Least-Squares Inverse Filtering of Speech

Bradford Gover†‡and James G. Ryan‡

†Department of Physics, University of Waterloo, Waterloo, ON N2L 3G1, Canada

‡Institute for Microstructural Sciences, National Research Council, Ottawa, ON K1A 0R6, Canada.

1 Introduction

When speech propagates through a room and is received by a microphone, the microphone output signal sounds unnatural and suffers reduced intelligibility (the so-called "barrel effect"). It is necessary to employ some sort of enhancement technique to process the microphone output so that the subjective impression of the subsequently reproduced speech is improved.

If the speech originates from a loudspeaker, then there are means to approximately invert the room's effect (see, for example, [1] and references therein). Most of these depend on the knowledge of the room impulse response as measured from the loudspeaker to the microphone, and the subsequent design of an appropriate inverse filter.

If, however, the source of the speech is a human talker, then it is not possible to determine this impulse response exactly. One way to attempt to enhance the live speech is to replace the talker with a loudspeaker, measure the impulse response, design an inverse filter, and then use it for the live speech. This introduces the question of the suitability of inverse filters for sources other than those from which they are designed.

This paper will illustrate some of the problems with such an approach. First, the technique of least-squares inversion will be summarized. Next, some actual room measurements using a loudspeaker and a mannequin will be presented. The mannequin is used to represent a real talker, and the results of its attempted dereverberation with the filter designed from the loudspeaker measurement will be presented and discussed.

2 Least-Squares Inverse Filtering

A least-squares inverse filter, when applied to a system impulse response, yields a result which is the best least-squares approximation to an ideal impulse. To invert a known impulse response $g(n)$, a filter $h(n)$ is sought such that $g(n) * h(n) = \delta(n - n_0)$, where n_0 is some delay. This can be written

$$\mathbf{Gh} = \mathbf{d}, \tag{1}$$

where $\mathbf{G}$ is the convolution matrix of $g(n)$, and the elements of the vectors $\mathbf{h}$ and $\mathbf{d}$ are the values of the time signals $h(n)$ and $\delta(n - n_0)$. The least-squares solution is obtained by minimizing the L_2-norm of the error vector $\mathbf{e} = \mathbf{Gh} - \mathbf{d}$. This occurs when it is orthogonal to the subspace spanned by the columns of $\mathbf{G}$ [2], so that

$$\mathbf{G}^T\mathbf{e} = \mathbf{G}^T\mathbf{Gh} - \mathbf{G}^T\mathbf{d} = 0, \tag{2}$$

or

$$\mathbf{Rh} = \mathbf{z}, \tag{3}$$

where $\mathbf{R} = \mathbf{G}^T\mathbf{G}$ is the correlation matrix of $g(n)$ and $\mathbf{z} = \mathbf{G}^T\mathbf{d}$ is the cross-correlation vector of $g(n)$ and the desired signal $\delta(n - n_0)$. The solution of Eq. (3) yields the desired filter.

3 Room Impulse Response Inversion

A room impulse response measuring system usually involves a signal chain comprising: pre-amplifiers, a loudspeaker, the room, a microphone, and post-amplifiers. In this case, the measured response, $g_{meas}(n)$, contains the responses of the amplifiers and transducers,

$$g_{meas} = g_{pre} * g_{spkr} * g_{room} * g_{mic} * g_{post}, \tag{4}$$

with the obvious definitions of the impulse responses on the right-hand side. Clearly for an inverse to have any chance of being independent of the system used to measure it, it must be an inverse of the room response, $g_{room}(n)$ only. If an inverse of the room response, say $h_{room}(n)$, is found, then the result of filtering the measured response is

$$g_{meas} * h_{room} = g_{pre} * g_{spkr} * g_{mic} * g_{post}, \tag{5}$$

which is exactly the anechoic response of the measuring system. Therefore, to design an inverse of the room response only, the least-squares criterion Eq. (1) should be changed to

$$\mathbf{Gh} = \mathbf{s}, \tag{6}$$

where $\mathbf{s}$ is the vector corresponding to $s(n - n_0)$, the delayed anechoic response of the measuring system. The solution for such a filter is found by solving Eq. (3), with $\mathbf{z} = \mathbf{G}^T\mathbf{s}$.

It is known that the room impulse response, $g_{room}(n)$, depends not only on the positions of the source and receiver, but also on the source directivity [3]. It is further known that due to the nonminimum-phase nature of the room impulse response, an exact inverse does not exist [4]. It is explored herein how detrimental the consequences of these facts are on dereverberation by inverse filtering.

4 Measurements

Room impulse response measurements were made with MLSSA, a maximum-length sequence analyzer. The sampling rate was 22 050 Hz and the MLSSA anti-alias filters (8th-order Chebyshev) had a bandwidth of 7.35 kHz. The MLS was 65 535 points long (16th order) and the measured impulse responses were 32 768 points long.

The room used for measurements was rectangular in shape with dimensions 4.65 × 6.70 × 2.44 m, smooth walls and ceiling, and carpeted floor. The room contained a typical amount of furniture, and had a reverberation time of approximately 350 ms. A Panasonic 6 mm electret microphone was located at 2.71 × 4.33 × 1.14 m. Measurements were taken with two sources: a PSB Alpha SE loudspeaker and a B&K HATS mannequin, which is intended to reflect the directional characteristics of a human talker. The sources were positioned at 2.07 × 2.31 × 0.82 m.

Additionally, the same equipment was set up in the same arrangement in an anechoic chamber and the measurements repeated.

The measured room and anechoic impulse responses and corresponding magnitude spectra are shown in the top two curves of Figures 1 (loudspeaker) and 2 (mannequin).

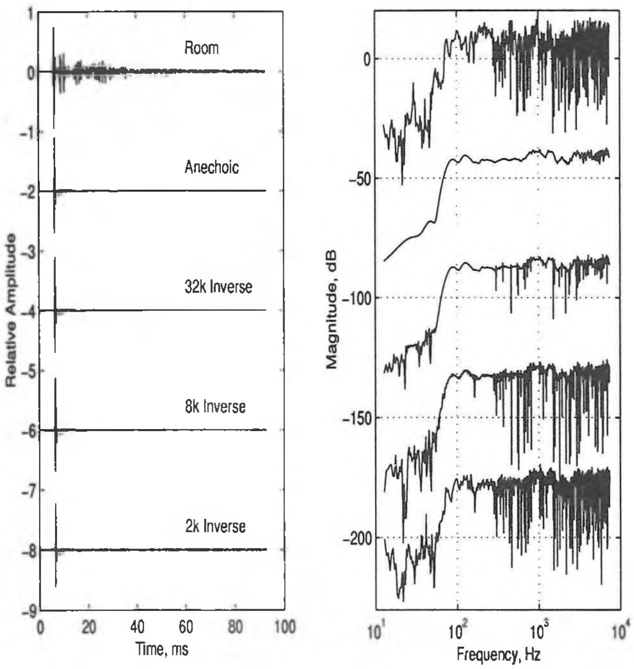

Figure 1: Room and Anechoic Impulse Responses and Magnitude Spectra for Loudspeaker (top two curves) and Equalized Outputs for Different Lengths of Inverse FIR.

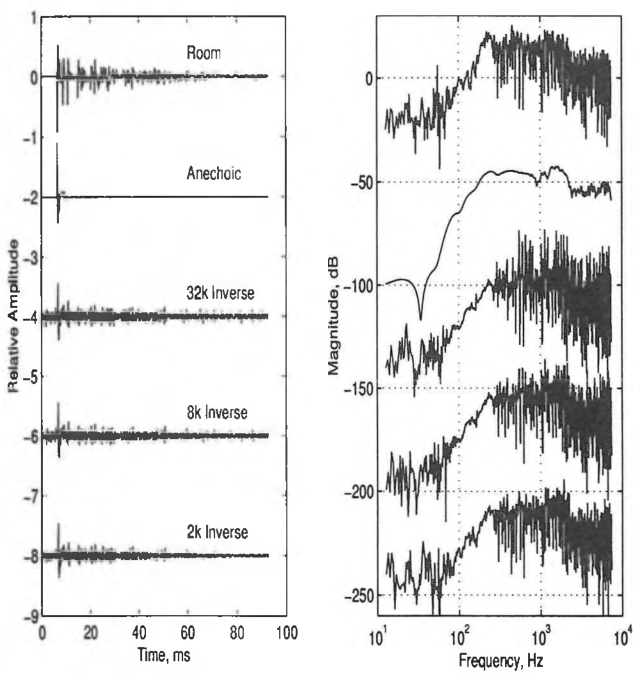

Figure 2: Room and Anechoic Impulse Responses and Magnitude Spectra for Mannequin (top two curves) and Equalized Outputs for Room Inverse Filter Designed with Loudspeaker.

5 Results

The lower three curves in Figures 1 and 2 show the result of post-filtering the impulse response as measured in the room, for the loudspeaker and the mannequin respectively, with inverse filters designed from the loudspeaker measurements.

For the loudspeaker, it is expected that longer inverse filters will yield a lower error [1], and this is supported by Figure 1. Notice, however, that some deep nulls in the room transfer function are not removed; this is not surprising since an all-zero (i.e., FIR) filter is not capable of removing zeros from a system transfer function, in particular nonminimum-phase zeros.

It is clear from Figure 2 that application of the inverse designed from the loudspeaker measurements does not effectively invert the room response for the mannequin measurements in any case. This mismatch is due to variations in the room response $g_{room}(n)$ due to different excitations.

Notice that the longer "inverse" filters are worse for dereverberation of the mannequin signals than the shorter ones. This is opposite from Figure 1, but is understandable since the longer filters are more sensitive to the fine details of the response from which they were designed, as opposed to the short ones which can equalize only gross distortions (such as room resonances), which may be common to both room responses.

6 Summary

The task of dereverberating received speech in a hands-free system is quite daunting. It has been shown that even for the case where the impulse response from source to receiver is known, the dereverberation is difficult and imperfect. Furthermore, an inverse filter designed from measurement of the room impulse response does not appear usable with a different source. In fact, if too long a filter is used, the speech can be degraded far more than if no equalization had been attempted at all. The usefulness of the least-squares inversion technique for dereverberation appears limited. Other techniques, such as beamforming, are known to be much less sensitive to source characteristics, and may prove to be the preferred choice.

Acknowledgements

The authors would like to thank Dr. Gilles Daigle at NRC and Dr. John Vanderkooy at UW for helpful criticisms and suggestions.

References

[1] J.N. Mourjopoulos, "Digital Equalization of Room Acoustics," *J. Aud. Eng. Soc.*, **42**(11), pp.884–900, 1994.

[2] S. Haykin, Adaptive Filter Theory, 3rd ed., Prentice-Hall, Upper Saddle River, NJ, USA, 1996.

[3] J. Mourjopoulos, "On the Variation and Invertibility of Room Impulse Response Functions," *J. Sound Vib.*, **102**(2), pp.217–228, 1985.

[4] S.T. Neely and J.B. Allen, "Invertibility of a Room Impulse Response," *J. Acoust. Soc. Am.*, **66**(1), pp.165-169, 1979.