Nonminimum-Phase Equalization and Its Subjective Importance in Room Acoustics

Biljana D. Radlović, *Student Member, IEEE,* and Rodney A. Kennedy, *Member, IEEE*

***Abstract*—This paper investigates the perceptual significance of residual phase distortion due to an approximate equalization of the nonminimum-phase room response from a sound source to a microphone in a reverberant room. It is shown that disrupted phase relationships introduced by a minimum-phase equalization filter may have a detrimental effect on perceived sound quality. The subjective assessment of phase distortion on speech signals is related to an objective error criterion, newly introduced in this paper. An alternative approach to the minimum-phase/allpass decomposition based on iterative flattening of the room transfer function (RTF) magnitude is also presented, which overcomes potential numerical problems and provides more insight into subjective aspects of magnitude and phase equalization in the reduction of acoustic reverberation. Factors contributing to the results and practical implications for equalization are discussed.**

***Index Terms*—Architectural acoustics, cepstral analysis, deconvolution, nonminimum-phase equalization, phase distortion, phase equalizers.**

I. Introduction

THE transmission between a sound source and a microphone placed some distance away in a reverberant room is characterized by linear distortion of the amplitude and phase, caused by reflections from the room boundaries. In many communication situations, such as hands-free telephony, reflections from the room walls and other surfaces produce reverberant and echoed speech, which is of hollow, barrel-like quality and is less intelligible to the listener at the far end [1]. Compensation for irregularities in the source-to-microphone transfer function can be achieved by convolving the received signal with the impulse response of an inverse (equalization) filter, yielding a signal of the same quality as if the microphone were placed in close proximity to the sound source [2], [3].

The perceptual aspects of phase distortion have been dealt with exhaustively in the literature, though false explanations have often been given—back to the famous "phase law" postulated by Ohm about 150 years ago [4]. According to this law, the ear is insensitive to phase relations between spectral components, and thus, the subjective effects of a complex sound depend solely on its amplitude spectrum. The subsequent work has convincingly demonstrated that phase-relations of various harmonic components may have dramatic subjective effect on perceived sound quality [5]–[7]. As reported in [7], one can even play simple melodies only by varying the relative phases of selected harmonics of an acoustic waveform.

In the last two decades, an analogous question has been raised regarding the audibility of phase distortion introduced by audio system components in the recording/playback chain [8]–[12]. Most of the work done on this subject was able to arrive at a threshold of audibility for group delay distortion, indicating that even small deviations from phase linearity may cause audible effects on suitably chosen artificial test signals, such as bandlimited rectangular pulses [10], tone bursts [11], or broadband impulsive signals (clicks) [12]. No data, however, are available for *speech* signals when subjected to the same amounts of phase distortion in a reverberant environment.

There is widespread agreement in the literature that the listener's ear is less phase sensitive in a reverberant room than it is in an anechoic environment; it is asserted in [13] that "amplitude equalization (of the distortions caused in the electrical or electroacoustical system) may improve the quality of the sound transmission drastically whereas an additional phase equalization does not have an audible effect." This idea is based on the results of listening tests with two tones, differing only in their phase spectra.

In equalization of room acoustics, we are faced with the apparent contradiction that an approximate inversion of the room response function, in the form of magnitude compensation only, may lead to further degradation of the equalized speech signal. This situation arises when magnitude equalization of a mixed-phase response is accomplished by a minimum-phase inverse filter [14]. The phase effects in [14] are attributed to spikes in the group delay function of the allpass component remaining after equalization; our present understanding of these effects, however, is still incomplete and conjectural.

The main objective of this paper is to resolve the perceptual relevance of residual phase distortion in case of single-point equalization of a room response function. We use both perceptual and objective error criteria to assess the performance of an equalization system formed by a minimum-phase (magnitude) equalizer in cascade with an allpass (phase) equalization filter. Through our experimental results, we establish a phase- and magnitude-dependent criterion for estimating objectively the degree of phase distortion that causes an audible degradation of a partially equalized speech signal.

The paper is organized as follows. A brief theoretical background of the inversion (deconvolution) problem of nonminimum-phase room responses is given in Section II. Section III presents a new cepstrum-based algorithm for achieving a minimum-phase inverse, based on iterative flattening of the room

Manuscript received June 3, 1999; revised March 17, 2000. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Dennis R. Morgan.

The authors are with the Telecommunications Engineering Group, Research School of Information Sciences and Engineering, Institute of Advanced Studies, Australian National University, Canberra ACT 0200, Australia (e-mail: rodney.kennedy@anu.edu.au; biljana@syseng.anu.edu.au).

Publisher Item Identifier S 1063-6676(00)09270-1.

transfer function (RTF) magnitude spectrum. A phase-equalization procedure is outlined in Section IV. In Section V-A, the effectiveness of several deconvolution algorithms is examined using both time-domain and frequency-domain error criteria. Section V-B investigates the effect of magnitude and phase equalization on subjective perception of an equalized speech signal. Sections V-C and V-D show the results in connection with the present study, derived from a set of experimental observations. Since sound equalization is normally carried out using digital filters, discrete-time signals, discrete Fourier transforms (DFTs), and the like are used throughout the paper.

II. Inverse-Filtering Concept

Let $h(n)$ be a causal and stable, nonminimum-phase impulse response between an acoustic source and a microphone placed some distance apart in a reverberant room. If the requirement for signal processing is that the waveform of the input (source) signal remains unchanged after passing through the equalized room transmission system, we may think of the equalization problem as that of finding an inverse impulse response function $p(n)$, such that

$$h(n) \otimes p(n) = \delta(n) \tag{1}$$

where

- n time index;
- $\otimes$ discrete linear convolution;
- $\delta(n)$ unit sample sequence ($\delta(n) = 1$ for $n = 0$, and $\delta(n) = 0$ for any other n).

Since the room response $h(n)$ is nonminimum phase, an inverse response, calculated from (1), is bound to be either unstable or noncausal, and therefore unrealizable in practice [14]. Such an inverse can be rendered causal and stable by introducing a delay factor, which compensates for the mixed-phase properties of the RTF [2].

The equalization procedure presented here makes use of a minimum-phase inverse filter for correction of RTF magnitude distortion, in cascade with an allpass (phase) equalization filter. If sufficiently large delay is introduced by the system, nearly perfect equalization of a room response function can be achieved using causal and stable finite impulse response (FIR) filters.

III. Magnitude Equalization

A few relevant points are set forth here as background to the following subsection, which proposes an alternative cepstrum-based approach for extraction and equalization of the minimum-phase component.

Let us denote the DFT of the sequence $h(n)$ by $H(k)$, where k is a frequency index. A nonminimum-phase room impulse response $h(n)$ can be represented as

$$h(n) = h_{\mathrm{mp}}(n) \otimes h_{\mathrm{ap}}(n) \tag{2}$$

where $h_{\mathrm{mp}}(n)$ is a minimum-phase sequence, such that its DFT has modulus $|H_{\mathrm{mp}}(k)| = |H(k)|$, and where $h_{\mathrm{ap}}(n)$ is an allpass sequence, such that $|H_{\mathrm{ap}}(k)| = 1$.

The convolved signals $h_{\mathrm{mp}}(n)$ and $h_{\mathrm{ap}}(n)$ can be separated by homomorphic transformation, which produces corresponding complex cepstra $\hat{h}_{\mathrm{mp}}(n)$ and $\hat{h}_{\mathrm{ap}}(n)$, combined by simple algebraic addition [15, pp. 768–771]. Causality of the complex cepstrum $\hat{h}_{\mathrm{mp}}(n)$, or, equivalently, the constraint that both the poles and zeros of the z-transform of the sequence $h_{\mathrm{mp}}(n)$ lie inside the unit circle, implies that the magnitude and phase of the frequency response $H_{\mathrm{mp}}(k)$ are related through the Hilbert transform relationship and thus the minimum-phase component can be recovered by knowing only the magnitude of the Fourier transform of the sequence $h(n)$ [15, pp. 664–674]. The effect of magnitude distortion can be perfectly removed in practice by convolving $h(n)$ with the minimum-phase inverse $g_{\mathrm{mp}}(n)$, that is

$$h_{\mathrm{ap}}(n) = h(n) \otimes g_{\mathrm{mp}}(n) \tag{3}$$

where $g_{\mathrm{mp}}(n)$ represents the inverse DFT of $G_{\mathrm{mp}}(k)$, given by $G_{\mathrm{mp}}(k) = 1/H_{\mathrm{mp}}(k)$.

In the case of ill-conditioned inversion problems, the conventional cepstral approach outlined above can lead to an inaccurate estimate of $h_{\mathrm{ap}}(n)$: even though all the poles and zeros of $H_{\mathrm{mp}}(k)$ lie inside the unit circle which renders it suitable for inversion, some of the zeros of $H(k)$, reflected to their conjugate reciprocal locations inside the unit circle in forming $H_{\mathrm{mp}}(k)$, as well as the original minimum-phase zeros of $H_{\mathrm{mp}}(k)$, may lie close to the unit circle and therefore cause a minimum-phase inverse of very long duration. This in turn implies that, given the finite-point nature of DFT algorithms, computation of the allpass sequence $h_{\mathrm{ap}}(n)$ may be inaccurate, due to time aliasing in the circular convolution of the sequences $h(n)$ and $g_{\mathrm{mp}}(n)$ [15, pp. 542–546], [16].

The limitations of the conventional cepstral analysis referred to above motivate our approach in the following subsection.

A. Minimum-Phase/Allpass Decomposition: Novel Approach

An alternative approach proposed in this section derives from the basic concept of the standard homomorphic technique [14], [15, pp. 768–772]. The algorithm performs an iterative extraction of the minimum-phase component, and iterative flattening of the RTF magnitude spectrum, by using the additive properties of the complex cepstrum. The aim is to achieve a sequential flattening of the log-magnitude response that enables us to monitor at each step the characteristics of a remaining mixed-phase component, while at the same time constraining the numerical complexity to within some predefined limit in each iteration.

The algorithm proposed here is based on the following decomposition of the complex cepstrum

$$\hat{h}(n) = \hat{h}_{\mathrm{mp}}^{(0)}(n) + \hat{h}_{\mathrm{mp}}^{(1)}(n) + \cdots + \hat{h}_{\mathrm{mp}}^{(L-1)}(n) + \hat{\tilde{h}}_{\mathrm{ap}}(n) \tag{4}$$

or, in the time domain, the decomposition

$$h(n) = h_{\mathrm{mp}}^{(0)}(n) \otimes h_{\mathrm{mp}}^{(1)}(n) \otimes \cdots \otimes h_{\mathrm{mp}}^{(L-1)}(n) \otimes \tilde{h}_{\mathrm{ap}}(n) \tag{5}$$

followed by successive deconvolution with $g_{\mathrm{mp}}^{(0)}(n)$, $g_{\mathrm{mp}}^{(1)}(n)$, $\cdots$, $g_{\mathrm{mp}}^{(L-1)}(n)$, where $g_{\mathrm{mp}}^{(l)}(n)$, $(l = 0, 1, 2, \cdots, L-1)$, is the inverse DFT of $G_{\mathrm{mp}}^{(l)}(k)$, given by $G_{\mathrm{mp}}^{(l)}(k) = 1/H_{\mathrm{mp}}^{(l)}(k)$, L is a nonnegative integer, and $\tilde{H}_{\mathrm{ap}}(k)$ denotes the DFT of

the mixed-phase impulse response $\tilde{h}_{\rm ap}(n)$ resulting from an approximate minimum-phase/allpass separation.

The procedure may be stated as follows.

1) Compute the DFT of $h(n)$, as

$$H(k) = \sum_{n=0}^{N-1} h(n) e^{-j(2\pi/N)kn} \tag{6}$$

where N represents the number of DFT points (assumed to be an even integer), and $k = 0, 1, 2, \cdots, N-1$.

2) Compute the even part of the complex cepstrum of the sequence $h(n)$, as

$$\hat{h}_{\rm e}(n) = \frac{1}{N} \sum_{k=0}^{N-1} \log|H(k)| e^{j(2\pi/N)kn} \tag{7}$$

where the complex cepstrum of the sequence $h(n)$ is defined as the inverse DFT of $\log[H(k)]$.

3) Calculate the corresponding complex cepstrum of the minimum-phase sequence $h_{\rm mp}(n)$, as

$$\hat{h}_{\rm mp}(n) = \begin{cases} \hat{h}_{\rm e}(n), & n = 0,\ N/2, \\ 2\hat{h}_{\rm e}(n), & 1 \le n < N/2, \\ 0, & N/2 < n \le N-1. \end{cases} \tag{8}$$

4) Set the iteration index l to 0.

5) Calculate the complex cepstrum $\hat{h}_{\rm mp}^{(l)}(n)$ corresponding to a minimum-phase factor $h_{\rm mp}^{(l)}(n)$, as

$$\hat{h}_{\rm mp}^{(l)}(n) = \hat{h}_{\rm mp}(n)/2^{l+1}. \tag{9}$$

6) Compute the DFT of $\hat{h}_{\rm mp}^{(l)}(n)$, as

$$\hat{H}_{\rm mp}^{(l)}(k) = \hat{H}_{\rm mp}(k)/2^{l+1} \tag{10}$$

where

$$\hat{H}_{\rm mp}(k) = \sum_{n=0}^{N-1} \hat{h}_{\rm mp}(n) e^{-j(2\pi/N)kn}. \tag{11}$$

7) Compute the minimum-phase part $H_{\rm mp}^{(l)}(k)$

$$H_{\rm mp}^{(l)}(k) = \exp\left[\hat{H}_{\rm mp}^{(l)}(k)\right]. \tag{12}$$

8) Calculate the remaining mixed-phase part

$$H^{(l+1)}(k) = H^{(l)}(k)/H_{\rm mp}^{(l)}(k), \tag{13}$$

and the corresponding time–domain response function

$$h^{(l+1)}(n) = h^{(l)}(n) \otimes g_{\rm mp}^{(l)}(n) \tag{14}$$

where $H^{(0)}(k) = H(k)$, and $h^{(0)}(n) = h(n)$.

9) Increase l by one and go to step 5).

The iteration process is repeated until the amplitude-spectrum distortion is reduced below some desired threshold level. For the purposes of our analysis it suffices to assume that an optimum number of iterations L has been chosen for computations.

The operations (9), (12), and (13) in the iterative procedure imply that filtering the room response with the minimum-phase inverse obtained from $\hat{h}_{\rm mp}^{(0)}(n)$, would mean halving the deviations in the log-magnitude spectrum and altering the phase by $(1/2)\arg[H_{\rm mp}(k)]$. This can verified by the following set of operations:

$$\begin{aligned}
(10) &\Longrightarrow \hat{H}_{\rm mp}^{(0)}(k) = (1/2)\hat{H}_{\rm mp}(k), \\
(12) &\Longrightarrow H_{\rm mp}^{(0)}(k) = \{\exp[\hat{H}_{\rm mp}(k)]\}^{1/2} = \{H_{\rm mp}(k)\}^{1/2}, \\
(13) &\Longrightarrow H^{(1)}(k) = H(k)/[|H_{\rm mp}(k)|^{1/2} e^{j\arg[H_{\rm mp}(k)]/2}] \\
&\qquad = |H_{\rm mp}(k)|^{1/2} e^{j\{\arg[H(k)] - \arg[H_{\rm mp}(k)]/2\}}.
\end{aligned}$$

Thus, $\log|H^{(1)}(k)| = (1/2)\log|H(k)|$, and $\arg[H^{(1)}(k)] = \arg[H(k)] - (1/2)\arg[H_{\rm mp}(k)]$. Further analogous filtering of the remaining mixed-phase response would lead to suppression of 3/4 of the initial log-magnitude distortion, with an additional phase shift of $(1/4)\arg[H_{\rm mp}(k)]$ introduced by this operation. At each frequency, after L iterations, the log-magnitude deviation from a constant (zero) level will be reduced by a factor of $(1/2)^L$. The iterative method is guaranteed to converge, as the complex cepstrum $\hat{h}_{\rm mp}^{(l)}(n) = \hat{h}_{\rm mp}(n)/2^{l+1}$ approaches zero with increasing iteration number l, meaning that for large L, $\tilde{h}_{\rm ap}(n) \cong h_{\rm ap}(n)$.

Concerning the length of the sequences, we can deduce the following result. Let $h_{\rm mp}(n)$ be a finite-point sequence of length P, i.e., the values of $h_{\rm mp}(n)$ are known to be zero except for P consecutive points. Because $\hat{h}_{\rm mp}(n) = \hat{h}_{\rm mp}^{(0)}(n) + \hat{h}_{\rm mp}^{(0)}(n)$ (or, $H_{\rm mp}(k) = H_{\rm mp}^{(0)}(k) \cdot H_{\rm mp}^{(0)}(k)$, in the frequency domain), it follows, from the additive properties of the complex cepstrum, that $h_{\rm mp}(n) = h_{\rm mp}^{(0)}(n) \otimes h_{\rm mp}^{(0)}(n)$. Thus, the length of $h_{\rm mp}^{(0)}(n)$ will be approximately half that of $h_{\rm mp}(n)$, that is, $(P+1)/2$. After each iteration, the duration of the remaining minimum-phase part of $h_{\rm mp}(n)$ is halved, so after L iterations the residual minimum-phase component of the mixed-phase response will be approximately of length $P/2^L$. Given that $G_{\rm mp}^{(l)}(k) = 1/H_{\rm mp}^{(l)}(k)$, this will also be valid for the corresponding minimum-phase sequences $g_{\rm mp}^{(0)}(n)$, $g_{\rm mp}^{(1)}(n)$, $\cdots$, $g_{\rm mp}^{(N-1)}(n)$, with respect to the length of $g_{\rm mp}(n)$.

The particular choice of the iterative method of minimum phase inversion over the standard procedure is governed by reduced time-domain aliasing given a fixed number of DFT points. We note that the most critical parameter is the length of the sequence $g_{\rm mp}^{(0)}(n)$, being approximately equal to half the length of the sequence $g_{\rm mp}(n)$. If the duration of $g_{\rm mp}^{(0)}(n)$ is still too long to avoid the effects of circular convolution, the complex cepstrum $\hat{h}_{\rm mp}(n)$ can be decomposed further into pieces of manageable size and processed using the same type of cepstral analysis, with an appropriate convergence criterion.

Thus, partial equalization of the room transfer function, with reduced deviation of the magnitude spectrum, can be achieved according to the operation

$$R(k) = H(k) G_{\rm mp}^{(0;\,L-1)}(k) \tag{15}$$

where $R(k)$ represents the frequency response of a room transmission system in cascade with the minimum-phase equalization filter, given by

$$G_{\rm mp}^{(0;\,L-1)}(k) = G_{\rm mp}^{(0)}(k) G_{\rm mp}^{(1)}(k) \cdots G_{\rm mp}^{(L-1)}(k). \tag{16}$$

The optimal solution for $g_{\mathrm{mp}}^{(0:\,L-1)}(n)$ requires a causal FIR filter with the coefficients defined by convolution of the minimum-phase sequences $g_{\mathrm{mp}}^{(l)}(n)$, for $l = 0, 1, \cdots, L-1$.

Rather than restricting our attention to the effect of the number of DFT points N, we present, in the following subsection, through an example, the main advantages gained with the iteration strategy.

B. Example of Magnitude Equalization Using Novel Approach

For the purpose of our current investigation, an acoustic impulse response was measured from a loudspeaker to a microphone in a general-purpose rectangular listening room. The room volume was 87 m^3 and the length, width, and height were 6.81, 4.70, and 2.72 m, respectively. The room ceiling was mostly reflecting, and the floor mostly carpeted. The background noise level (due to ventilation equipment, etc.) was very low. The loudspeaker was positioned at a distance of 1.2 m to the side wall and 1.5 m to the back wall, measured from the reference position on the front baffle. The microphone was positioned approximately 1.5 m away from the center of the room (toward the opposite corner), at a distance 3.8 m from the loudspeaker. Both the loudspeaker and the microphone were at a height of 1.3 m from the floor.

The reverberation time of the room (for a 60 dB decay) was estimated from a simple model of the exponential reverberation decay in a room [17]. First, we measured the impulse response for several source-microphone arrangements in the room, while keeping the source-to-microphone distance constant, at 3.8 m (direct-to-reverberant energy ratio -9.5 dB). For each of the measured room responses, we determined the frequency ω_c for which the modulation transfer function (MTF) of the transmission system equals 0.707 [17]. The reverberation time of the room was obtained by averaging the different T_{60} calculated for each of the source-microphone arrangements according to the formula: $T_{60} = 13.8/\omega_c$. The estimated reverberation time was 0.28 s.

The room impulse response was acquired at a 48 kHz sampling rate using the Huron DSP system, and bandlimited to 0–4000 Hz using MATLAB's Audio Tool. For purposes of achieving higher accuracy and finer spectral sampling of the Fourier transform, the downsampled (at 8 kHz, by MATLAB's Audio Tool) room response was zero-padded by 13 384 points, resulting in a sequence $N = 16\,384$ samples long, so that a 16 384-point FFT is computed. In the following example, we choose $L = 6$.

The measured impulse response, illustrated in Fig. 1(a), is nonminimum phase, which is evident from Fig. 2(a), showing a small section of the Nyquist plot of the room tranfer function, with the loops encircling the origin in the DFT plane. Fig. 1(b) shows the impulse response of the minimum-phase equalization filter, $g_{\mathrm{mp}}(n)$, which extends over a time interval exceeding more than twice that of the original (measured) impulse response. The minimum-phase inverse $g_{\mathrm{mp}}^{(0)}(n)$ obtained in the first iteration of the alternative approach is shown in Fig. 1(c). The implications of the minimum-phase inverse-filter length have been discussed at the beginning of Section III.

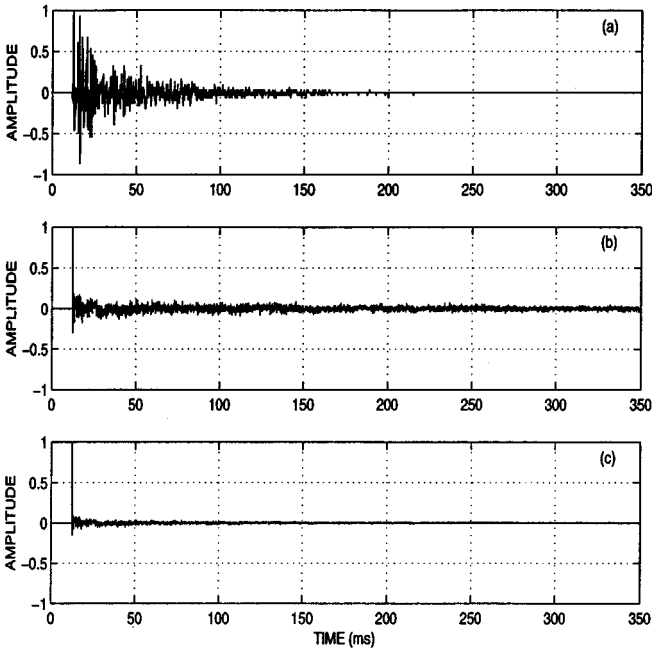

Fig. 1. Upper trace: (a) measured impulse response from the loudspeaker to the microphone position in a room; lower traces: (b) impulse response of the minimum-phase inverse calculated using standard approach to the cepstral decomposition; (c) impulse response of the minimum-phase inverse calculated in first iteration, using alternative approach.

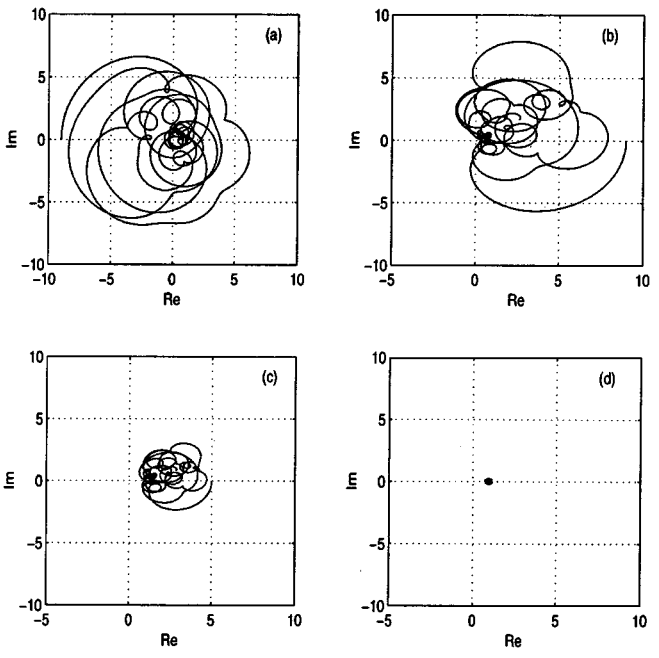

Fig. 2. Nyquist plots of (a) transfer function of a room, (b) minimum-phase part of the transfer function, (c) minimum-phase component calculated in the first iteration, and (d) in the third iteration, all over the frequency range 0–250 Hz.

Fig. 2(b) shows the locus of the transfer function $H_{\mathrm{mp}}(k)$ in the complex plane over the frequency range 0–250 Hz. Fig. 2(c) and (d) reveal some positive attributes of the minimum-phase functions $H_{\mathrm{mp}}^{(l)}(k)$ obtained by the iterative procedure outlined above, as their Nyquist plots: 1) move away from the origin with

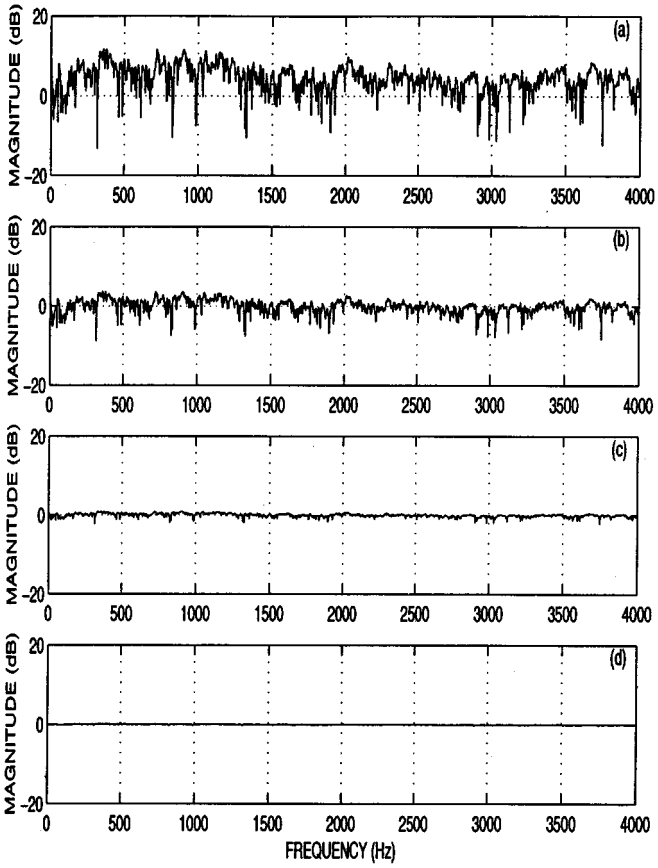

Fig. 3. Flattening of magnitude spectra: (a) measured room response; convolved room/equalizer response as a result of (b) one iteration, (c) three iterations, and (d) six iterations.

increasing l, meaning shorter duration of the corresponding impulse responses in the time domain; 2) reduce in size, meaning less peak-to-valley differences in the corresponding magnitude spectra in the frequency domain.

Fig. 3(b)–(d) demonstrate iterative flattening of the magnitude spectrum of the original room transfer function shown in Fig. 3(a). After only a few iterations, peaks and dips in the room transfer function have been largely removed; after six iterations [Fig. 3(d)], the magnitude spectrum is almost completely flat.

IV. Phase Equalization

The allpass component of the room transfer function has a flat magnitude spectrum but carries a significant portion of the reverberant energy. Since $H_{\rm ap}(k)$ is of nonminimum phase, its exact, zero-delay inverse would consequently have unstable poles, i.e., poles that lie outside of the unit circle. As a consequence, the direct inversion of the allpass component is not possible in practice, since it leads to unstable filter realizations [14].

If the purpose of equalization is to restore the shape of a received pulse to resemble that of a time-delayed transmitted pulse, the problem of phase distortion can be simply solved by convolving the allpass sequence (obtained at the output of the minimum-phase equalization filter) with its time-reversed version, which is equivalent to multiplying sequence $H_{\rm ap}(k)$ with its complex-conjugate, $H^*_{\rm ap}(k)$ in the frequency domain. This process is also known as matched filtering [8].

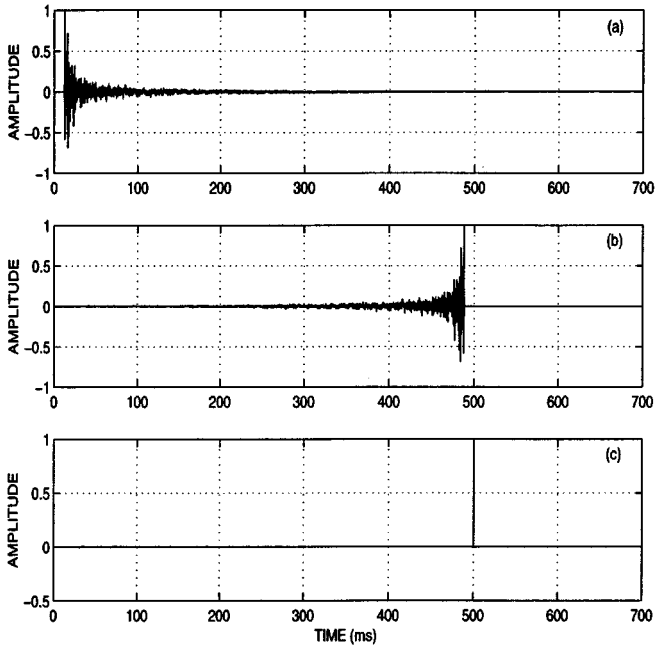

Fig. 4. Post processing of magnitude-equalized impulse response with matched-filter phase equalization. (a) Room impulse response equalized by the minimum-phase inverse filter obtained in six iterations, (b) impulse response of the pulse-shaping allpass matched filter, and (c) equalized room impulse response.

The application of this technique is illustrated in Fig. 4, for the measured room response considered in the previous section. The top trace of this figure shows the time sequence at the output of the minimum-phase equalization filter obtained in the first six iterations, which corresponds, approximately, to the allpass component of the measured room impulse response. This signal is convolved with its time-reversed version, shown in Fig. 4(b). The overall impulse response of the equalized room transmission system is calculated as

$$s(n) = h(n) \otimes p(n) \tag{17}$$

where $p(n)$ is the impulse response of the equalization system, given by

$$\begin{aligned} p(n) &= g^{(0)}_{\rm mp}(n) \otimes g^{(1)}_{\rm mp}(n) \otimes \cdots \otimes g^{(5)}_{\rm mp}(n) \otimes a(n) \\ &= g^{(0:5)}_{\rm mp}(n) \otimes a(n). \end{aligned} \tag{18}$$

In (18), $a(n)$ represents the impulse response of the pulse-shaping (matched) filter, such that $a(n) = h^{(6)}(M-n)$ and $a(n) = 0$ for $n < 0$. The processing delay and the equalization error will depend on the effective matched-filter length, $M_{\rm eff} = M + 1$. In the present example, $M_{\rm eff} = 3800$.

The conceptually simple method outlined above does not require calculation of the phase response; however, any possible ripples in the magnitude spectrum at the output of the minimum-phase magnitude equalizer, will be amplified at the output of the allpass phase equalizer system. The effect of equalizer performance on the subjective perception of speech will be discussed in the following section.

V. Effects of Equalization on Speech Quality: Subjective Tests and Objective Measures

The preceding sections show that nearly perfect dereverberation of speech signals *is possible* if the filters used for deconvolution are of sufficient length. Although an error formula that estimates deviation of the equalized room response from an ideal (or desired) response can be a useful tool in assessing the performance of the equalization system, it bears no simple relation to the perceived quality of the speech after equalization. Sections V-A and V-C consider objective measures, i.e., distortion associated with RTF deconvolution. Sections V-B and V-D are aimed at clarifying some subjective aspects of speech dereverberation in room acoustic systems, in connection with results in Sections V-A and V-C. For consistency with the preceding sections, the reverberant speech signal has been produced by convolving the clean speech with the room impulse response shown in Fig. 1(a).

A. Objective Measure of Equalizer Performance Based on the Error Function

The effectiveness of the equalization procedure described in Sections III-A and IV was evaluated using

1) frequency-domain error function, which estimates the standard deviation of the magnitude response from a constant level S_{dB} (in decibels)

$$\Delta = \left[\frac{1}{N}\sum_{k=0}^{N-1}(10\log_{10}|S(k)| - S_{\mathrm{dB}})^2\right]^{1/2} \tag{19}$$

with S_{dB} given by

$$S_{\mathrm{dB}} = \frac{1}{N}\sum_{k=0}^{N-1} 10\log_{10}|S(k)|. \tag{20}$$

2) A time-domain error function, which estimates the sum of the squares of the error signal $e(n)$ between the desired system impulse response and the (partially) equalized room impulse response

$$\varepsilon = \frac{1}{I}\sum_{n=0}^{I} e^2(n) = \frac{1}{I}\sum_{n=0}^{I}(\delta[n - N_{\mathrm{d}}] - s(n))^2 \tag{21}$$

with I being the length of the output sequence $s(n)$ and N_{d} the modeling delay.

The frequency-domain error function is calculated for the cases of magnitude equalization shown in Fig. 3, using equation (19). The quantitative effects of truncating the matched-filter length (M_{eff}) from the previous section are estimated using the error function given by (21), which we term the error energy. Results are shown in Figs. 5 and 6.

B. Perception of Phase Distortion: Experiment I

The experiment was carried out in an almost echo-free room, with the listener facing toward the loudspeaker set up in the manner shown in Fig. 7. The frequency response of the loudspeaker was within ±3 dB over the working range of 50 Hz to 12.5 kHz. (The loudspeaker was actually a 110-mm cone-type drive unit mounted in a totally enclosed cabinet of 200-mm edge length.)

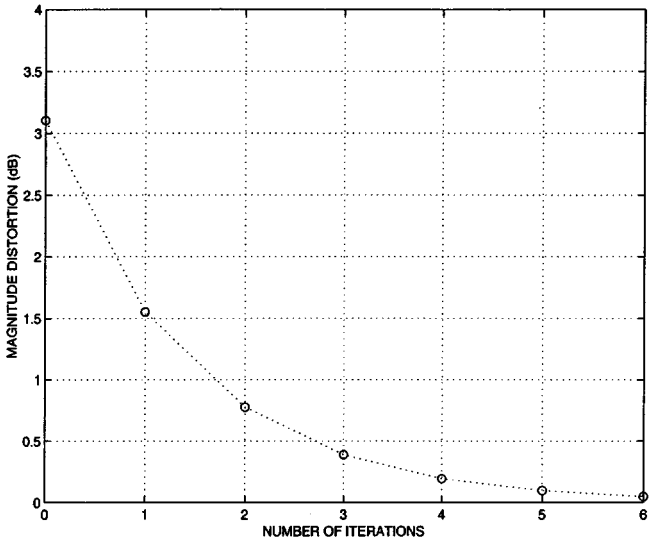

Fig. 5. Residual frequency-domain magnitude distortion as a function of the number of iterations.

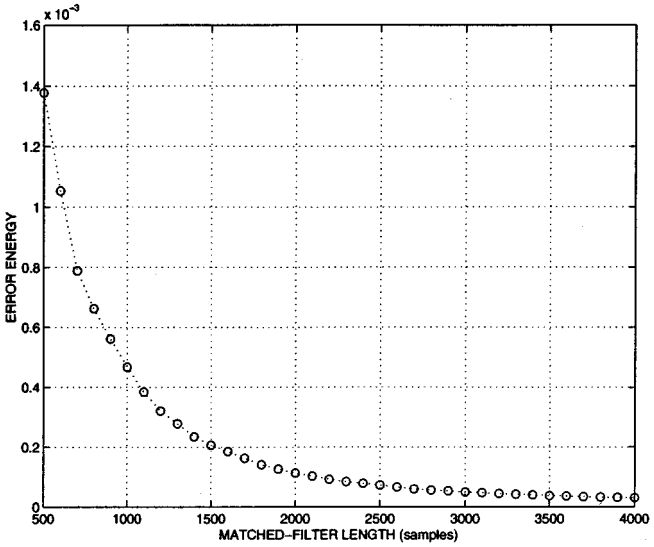

Fig. 6. Time-domain error energy as a function of the matched-filter length (M_{eff}).

A reverberated speech signal 10 s in duration was convolved with the inverse responses calculated in Sections III-A and IV. The test signals were reproduced at a sound level of approximately 64 dB at the position of a listener. A qualitative assessment of the reproduced speech signals is based on subjective judgment of the authors (both with normal hearing), after several listening sessions over a few consecutive days.

Case 1: The test signal was obtained by convolving the reverberant speech with the time-domain response of the minimum-phase magnitude equalization filter obtained in first iteration [see Fig. 3(b)]. On listening alternatively to the reverberant speech and the test signal, a noticeable difference due to reverberation suppression in the second case was perceived, with the acoustic sensation being qualitatively similar to that when switching between monaural and binaural listening in a reverberant room (so called "squelching" effect) [18].

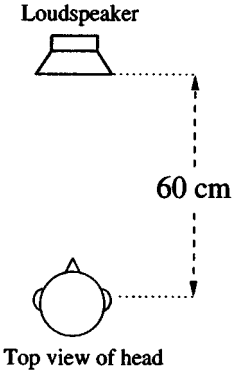

Fig. 7. Experimental setup.

Case 2: The test signal was obtained by convolving the reverberant speech with the minimum-phase magnitude equalization filter calculated in three iterations [see Fig. 3(c)]. The effects of phase distortion first become detectable in this case, with a resounding metalic noise, resembling a bell chime, heard in the background; exactly the same effects have been reported in [14]. No significant echo suppression with respect to *Case 1*, however, has been observed.

Case 3: As in the previous cases, but with six iterations [see Fig. 3(d)]. The metalic noise becomes more prominent, affecting the overall perceived speech quality.

Case 4: The test signal was obtained by convolving the reverberant speech with the impulse response of the magnitude equalization filter calculated in 6 iterations and with the impulse response of the phase-equalization matched filter, producing the overall impulse response shown in Fig. 4(c). The qualitative change is perceived as a sharpening or brightening of the speech signal, with the reverberation completely suppressed. The speech source sounded closer, giving no impression of a closed space, as in *Case 1*. There seemed to be no difference compared to the sensation of the recorded clean speech, except for a low-frequency disturbance heard in the background, which was almost unnoticeable when not consciously listened to.

Case 5: Listening tests analogous to that of *Case 4* were repeated for different lengths of the pulse-shaping filter. With $M_{\text{eff}} \simeq 2000$, the lack of symmetry between the two allpass impulse responses causes the disturbing sounds to become more audible, similar to the metalic noise heard in *Cases 2, 3*. These effects become more pronounced and the speech signal more reverberant with further reduction of the matched-filter length.

It is important to note that the quantitative results shown in Figs. 5 and 6 do not reveal the presence of disturbing sounds, such in *Cases 2, 3, 5*—the only certain information provided by the error functions given by (19) and (21) is the extent to which reverberant sound energy increases with variations from the response of the ideal equalized transmission path. In this context we note that both the quality of the sound *and* the low performance error are essential for successful equalization.

In closing this section, we may conclude

a) if the equalization is intended to suppress reverberation only partially, then shorter FIR filters (such as the one considered in *Case 1*) simplify the problem;

b) if the purpose of equalization is to cancel reverberation completely, the problem can be solved only by using FIR filters of sufficient (in general, extremely long) overall length (such as in *Case 4*).

C. Time-Delay Measure

In *Cases 3–5*, spectral components of the original speech signal arrive at the output of the equalization system with the same attenuation but at different times. A common measure of this "time delay" of frequency components passed through a linear time-invariant system is the group delay function, defined as the negative rate of change of the phase shift with respect to angular frequency ω. The deviation of the group delay from a constant value indicates the presence of phase distortion.

The human ear, as a frequency analyzer, tends to respond to each component of the received signal separately. It has been asserted in [14] that the phase distortions become audible when the deviations from a constant group delay exceed the time interval corresponding to the time constant of the ear. The time constant of the ear has values that range between 30–200 ms [5]. In this time interval, the amplitude of the oscillations of the ear, in response to the sudden stopping of a tone, falls to a proportion of its initial value equal to the ratio $1/e$ or 0.368. The experiments on the fastest variation in the pressure waveform that the ear can follow, carried out with specially constructed digital signals having a nearly flat energy spectrum (Huffman sequences) [19], show this minimum value to be about 2 ms, independently of the frequency region in which the delay in energy occurred.

Fig. 8(a) shows the group delay function of the measured room response and the group delay after equalization with the minimum-phase inverse filter $G_{\text{mp}}^{(0:\,L-1)}(n)$ [see (15), (16)], calculated as

$$\tau_{\text{g}}(k) = -\text{Im}\left[\frac{R'(k)}{R(k)}\right] \tag{22}$$

$$R'(k) = -j\sum_{n=0}^{N-1} nr(n)e^{-j(2\pi/N)kn} \tag{23}$$

where $R(k)$ is as defined earlier, $R'(k) = d[R(\omega)]/d\omega|_{\omega=2\pi k/N}$, and $r(n)$ represents the inverse DFT of $R(k)$. The constant part of the group delay function (the delay of the direct sound) is not included in the plots.

It is evident from Fig. 8 that the spikes in the group delay function are well above the values associated with the time constant of the ear, even in the case of the measured room response—Fig. 8(a)—where the most pronounced group delay peaks are at 47, 2900, 3750 Hz, with corresponding peak values of 1, 0.79, 1.3 s. The introduction of the minimum-phase magnitude equalization filter corresponds to an increase of $\tau_{\text{g}}(k)$ with respect to zero delay. Contrary to the assumption made in [14], it does not seem possible that any useful conclusion can be drawn regarding a relationship between the phase effects described above and the corresponding plots of the time delay distortion.

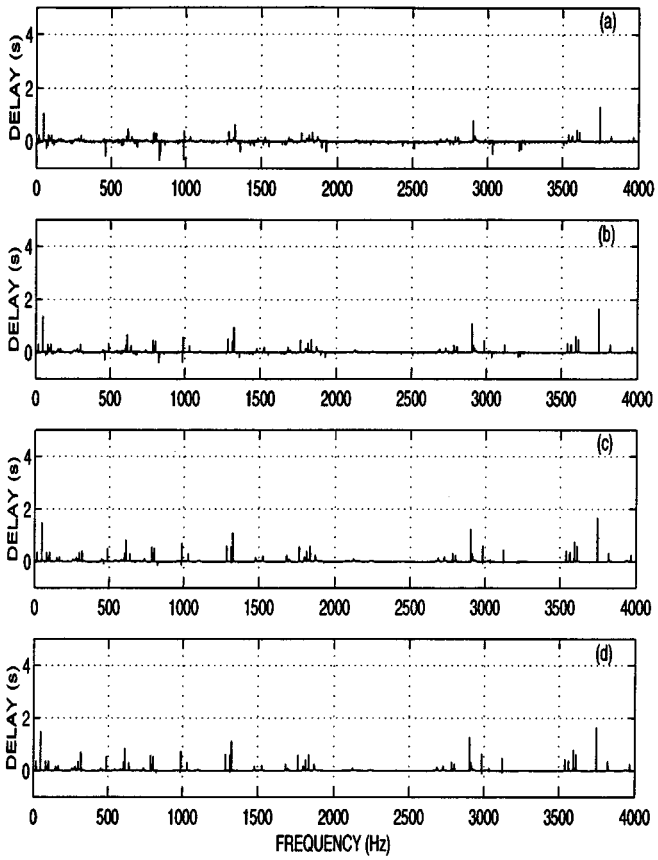

Fig. 8. Group delay of (a) measured room response, (b) after magnitude equalization with the minimum-phase inverse filter obtained in one iteration; (c) after three iterations; and (d) after six iterations.

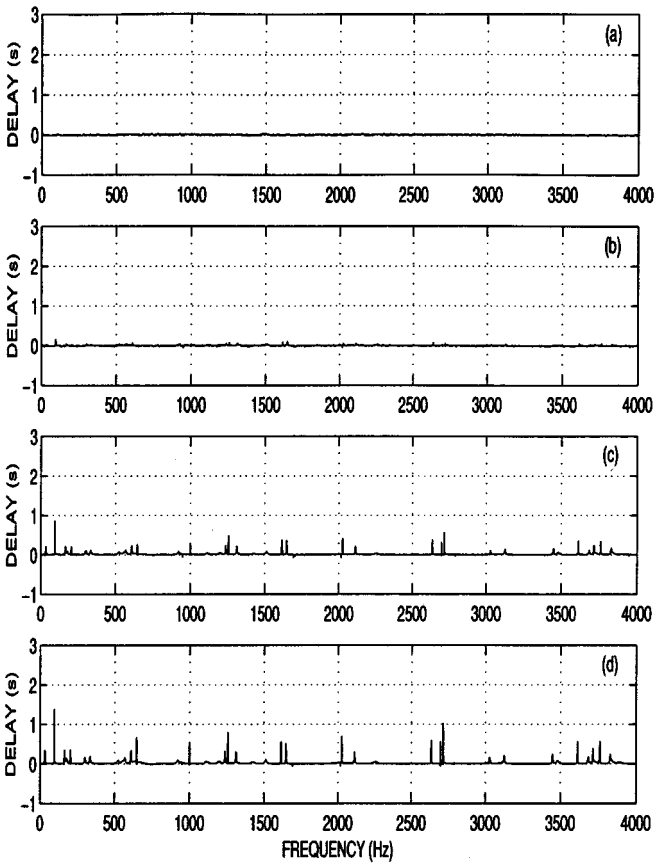

Fig. 9. Modified group delay function of the measured room response (a), and after magnitude equalization with the minimum-phase inverse filter obtained in (b) one iteration, (c) three iterations, and (d) six iterations.

It is plausible, from studies on human hearing [5], to suggest that some *perceptually meaningful* measure of phase distortion ought to be evaluated in terms of the magnitudes of the frequency components presented to the ear. For frequencies not corresponding to the RTF magnitude peaks, the response of the hearing mechanism will be shifted down toward the level of low audibility (frequency-domain masking).

We set up the following criterion, which was established through a series of experiments.

Definition 1: Let $R(\omega)$ be the frequency response of the equalized transmission path between source and receiver in a reverberant room. The *modified group delay function* is defined by multiplying the group delay function $\tau_{\mathrm{g}}(\omega) = -d[\arg(R(\omega))]/d\omega$ by the magnitude spectrum of the partially equalized transmission path, $|R(\omega)|$, scaled by a constant factor $\max(|R(\omega)|)$, i.e.,

$$\tau_{\mathrm{M}}(\omega) = -\frac{d[\arg(R(\omega))]}{d\omega}\frac{|R(\omega)|}{\max(|R(\omega)|)} \tag{24}$$

or equivalently, in the discrete-time frequency domain

$$\tau_{\mathrm{M}}(k) = -\mathrm{Im}\left[\frac{R'(k)}{R(k)}\right]\cdot\frac{|R(k)|}{\max(|R(k)|)}. \tag{25}$$

This sets the values of τ_{M} and τ_{g} equal at frequency corresponding to the most pronounced resonance peak.

Proposition 1: The modified group delay function of the transmission system must not exceed the time constant of the ear for the phase distortion to be inaudible. □

The modified group delay function is plotted in Fig. 9. We observe that while magnitude distortion decreases with the number of iterations simultaneously (Fig. 5), phase distortion becomes more audible as visualised by the modified group delay function. It can be hypothesized that the correlation between the magnitude and phase response of the RTF maintains the factor τ_{M} below the audible levels in a reverberant environment [see Fig. 9(a)].

D. Perception of Phase Distortion: Experiment II

After considerable work on possible techniques that could verify the applicability of the proposed predictor (e.g., [9], [19], [20]), the one described below was found to be the most satisfactory in regard to the consistency of the results and ease of measurement.

The experiment was carried out with 13 listening subjects, of ages 20 to 35 years, each with normal hearing. The subjects were asked to judge on the quality of partially equalized reverberated speech signals (corresponding to the each iteration step of the magnitude equalization procedure described in Section III-A), relative to that of the clean speech signal and the original reverberated speech signal used in Section V-B.

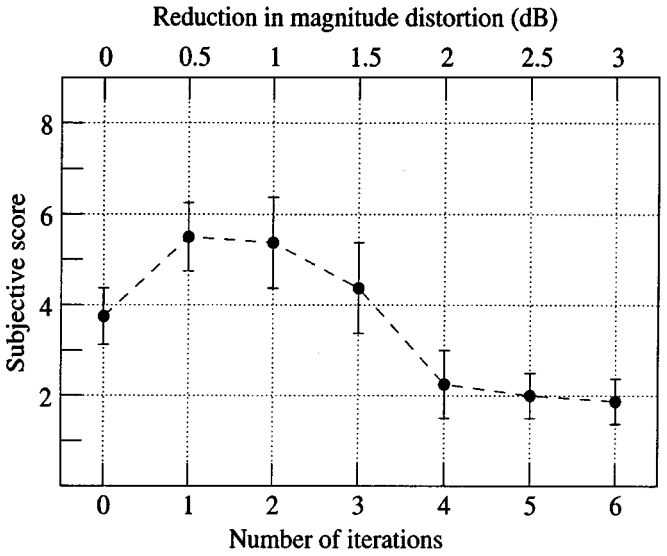

Fig. 10. Subjective scores of the sound quality as a function of the number of iterations. Each data point (filled circles) represents the average of 78 observations: six observations by each of 13 listeners. The vertical bars indicate standard deviation.

The test signals were presented to the ears of a subject through headphones, at a comfortable listening level. The first signal was always chosen to be the clean speech, while the reverberated and partially equalized speech signals were presented in a random order of succession to the listener. The quantification of subjective judgments was performed according to the following scale:

8—Good
7—
6—Fair
5—
4—Poor
3—
2—Bad
1—

where the number 8 denotes a sound quality equal to that of the reverberation-free speech sample.

Each subject completed two one-hour training sessions before data were recorded. The final result was taken as the mean of the individual results for 13 subjects, collected over six trials. The data are plotted in Fig. 10 as a function of number of iterations, ranging from 0 (unequalized signal) to 6 (approximately 3 dB reduction in magnitude distortion).

It can be seen that the judgment curve in Fig. 10 follows the same qualitative trends as the authors' subjective findings reported in Section V-B. This bears out the indication that the proposed predictor is a more appropriate phase-distortion measure than the commonly used measures of phase nonlinearity, usually expressed in terms of the group delay function [8], [9].

VI. Conclusion

In equalization of room acoustics, it is common to use an error function based on a difference between the desired and achieved system responses as an objective measure of performance. A mathematical, rather than perceptually relevant, criterion is for the equalized signal to be as exact a copy as possible of the sound produced at the speaker's position. However, this often results in nonrobust correction of the transmission path distortion and excessive inverse filter lengths.

The present work clarifies some subjective aspects of speech equalization in reverberant rooms. We have presented theoretical and experimental results concerning the audibility of phase distortion in a partially equalized room transmission system. An iterative homomorphic technique was used to track the effects of group delay distortion produced by a minimum-phase inverse filter in equalization of the nonminimum-phase room response. A theory based on the assumption that the phase effects derive from the interrelationship between the magnitude and time delay distortion was developed and shown to be in reasonable agreement with the subjective evaluation of the equalized-speech quality. A criterion has been introduced in this paper that subjectively suppresses reverberation by reducing variations of the RTF magnitude spectrum, while keeping the phase distortion below the threshold of audibility. The results suggest the possibility of achieving partial dereverberation of speech signals by using shorter filter lengths than required by an exact inverse filtering operation.

Finally, we point out that "imperfections" of equalization that we are capable of measuring quantitatively, need not always be considered essential by the listener. Rather, dereverberation is better formulated to take due regard of the underlying physiological mechanism of human perception. In this way, simpler signal processing structures can be employed, such as shorter filter lengths, reduced numerical demands, and more robust processing, less affected by parameter variations.

Acknowledgment

The authors are grateful for the assistance offered by many patient subjects who participated in the experiments. The authors also wish to thank to reviewers for their helpful comments and suggestions.

References

[1] E. Hänsler, "The hands-free telephone problem—An annotated bibliography," *Signal Process.*, vol. 27, pp. 259–271, 1992.
[2] J. N. Mourjopoulos, "Digital equalization of room acoustics," *J. Audio Eng. Soc.*, vol. 42, pp. 884–900, Nov. 1994.
[3] J. Mourjopoulos, P. M. Clarkson, and J. K. Hammond, "A comparative study of least-squares and homomorphic techniques for the inversion of mixed phase signals," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 1982, pp. 1858–1861.
[4] G. S. Ohm, "Noch ein paar worte über die definition des tones," *Ann. Phys. Chem.*, vol. 62, pp. 1–18, 1844.
[5] S. S. Stevens and H. Davis, *Hearing: Its Psychology and Physiology*. New York: Wiley, 1948, p. 287.
[6] M. R. Schroeder, "Models of hearing," *Proc. IEEE.*, vol. 63, no. 9, pp. 1332–1350, Sept. 1975.
[7] ——, "New results concerning monaural phase sensitivity," *J. Acoust. Soc. Amer.*, vol. 31, p. 1579, 1959.
[8] D. Preis, "Phase distortion and phase equalization in audio signal processing—A tutorial review," *J. Audio Eng. Soc.*, vol. 30, no. 11, pp. 774–794, Nov. 1982.
[9] S. P. Lipshitz, M. Pocock, and J. Vanderkooy, "On the audibility of midrange phase distortion in audio systems," *J. Audio Eng. Soc.*, vol. 30, pp. 580–595, Sept. 1982.
[10] J. Blauert and P. Laws, "Group delay distortions in electroacoustical systems," *J. Acoust. Soc. Amer.*, vol. 63, pp. 1478–1483, May 1978.
[11] L. R. Fincham, "The subjective importance of uniform group delay at low frequencies," *J. Audio Eng. Soc.*, vol. 33, no. 6, pp. 436–439, June 1985.

[12] J. A. Deer and P. J. Bloom, "Perception of phase distortion in all-pass filters," *J. Audio Eng. Soc.*, vol. 33, pp. 782–785, Oct. 1985.
[13] H. Kuttruff, "On the audibility of phase distortions in rooms and its significance for sound reproduction and digital simulation in room acoustics," *Acustica*, vol. 74, pp. 3–7, 1991.
[14] S. T. Neely and J. B. Allen, "Invertibility of a room impulse response," *J. Acoust. Soc. Amer.*, vol. 66, no. 1, pp. 165–169, July 1979.
[15] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
[16] H. Tokuno, O. Kirkeby, P. A. Nelson, and H. Hamada, "Inverse filter of sound reproduction systems using regularization," *IEICE Trans. Fund.*, vol. E80-A, pp. 809–820, May 1997.
[17] M. R. Schroeder, "Modulation transfer functions: Definition and measurement," *Acustica*, vol. 49, pp. 179–182, 1981.
[18] W. Koeing, "Subjective effects in binaural hearing," *J. Acoust. Soc. Amer.*, vol. 22, pp. 61–62, Jan. 1950.
[19] D. M. Green, "Temporal acuity as a function of frequency," *J. Acoust. Soc. Amer.*, vol. 54, no. 2, pp. 373–379, 1973.
[20] H. Haas, "Über den einfluss eines einfachechos auf die hörsamkeit von sprache," *Acustica*, vol. 1, pp. 49–58, 1951.

Biljana D. Radlović (S'99) received the B.S. degree in electrical engineering from the University of Niš, Yugoslavia, in 1992. She is currently pursuing the Ph.D. degree at Telecommunications Engineering Group, Australian National University, Canberra.

From 1993 to 1996, she was a Research Assistant with the Department of Telecommunications, University of Niš. Her research interests are in the fields of acoustics and signal processing.

Rodney A. Kennedy (S'86–M'88) was born in Sydney, Australia, in 1960. He received the B.E. (Hons.) degree in electrical engineering from University of New South Wales, Australia, in 1982, the M.E. degree in digital control theory from the University of Newcastle, Australia, in 1986, and the Ph.D. degree in 1988 from the Department of Systems Engineering, Australian National University (ANU), Canberra.

He is Professor and Head of the Telecommunication Engineering Group, Research School of Information Sciences and Engineering, ANU. His research interests are in the fields of digital communications, digital signal processing, and acoustical signal processing.

Dr. Kennedy is currently an Editor for Data Communications, IEEE TRANSACTIONS ON COMMUNICATIONS.