Invertibility of a room impulse response

Stephen T. Neely[a] and Jont B. Allen

Acoustics Research Department, Bell Laboratories, Murray Hill, New Jersey 07974
(Received 6 July 1979)

When a conversation takes place inside a room, the acoustic speech signal is linearly distorted by wall reflections. The room's effect on this signal can be characterized by a room impulse response. If the impulse response happens to be minimum phase, it can easily be inverted. Synthetic room impulse responses were generated using a point image method to solve for wall reflections. A Nyquist plot was used to determine whether a given impulse response was minimum phase. Certain synthetic room impulse responses were found to be minimum phase when the initial delay was removed. For these cases a mimimum phase inverse filter was sucessfully used to remove the effect of a room impulse response on a speech signal.

PACS numbers: 43.55.Br, 43.45.Bk

INTRODUCTION

When a conversation takes place inside a room, the speech signals are distorted by the presence of nearby reflecting walls. Sounds travel not only the direct path from source to receiver, but also reach the receiver after bouncing off one or more walls. The total "room effect" can be viewed as a convolution in the time domain of the speech signal with a room impulse response. This room effect is perceived as echo and reverberation and is often undesirable. One scheme for removing the room effect is to pass the distorted speech through a second filter which exactly inverts the effect of the room. The purpose of this paper is to characterize a room impulse response in terms of parameters relevant to its invertibility.

An impulse response, like any other finite energy time function, can be characterized by the magnitude and phase of its Fourier transform. If $H(\omega)$ is the Fourier transform of $h(t)$ and $\phi(\omega)$ is the phase of $H(\omega)$, then

$$H(\omega) = |H(\omega)| \exp[i\phi(\omega)]. \tag{1}$$

For a certain class of functions known as minimum phase functions, $\phi(\omega)$ can be uniquely determined from $|H(\omega)|$. A function is said to be minimum phase if its Laplace transform contains no poles or zeroes in the right half-plane. When a function is minimum phase, the log magnitude and phase of the Fourier transform are related through the Hilbert transform.[1] This relationship depends upon the fact that the log of the Laplace transform is analytic in the right half-plane for minimum phase functions.

Let $h(t)$ be an arbitrary room impulse. We expect $h(t)$ to be a stable, casual, but, in general, nonminimum phase impulse response. The phase of the Fourier transform $\phi(\omega)$ can be expressed as the sum of a minimum phase component $\phi_m(\omega)$ (as determined from $|H(\omega)|$) and a component which represent the deviation from minimum phase $\phi_a(\omega)$:

$$\phi(\omega) = \phi_m(\omega) + \phi_a(\omega). \tag{2}$$

With this notation $H(\omega)$ can be factored into a minimum phase and an allpass part.

$$H(\omega) = M(\omega)A(\omega), \tag{3}$$

where

$$M(\omega) = |H(\omega)| \exp[i\phi_m(\omega)], \tag{4}$$

and

$$A(\omega) = \exp[i\phi_a(\omega)]. \tag{5}$$

Note that $|A(\omega)| = 1$ for any $\phi_a(\omega)$. When $H(\omega)$ is minimum phase $\phi_a(\omega) = 0$, which implies that $A(\omega) = 1$.

Minimum phase impulse responses are of particular interest because their inverses are guaranteed to be minimum phase and casual. (The truth of this statement can be seen by considering the s plane of the Laplace transform where the inverse filter replaces poles with zeros and vice versa. If the original impulse has poles and zeros only in the left half-plane, then its inverse must have poles and zeros only the left half-plane.) If a room impulse response is minimum phase, then an inverse filter will exist capable of completely removing the room's effect from a speech signal. Furthermore, this inverse filter can be determined from knowledge of only the magnitude of the room's frequency response (i.e., unknown phase), which can be estimated from the signal power spectra.

If $A(\omega)$ is not identically equal to one, then it will be (as a consequence of its definition) nonminimum phase. Thus, $H(\omega)$ will not be minimum phase and its inverse may be either unstable or acasual. If, however, $A(\omega)$ represents a "pure delay" it will introduce no perceptual distortion in the speech signal. In terms of an impulse response, "pure delay" is defined as an allpass function with a group delay $\tau_{ga}(\omega)$ which is constant for all frequencies, i.e.,

$$\tau_{ga}(\omega) = -[d\phi_a(\omega)/d\omega] = \text{constant}. \tag{6}$$

If a room impulse response were minimum phase with pure delay, then an inverse filter would only need to remove the minimum phase component.

The worst case, for inverse filters, is a room impulse for which $A(\omega)$ has a group delay which is not independent of frequency. Perceptually effective inverse filters may exist for this case, but will not be

[a] Present address: Computer Systems Laboratory, 724 South Euclid Avenue, St. Louis, MO 63110.

 0001-4966/79/070165-05$00.80

considered in this paper.

To investigate the usefulness of the minimum phase inverse filter, synthetic room impulse responses were generated by computer and separated into their minimum phase and allpass components. It was found that certain room impulses are truly minimum phase with pure delay. For these cases a minimum phase inverse filter was effective in removing the "room effect" from a distorted speech signal.

I. METHODS

Synthetic impulse responses were generated by an existing computer program[2] on a Data General S200 Eclipse computer. This synthesis program accepts values for room size (length, width, and height) source location, and reflectivity for each of the six walls. The synthetic impulse responses are found by using a point image method to solve for wall reflections. Duration of the computed responses was 2048 samples (or 204.8 ms assuming a 10 kHz sampling rate).

The minimum phase component of the impulse response was determined by zeroing the cepstrum for negative quefrencies.[3] For a finite sequence $h(n)$ (such as the truncated room impulse responses) a real, periodic approximation to the cepstrum $c_p(n)$, is defined by the following equations:

$$H(k) = \mathrm{DFT}[h(n)] = \sum_{n=0}^{N-1} h(n) \exp[-i(2\pi/N)kn], \quad (7)$$

$$C(k) = \log|H(k)|, \quad (8)$$

$$c_p(n) = \mathrm{DFT}^{-1}[C(k)] = \frac{1}{N}\sum_{k=0}^{N-1} C(k) \exp[i(2\pi/N)kn], \quad (9)$$

where DFT indicates a discrete Fourier transform. The minimum phase component $M(k)$ of the frequency response $H(k)$ is computed from $c_p(n)$ in the following manner. Since $c_p(n)$ is a periodic function of n, the cepstrum may be effectively zeroed for negative quefrencies by setting the second half of $c_p(n)$ to zero. The modified cepstrum is transformed back to the frequency domain to obtain $M(k)$. This procedure is described by the following equations. If

$$\hat{m}(n) = \begin{cases} c_p(n), & n = 0, N/2, \\ 2c_p(n), & 1 \le n < N/2, \\ 0, & N/2 < n \le N-1, \end{cases} \quad (10)$$

and

$$\hat{M}(k) = \mathrm{DFT}[\hat{m}(n)] = \sum_{n=0}^{N-1} \hat{m}(n) \exp[-i(2\pi/N)kn], \quad (11)$$

then

$$M(k) = \exp[\hat{M}(k)]. \quad (12)$$

In the limit as N becomes very large, this procedure for determining $M(k)$ is equivalent to using the Hilbert transform relation (mentioned above) to compute the minimum phase $\phi_m(k)$ from $\log|H(k)|$ and letting $M(k) = |H(k)|\exp[i\phi_m(k)]$.

The allpass component $A(k)$ of the frequency response is then computed by dividing out the minimum phase component.

$$A(k) = H(k)/M(k). \quad (13)$$

Because of the special significance of a pure delay, the group delay of the allpass component was of particular interest. Group delay was computed by the formula

$$\tau_{ga}(k) = -\mathrm{Im}\left[\frac{A'(k)}{A(k)}\right], \quad (14)$$

where $A'(k)$ is the frequency derivative of $A(k)$, $\mathrm{Im}(\cdot)$ denotes the imaginary part,

$$A'(k) = -i\,\mathrm{DFT}[na(n)] = -i\sum_{n=0}^{N-1} na(n) \exp[-i(2\pi/N)kn], \quad (15)$$

and

$$a(n) = \mathrm{DFT}^{-1}[A(k)] = \frac{1}{N}\sum_{k=0}^{N-1} A(k) \exp[i(2\pi/N)kn]. \quad (16)$$

The motivation for determining the minimum phase component of the impulse response was the relative ease of computing a minimum phase inverse filter. The impulse response of the minimum phase inverse filter $g(n)$ was computed by taking the inverse DFT of the reciprocal of the minimum phase spectrum $M(k)$:

$$G(k) = 1/M(k), \quad (17)$$

$$g(n) = \mathrm{DFT}^{-1}[G(k)] = \sum_{k=0}^{N-1} G(k) \exp[i(2\pi/N)kn]. \quad (18)$$

To test the effectiveness of the inverse filter perceptually, a selected speech sample was filtered with the room impulse response and inverse filtered with the minimum phase inverse. The resultant speech was compared, in informal listening test, with the original speech. These comparisons will be discussed in Sec. III.

Finally, a necessary and sufficient condition was desired for determining whether or not a given impulse response was, indeed, minimum phase. For this purpose the Nyquist criterion was used to detect the presence of nonminimum phase zeros. The Nyquist criterion is based on a mapping theorem of Cauchy. If a complex variable z in the z plane describes a contour C_1 in a positive sense, then $F(z)$, a function of the complex variable z, will describe a contour C_2 in the $F(z)$ plane, which will encircle the origin N times in a positive direction, where N is the difference between the number of zeros and poles of $F(z)$ enclosed C_1 (Ref. 4).

The use of the Nyquist criterion for discrete time systems is essentially the same as for continuous systems except that instead of detecting zeros in the right-half complex z plane (of the Laplace transform), we will detect zeros outside the unit circle in the z plane (of the z transform). If $F(z)$ is the z transform of a stable, casual time sequence $f(n)$, then $F(z)$ has no poles outside the unit circle. Consequently, the number of encirclements by $F(z)$ of the origin in the $F(z)$ plane (as z describes the unit circle in the z plane) is precisely the number of zeros of $F(z)$ exterior to the unit circle in the z plane (i.e., the number of nonminimum phase zeros).

The $\mathrm{DFT}[f(n)] = F(k)$ is equal to $F(z)$ sampled at a finite number of points on the unit circle in the z plane. So, as N becomes large, the number of nonminimum phase zeros will be equal to the number of times $F(k)$ encircles the origin in the DFT plane, provided that the phase of $F(k)$ does not change by more than π between frequency points. To increase the density of (and consequently reduce the amount of phase change between) frequency points, the synthetic room impulse responses were extended in time by appending an interval of zero response. This has the effect of increasing N both in the time domain and in the frequency domain when the DFT is taken.

The test for the minimum phase property, therefore, reduces to the following procedure. A polar plot (Nyquist plot) is made of the DFT of the impulse response which has been shifted in the time domain to remove any pure delay: If the plot does not encircle the origin, then the shifted impulse response must be minimum phase.

II. RESULTS

All synthetic impulse responses for this paper were generated for a room 130 samples long, 110 samples wide, and 70 samples high which we denote by the vector (70, 110, 130). Assuming a sound velocity of 1000 ft/s and a sampling rate of 10 kHz, these dimensions correspond to a room approximately 13 ft × 11 ft × 7 ft, which is about the size of a small office.

Within this room the source was first placed at coordinates (10, 20, 30) with the receiver at (40, 50, 60) (Fig. 1). A distance of about 52 samples (or about 5 ft) separated the source and receiver. With 10% reflectivity on all six walls, the room impulse response was found to be strictly minimum phase within a pure delay.

The reflectivity of the walls in this room was then raised until the room became nonminimum phase. A reflectivity of 36% still produced a delayed minimum phase impulse response, according to the Nyquist test, while a reflectivity of 37% produced a delayed nonminimum phase impulse response. The threshold for nonminimum phase behavior was thus determined to be about 37% (note that a typical wall reflection coefficient is 70%–90%).

The source and receiver were moved toward the center of the room while maintaining the same separation. The source was placed at coordinates (20, 20, 20) and the receiver at (50, 50, 50). As a result, the threshold was lowered to about 33%.

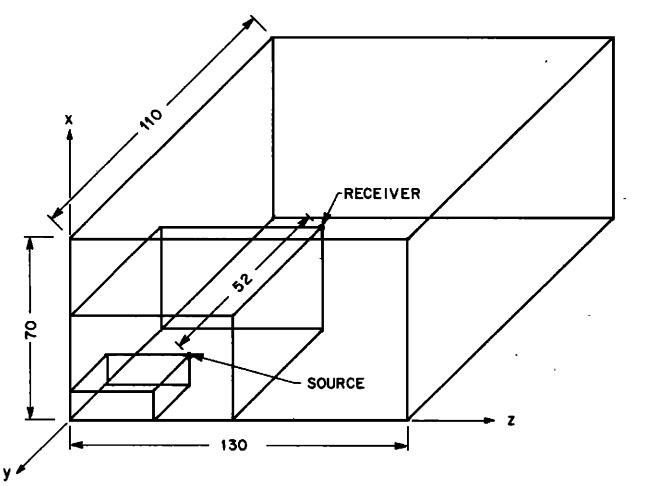

FIG. 1. Primary location for source and receiver: room size is (70,110,130), source location is (10,20,30), receiver location is (40,50,60). Source–receiver distance is therefore 52 samples.

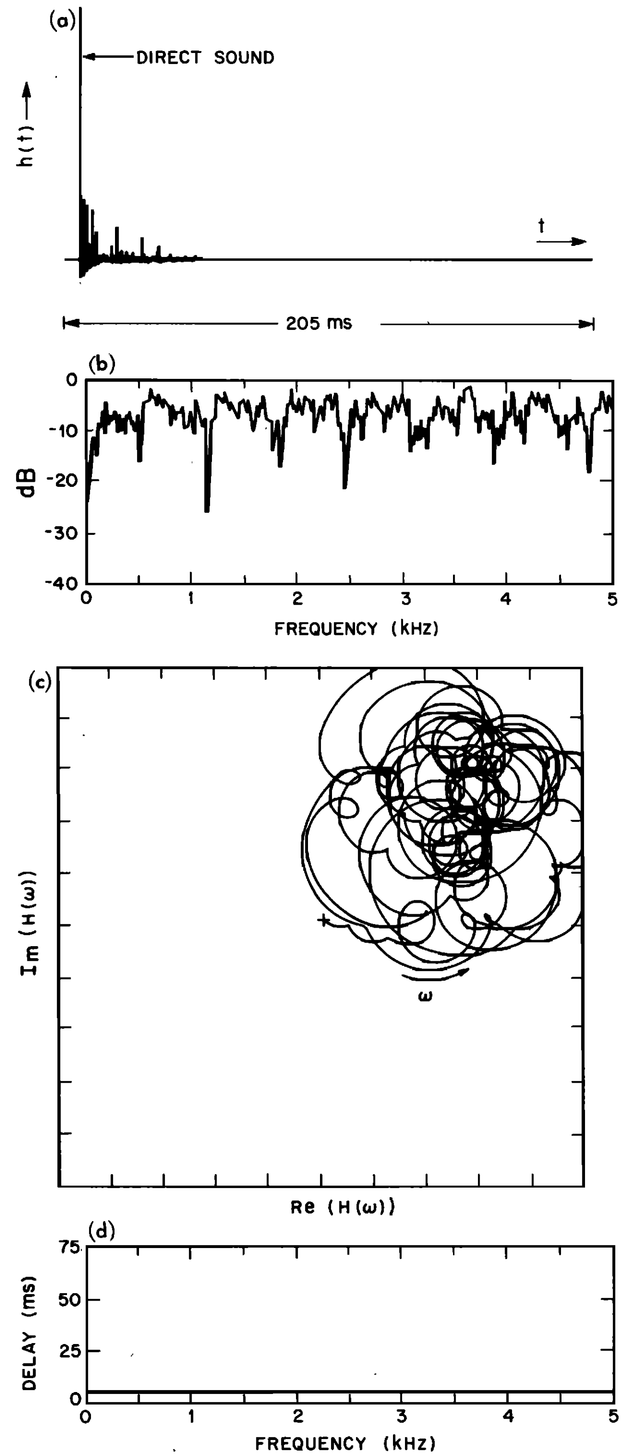

FIG. 2. Room impulse response for primary location with 35% reflectivity (minimum phase). (a) impulse response, (b) frequency response, (c) nyquist plot (of frequency response, and (d) group delay.

The source was then returned to its original location (10, 20, 30) and the receiver was placed at (50, 60, 70), a separation of about 69 samples. For this case, the threshold for nonminimum phase behavior was lowered to about 34%.

Finally, with the source at (10, 20, 30), the receiver was placed at (30, 40, 50), a separation of about 35 samples. The threshold for nonminimum phase behavior was increased in this case to nearly 40%.

Typical examples were chosen to test the effectiveness of a minimum phase inverse filter. A minimum phase room with 35% reflectivity is shown in Fig. 2(a) and a nonminimum phase response for a room with 40% reflectivity is shown in Fig. 3(a). These two responses were used to filter a selected speech sample. The filtered speech was then inverse filtered with the minimum phase inverse, which was computed from the log-magnitude frequency response. The effect of the inverse filter for the nonminimum phase case is shown in Fig. 4.

III. DISCUSSION OF RESULTS

The results of the previous section show that certain impulse responses for our synthetic rooms are minimum phase within a pure delay. For a given room size and fixed source and receiver locations, there was a threshold for the reflectivity. When the reflectivity (for all six walls) was kept below this threshold, the room impulse response had a minimum phase behavior. But, if the reflectivity was increased beyond this threshold value, the room impulse had a nonminimum phase behavior. This threshold was not constant for a given room size. The reflectivity threshold was *lowered* when: (1) The "near" walls were equidistant from the source and equidistant from the receiver; and (2) The separation between source and receiver was increased. Perceptually, the increase in echo and reverberation with larger reflectivity is evident, but there was no apparent qualitative difference when the reflectivity threshold was crossed.

When the speech sample was filtered by the minimum phase impulse response and inverse filtered, the resultant speech sounded identical to the original. However, when the speech sample was filtered by the nonminimum phase impulse response and inverse filtered with the minimum phase inverse, there was a distinctive difference. The room effect had been removed, but a tone, much like a bell chime, sounded in the background. This impulse response had two, narrow bandwidth, nonminimum phase zeros (at about 1200 and 2500 Hz) which the inverse filter could not remove [see Figs. 4 and 4 (c)].

The allpass component of the impulse response is of interest because it is the component of the room effect which cannot be removed by the minimum phase inverse filter. By definition, the allpass component cannot be deduced if only the magnitude of the frequency response is known. The allpass group delay was the most useful representation of the allpass component [see Figs. 2(d) and 3(d)]. In the minimum phase case [Fig. 2(d)] the

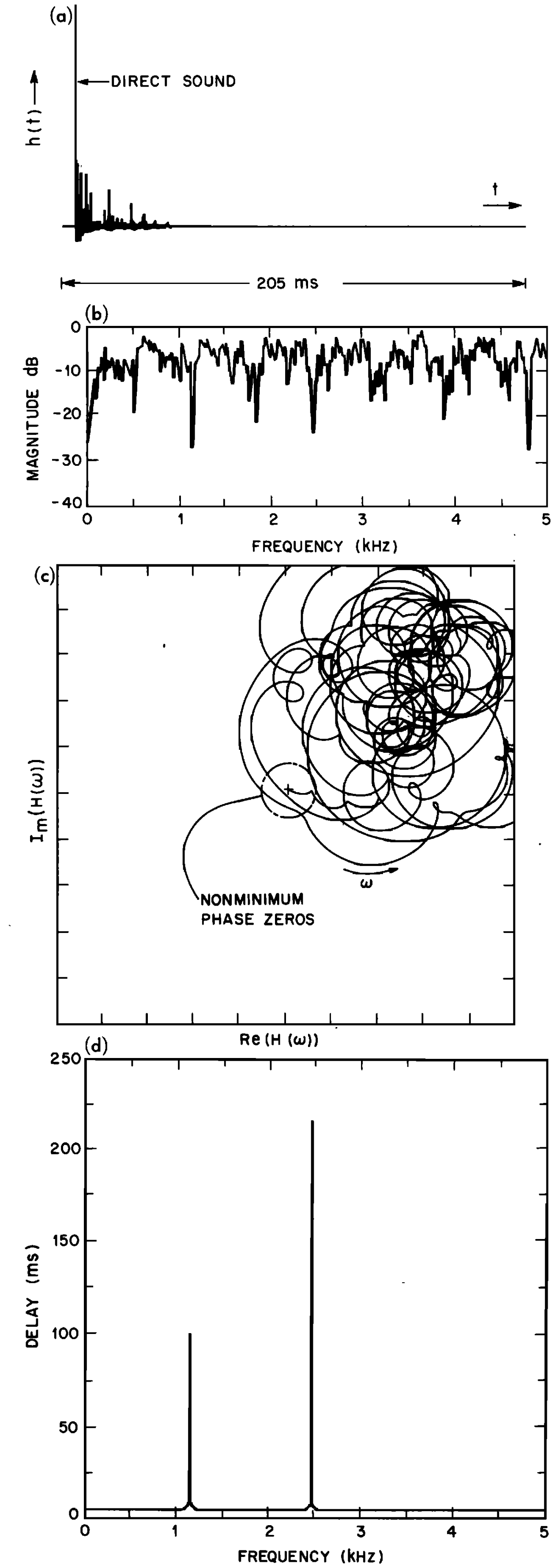

FIG. 3. Room impulse response for primary location with 40% reflectivity (nonminimum phase). (a) impulse response, (b) frequency response, (c) nyquist plot (of frequency response), and (d) group delay.

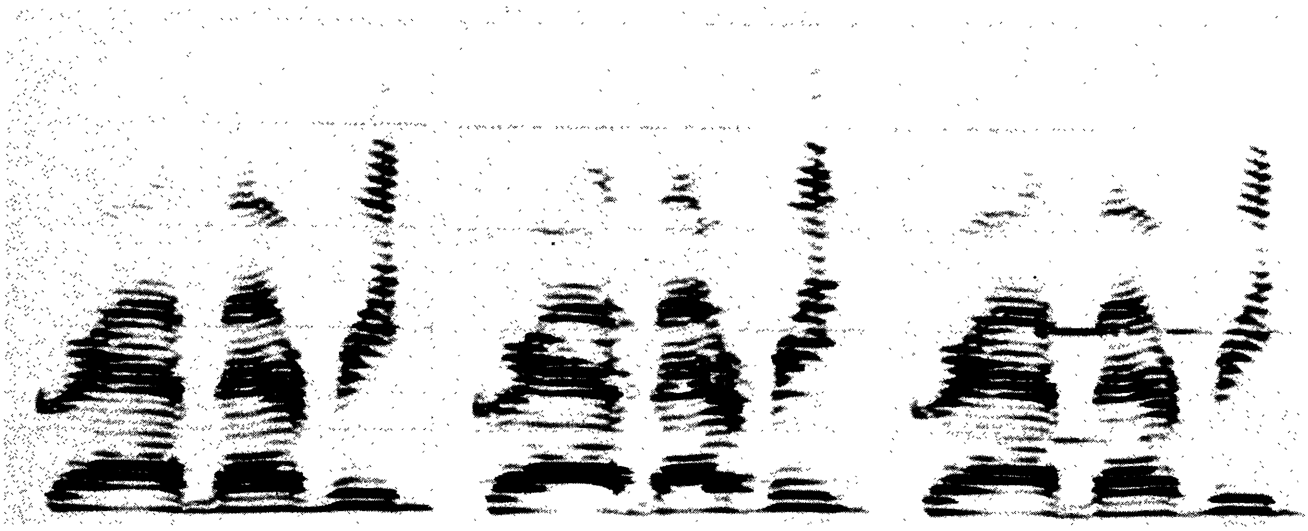

FIG. 4. Voice prints of speech at various points in the process. (a) original speech, (b) speech which has been reverberated by nonminimum phase room, and (c) inverse filtered speech. Note reverberant tails in 4(b) and the two tones in 4(c) at 1.2 and 2.5 kHz. These tones are due to the allpass component of the reverberation.

group delay was 5.2 ms for all frequencies. (This was the delay of the direct sound, i.e., 52 samples.) The nonminimum phase case [Fig. 3(d)] was dramatically different at the frequencies of the nonminimum phase zeros. This deviation from constant group delay caused the perceived tone described above. The Fourier transform spectrum of an allpass filter is always flat, but the ear acts as a filter bank and thus forms a short-term spectrum. An allpass filter does not have a flat short-term spectrum when analyzed by a filter bank if the deviation from constant group delay becomes greater than the filter decay time.

Inspection of the impulse response in the time domain [see Fig. 2(a)] does give some clue as to whether it is minimum phase. The first sample (after the initial delay) must be larger than all other samples and the response should decay rapidly. Of course, the log-magnitude frequency response [see Figs. 2(b) and 3(b)] can give no indication as to whether the impulse response is minimum phase, since *every* log-magnitude frequency plot has a corresponding minimum phase response.

Inspection of the Nyquist plot [see Figs. 2(c) and 3(c)] makes determination of minimum phase a trivial matter. It must be kept in mind, however, that the phase can not change rapidly between frequency points if the interpretation is to be valid. This condition was adequately satisfied whenever the response was minimum phase. In any case, the Nyquist plot is a much better indicator of minimum phase than the allpass group delay which is somewhat prone to aliasing and numerical errors.

By comparison of synthetic room impulses with physically measured room impulses, it was estimated that real, typical offices have walls with reflectivity on the order of 90%. The synthetic room presented in this paper must have source and receiver within seven samples of each other (about 8 in) in order to raise the reflectivity threshold as high as 90%. This implies that a typical room would have a nonminimum phase effect on a speech signal when the receiver is more than 8 in. from the source. For this reason other methods of dereverberation are necessary under typical room conditions.[5]

---

[1]B. Gold and C. M. Rader, *Digital Processing of Speech Signals* (McGraw–Hill, New York, 1969).

[2]J. B. Allen and D. A. Berkley, "Image Method for Efficiently Simulating Small-Room Acoustics," J. Acoust. Soc. Am. **65**, 943–950 (1979).

[3]A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing* (Prentice–Hall, Englewood Cliffs, NJ, 1975).

[4]R. Saucedo and E. E. Schiring, *Introduction to Continuous and Discrete Control Systems* (Macmillan, New York, 1968).

[5]J. B. Allen, D. A. Berkley, and J. Blauert, "Multimicrophone signal processing technique to remove room reverberation from speech signals," J. Acoust. Soc. Am. **62**, 912–915 (1977).