqrt{R^2+r^2-2Rr\cos\theta}} \\ &= \frac{1}{(4\pi R)^2}\frac{\sin(kr)}{kr} \end{aligned} \tag{34}$$

leading to

$$E\{\widetilde{P}_{f\mathrm{d}}P^*_{f\mathrm{d}}\} = \frac{\Pi\rho c}{4\pi R^2}\frac{\sin(kr)}{kr}. \tag{35}$$

Finally, returning again to the expression (25), we utilize the well-known formula for the normalized correlation function of the complex sound pressure amplitude at two points separated by a distance r in space [24], [25]

$$\frac{E\{\widetilde{P}_{f\mathrm{r}}P^*_{f\mathrm{r}}\}}{E\{P_{f\mathrm{r}}P^*_{f\mathrm{r}}\}} = \frac{\sin(kr)}{kr}. \tag{36}$$

Because the second term in (25) turns out to be a real function of r, the same result must be valid for the third term in (25).

These results can now be inserted in (25) to obtain an approximate solution to the power of the error signal at single frequency.

Proof of Theorem 2

Equation (8) may be expanded to give

$$\begin{aligned} W_f &\cong E\{\frac{\widetilde{G}_f^*\widetilde{G}_f}{G_f G_f^*}(1-\varepsilon)^2 - \frac{\widetilde{G}_f}{G_f}(1-\varepsilon)e^{-j\kappa} \\ &- \frac{\widetilde{G}_f^*}{G_f^*}(1-\varepsilon)e^{j\kappa} + 1\}. \end{aligned} \tag{37}$$

By an analysis analogous to that in the Proof of Theorem 1, one can conclude that $E\{\widetilde{G}_f/G_f\} = E\{\widetilde{G}_f^*/G_f^*\}$. Thus, we may write

$$\begin{aligned} W_f &\cong E\{(1-\varepsilon)^2\}E\{\frac{\widetilde{G}_f^*\widetilde{G}_f}{G_f G_f^*}\} \\ &- 2E\{(1-\varepsilon)\cos\kappa\}E\{\frac{\widetilde{G}_f}{G_f}\} + 1. \end{aligned} \tag{38}$$

The ε and κ are independent random variables with mean value $E\{\varepsilon\} = E\{\kappa\} = 0$ and variances $\sigma_\kappa^2 \ll 1$, $\sigma_\varepsilon^2 \ll 1$. Because both ε and κ are assumed to be small, we can use a series expansion to find:

$$E\{\cos\kappa\} \cong E\{1 - \kappa^2/2\} = 1 - \sigma_\kappa^2/2. \tag{39}$$

Substituting this approximate expression into (38), and collecting the terms as in (3), one obtains an approximate solution to the mean-squared error as given by (9).

Acknowledgment

The authors would like to thank Gary Elko for motivating this problem and David Lubman for advice on the statistical room acoustics literature. Also, the authors are thankful to reviewers for their helpful suggestions and discussions.

References

[1] S.J. Elliott and P.A. Nelson, "Multiple-point equalization in a room using adaptive digital filters," *J.Audio Eng.Soc.*, vol. 37, no. 11, pp. 899–907, Nov. 1989.

[2] L.M. van de Kerkhof and W.J.W. Kitzen, "Tracking of a time-varying acoustic impulse response by an adaptive filter," *IEEE Trans.Signal Processing*, vol. 40, no. 6, pp. 1285–1294, June 1992.

[3] M.R. Schroeder, "Statistical parameters of the frequency response curves of large rooms," *J.Audio Eng.Soc.*, vol. 35, no. 5, pp. 299–305, May 1987.(Originally published in *Acustica*, vol.4, pp. 594–600, 1954.)

[4] J. Homer, R.R. Bitmead, and I. Marlees, "Quantifying the effect of dimension on the convergence rate of the LMS adaptive FIR estimator," *IEEE Trans.Signal Processing*, vol. 46, no. 10, pp. 2611–2615, Oct. 1998.

[5] J. Mourjopoulos, "On the variation and invertibility of room impulse response functions," *J.Sound Vib.*, vol. 102, no. 2, pp. 217–228, 1985.

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

[7] Y. Haneda, S. Makino, and Y. Kaneda, "Common acoustical pole and zero modeling of room transfer functions," *IEEE Trans.Speech, Audio Processing*, vol. 2, no. 2, pp. 320–328, Apr. 1994.

[8] D. Lubman, "Fluctuations of sound with position in a reverberant room," *J.Acoust.Soc.Am.*, vol. 44, no. 6, pp. 1491–1502, 1968.

[9] A.H. Benade, "From instrument to ear in a room: Direct or via recording," *J.Audio Eng.Soc.*, vol. 33, no. 4, pp. 218–233, Apr. 1985.

[10] F. Asano, Y. Suzuki, and T. Sone, "Sound equalization using derivative constraints," *Acustica*, vol. 82, pp. 311–320, 1996.

[11] M.R. Schroeder, "Frequency correlation functions of frequency responses in rooms," *J.Acoust.Soc.Am.*, vol. 34, no. 12, pp. 1819–1823, 1962.

[12] R.V. Waterhouse, "Statistical properties of reverberant sound fields," *J.Audio Eng.Soc.*, vol. 43, no. 6, pp. 1436–1444, 1968.

[13] S.T. Neely and J.B. Allen, "Invertibility of a room impulse response," *J.Acoust.Soc.Am.*, vol. 66, No. 1, pp. 165–169, 1979.

[14] P.A. Nelson and S.J. Elliott, *Active Control of Sound*. London: Academic Press, 1992, pp. 384–387.

[15] J.B.Allen and D.A.Berkley, "Image method for efficiently simulating small-room acoustics," *J.Acoust.Soc.Am.*, vol. 65, no. 4, pp. 943–950, 1979.

[16] V. Knudsen and C. Harris, *Acoustical Designing in Architecture*. New York: Wiley, 1950, pp. 402–403.

[17] S.J. Elliott, P. Joseph, A.J. Bullmore and P.A. Nelson, "Active cancellation at a point in a pure tone diffuse sound field," *J.Sound Vib.*, vol. 120, no. 1, pp. 183–189, 1988.

[18] P. Joseph, S.J. Elliott and P.A. Nelson, "Near field zones of quiet," *J.Sound Vib.*, vol. 172, no. 5, pp. 605–627, 1994.

[19] P. Joseph, S.J. Elliott and P.A. Nelson, "Statistical aspects of active control in harmonic enclosed sound fields," *J.Sound Vib.*, vol. 172, no. 5, pp. 629–655, 1994.

[20] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 2, pp. 145–152, 1988.

[21] P.M. Morse and K.U. Ingard, *Theoretical Acoustics*. London: McGraw-Hill, 1968, pp. 311,555,582,591.

[22] M. Kendall and A. Stuart, *The advanced theory of statistics (vol.1)*, 4th ed. London: Charles Griffin Co.Ltd, 1976, p. 246.

[23] L.L. Beranek, *Acoustics.* American Institute of Physics, for the Acoustical Society of America, 1996, pp. 312–314.

[24] R.K. Cook, R.V. Waterhouse, R.D. Berendt, S. Edelman and M.C. Thompson, "Measurement of correlation coefficients in reverberant sound fields," *J.Acoust.Soc.Am.*, vol. 27, no. 6, pp. 1072–1077, 1955.

[25] K.J. Ebeling and K. Freudenstein, "On spatial averaging of sinusoidal sound intensities in reverberant rooms: Exact Distribution Functions," *Acustica*, vol. 46, pp. 18–25, 1980.

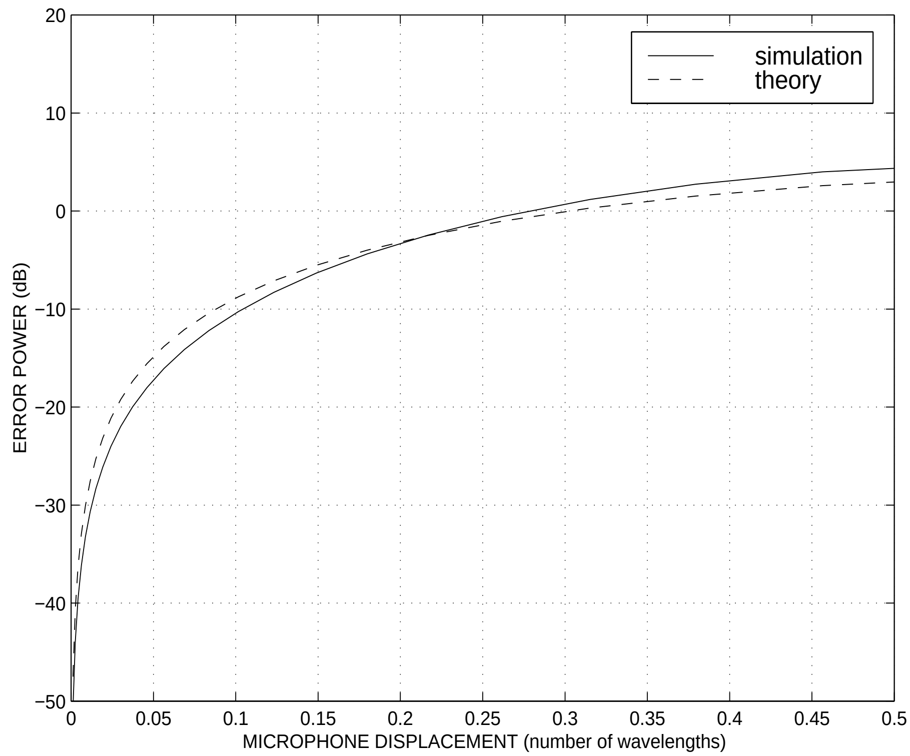

Fig. 1. Single-channel residual reverberant power as a function of displacement from the equalization point, at frequency 2000 Hz. An individual, narrowband image model simulation of a room with dimensions $(6.4 \times 5 \times 4)$ m^3, wall reflection coefficients 0.84, and direct-to-reverberant energy ratio -8.4 dB (source-to-microphone distance $R = 3$ m).

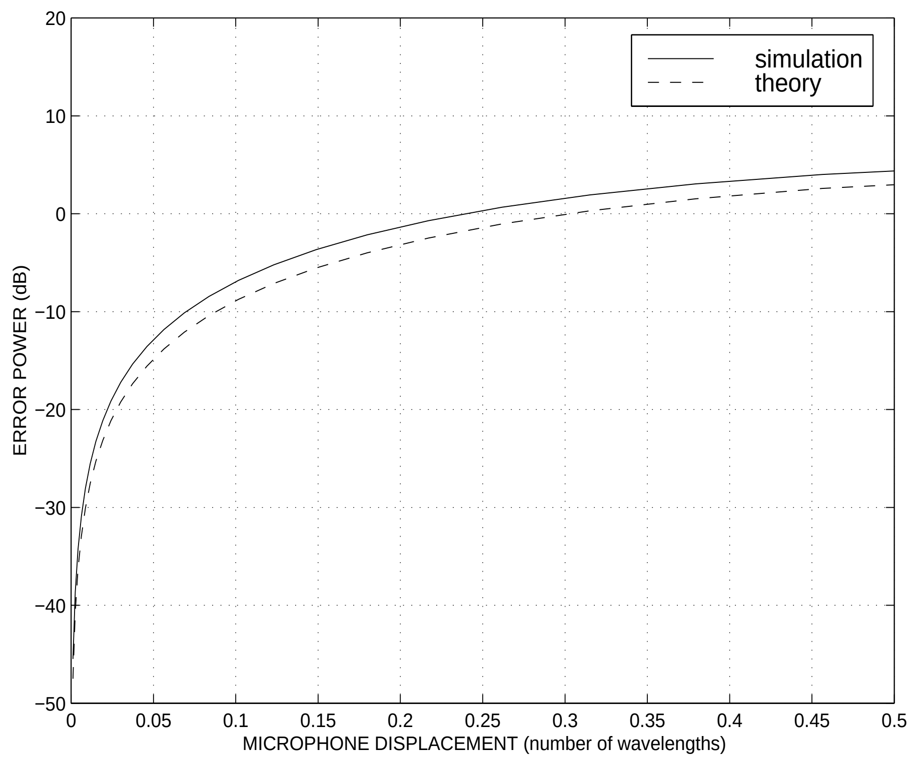

Fig. 2. Single-channel residual reverberant power as a function of displacement from the equalization point, at frequency 2000 Hz. Narrowband image model simulation (spatial average) of a room with dimensions $(6.4 \times 5 \times 4)$ m^3, wall reflection coefficients 0.84, and direct-to-reverberant energy ratio -8.4 dB (source-to-microphone distance $R = 3$ m).

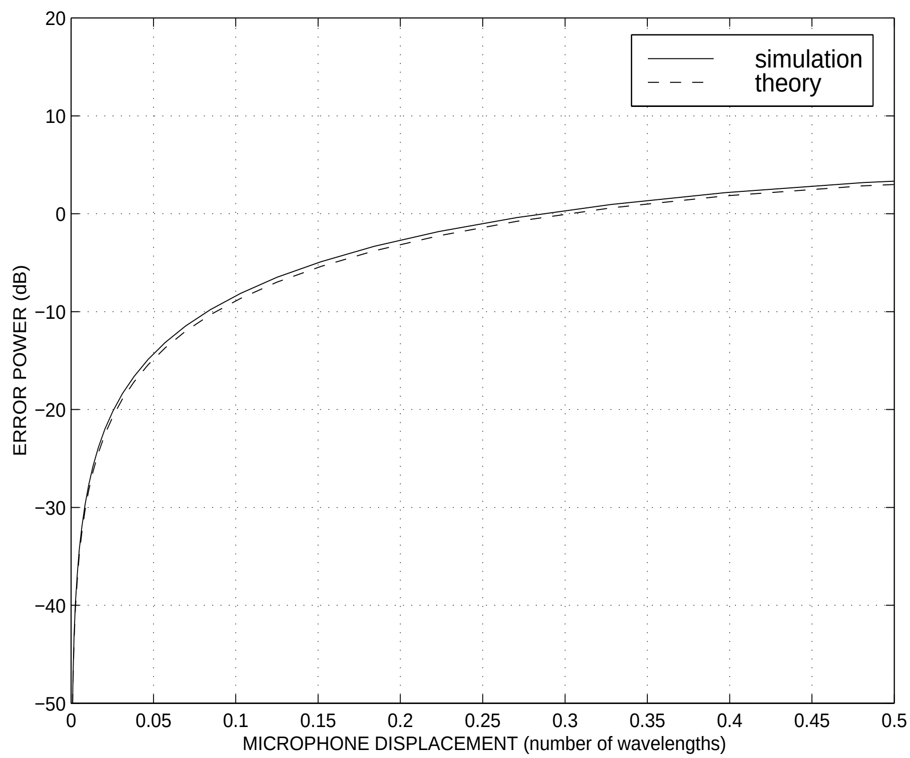

Fig. 3. Single-channel residual reverberant power as a function of displacement from the equalization point, at frequency 800 Hz. Narrowband image model simulation (spatial average) of a room with dimensions $(6.4 \times 5 \times 4)$ m^3, wall reflection coefficients 0.9, and direct-to-reverberant energy ratio -1.4 dB (source-to-microphone distance $R = 1$ m).

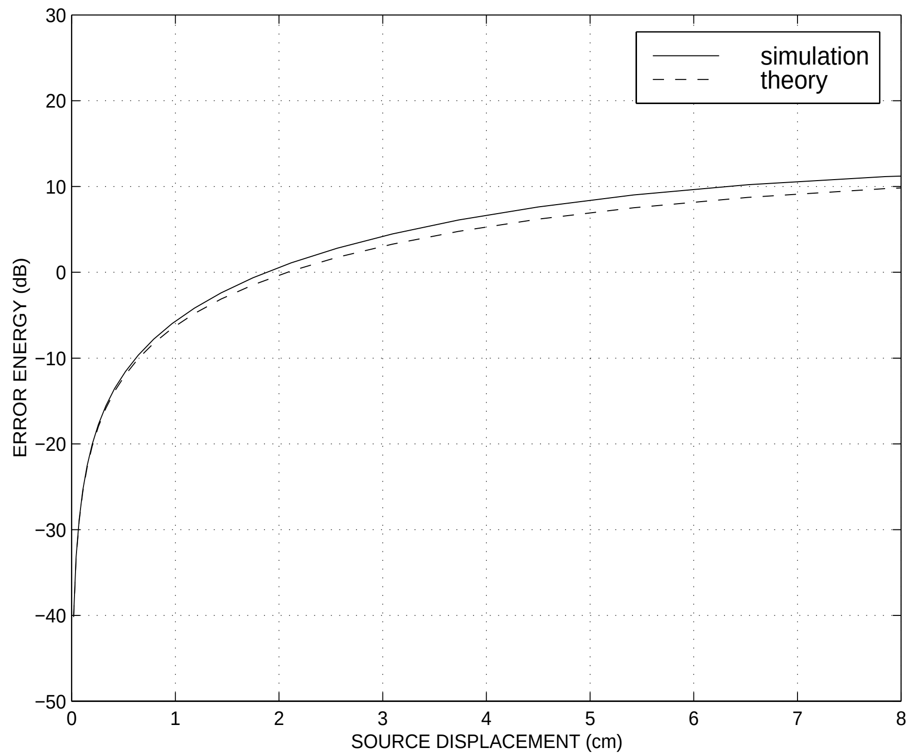

Fig. 4. Two-channel residual reverberant energy in the frequency range between 140 and 2700 Hz, as a function of the source displacement. Broadband image model simulation (spatial average) of a room with dimensions $(6.4 \times 5 \times 4)$ m^3, wall reflection coefficients 0.84, and direct-to-reverberant energy ratio -8.4 dB (source-to-microphone distance $R = 3$ m, for both equalization points).