CENTER FOR COMPUTER RESEARCH IN MUSIC AND ACOUSTICS

Department of Music
Report No. STAN-M-51

JUNE
(published at CCRMA May

REVERBERATION CANCELLATION IN MUSICAL SIGNALS USING ADAPTIVE FILT

Perry R. Cook

The technicques of adaptive filtering and noise cancellation are investigated as to their suitability for t purpose of removing reverberation from musical signals. Under the assumption that the reverber pattern is significantly uncorrelated from one location to another location within a room, a two micropho LMS adaptive filter system is tested.

First, reverberation models are studied, then various system models which represent the actual tv microphone pickup configuration are invetigated. Both causal and non-causal filters are tested w various values for the adaptation constant, and all pertinent data is tabulated. Best performance rest from a long, non-causal filter configuration.

This work was supported (in part) by the System Development Foundation. The views and conclusions containe this document are those of the author and should not be interpreted as necessarily representing the official polic either expressed or implied, of Stanford University or of the sponsoring foundation.

TABLE OF CONTENTS

INTRODUCTION

Reverberation is sound which reaches the listener (or listening device) through the reflection paths in a room or space. Typically a first few discrete echos (early reflections) occur, then these echos fuse into a more continuous sound. Often this continuous decaying sound is called reverberation, while the early reflections are called echo [1]. In both music and speech literature, the importance of room reverberation is a much considered topic. Using a speech signal, Lochner and Burger [2] determined that the perceived quality of a recording is affected, and often the intelligibility is greatly reduced by the room reverberation. Using musical signals, Sheeline [3] determined that a small amount of room reverberation helps the human spatial localization process, but excessive reverb destroys the ability to determine source location. Most authors [4] agree that the early reflections help in angular source localization (since their origins are related to the boundaries of the room, and thus help to orient the listener) and perform multipath filtering (called coloration). The later arriving diffuse reverberation tail is responsible for a different set of psychoacoustic effects, giving cues as to the room size and the source distance.

The purpose of this project is to research the removal of reverberation from a signal, using adaptive filter techniques. It is desirable to develop techniques for accomplishing this both in real time, and off line in a more controlled post-processing environment. Ideally this could be accomplished without any knowledge of the room itself or of the microphone placements used for sound pickup. Of course, the exact removal of all room effects is probably not possible, even if desirable, for many reasons. One such difficulty is that of removing a zero of the room impulse response. If the deconvolution filter is to cancel the zero, it must have a pole at the same frequency, and stability and noise enhancement problems can easily arise. However, the significant zeroes associated with a room response fall into the coloration (early reflection) category, and thus do not technically belong to that which we will call the reverberant tail.

Since the desired application is for music processing, the principle consideration in evaluating the performance of any system is the retention of the quality of the resulting perceived source instrument. The term 'perceived' is used because the final judgement must be made by human listeners, and many seemingly objectionable spectral operations are quite transparent to the human ear, while other minor changes to a signal are quite audible. If a process removes all room reverberation, but also destroys the character of the recorded instrument, then the process is not suitable. So as a qualitative figure of merit, if the process itself is audible, or if the perception of the timbre and attack of the instrument is significantly affected, then the process has failed.

APPLICATIONS

The applications for a reverberation removal processor are many, and more would arise if such a processor were readily available.

The first area of application is that of studio post-processing of live recordings. Typical modern practice is to record a performance live on a multitrack recorder (one track per instrument if possible), then take the multitrack tape back to a studio and mix it down to a master. Before mixing down, various sweetening operations can be performed; equalization, doubling of certain sections with studio recorded performances, actual replacement of some tracks with studio performances, and the addition of studio special effects [5]. Removal of the room reverberation from some or all tracks would provide the studio engineer with a powerful first step in post-processing, helping any of the above sweetening operations to be more effective. Another studio technique gaining some popularity is the use of a prerecorded track to control electronic music synthesizers via a real time pitch detector [6]. Such pitch tracking devices need the cleanest, most well behaved waveforms possible, and accurate removal of room reverberation could undoubtedly help the process.

Another possible application is that of noise reduction. Since the room reverberation is thought of as an unwanted noise, any background noise in the room (air conditioners, etc.) may also be eliminated as a side effect of the process.

Feedback control is another possible application. If the filte can truly invert the room response in real time, then the unstable standing wave patterns due to the multipath reflections could be removed.

Current research at CCRMA [7] and elsewhere in computerized automatic music transcription and performance study could be aided b removing the room signature from the signals before attempting any processing.

The final application is related to a particular class of instruments, those which are inherently tied to some room by their nature. Organs, carillons, and other such instruments are tied to whatever architecture into which they are built. In the case of carillons and other outdoor instruments, the nearby buildings make the reverberant boundaries for the listening space. In fact, the design and voicing of such instruments is usually dictated by the peculiarities of the buildings which house them. If adaptive room removal were available, such instruments could be studied 'outside' their shells, and interesting cross couplings of instruments with different rooms could be done. For example, perhaps the Stanford Church organ could be heard with the reverb patterns of Carnegie Hall, rather than those of Stanford Church.

HISTORY OF ATTEMPTS

Of course, the most effective method of achieving a recorded performance without room reverberation is to record in a room with nc reverberation. In fact, recording studios spend much money on architecture to ensure a non-invasive room sound. An unobjectionabl reverb pattern is usually preferred over no reverberation at all, because a purely anechoic room is actually not desirable, even if possible. In general, humans feel uncomfortable in a room which is completely anechoic, and musicians cannot play together and in tune in such an unnatural environment. Aside from these considerations, live performance recording by definition is only possible in real concert hall type rooms. Also, as mentioned before, many instrument are tied to their particular room, and thus cannot be recorded without that room's reverberations.

The live music sound system industry depends heavily on graphic and parametric equalizers, which are used for 'fixing' the room soun (as well as other special effects uses). Such equalizers are hand adjusted, with the human ear as the final judge, until the room 'sounds right'. Such filtering can only undo the most gross colorations, however, and typically does nothing about the time that the reverberation persists (a strong auditory cue as to room size an reflectivity).

Various non-linear gain control devices are commonly employed t remove background noise, and some applications to reverb removal hav been attempted. A noise gate is a two gain-state amplifier, which i at high gain when the input signal is greater than some arbitrary threshold level, and off (or greatly attenuated) otherwise. Some hysteresis and some slew in changing gain levels is usually employed to avoid amplitude modulation when the input signal is fluctuating about the threshold. An expander is an amplifier with a non-linear gain characteristic. Based on the average level of the input signal the expander changes gain in a non-linear and increasing fashion. Thus the loud sounds are made louder, and the soft ones softer. Since the reverberant tail is typically at a lower volume than the source, both of the above devices can attenuate reverb when there is no source present. The problem is that both of these processes are typically quite audible, often destroying the attack portion of musical events. Also there is sometimes an audible 'pumping' effect on repetitive sound patterns.

In 1977, Allen, Berkley, and Blauert [8] of Bell Labs attacked the problem of automatically filtering the room reverberation from prerecorded speech signals. They employed a multi-microphone proce which affects both the coloration due to early reflections, and the long term reverberant tail. The processor works in bands, first removing delays by a technique called phase shift and add. Then the gain of each band is adjusted using a normalized cross-correlation function. Frequency bands containing uncorrelated signals are shut off, and those containing correlated signals are left open. This technique relies on the assumption that the source component in each microphone is correlated with the other microphone signals, and that the reverberation in each microphone is uncorrelated with the reverberation in the other microphones. Binaural recordings were made in rooms with reverberation times ranging from .1 to 2 seconds, then these recordings were digitized to 12 bits and processed. For recordings made in rooms with reverb times greater than .5 seconds, the authors reported dramatic (no quantified results in the paper) reductions in reverberation.

Other attempts at solving the room response problem involve actually inverting the room's impulse response. In 1979, Neely and Allen [9] of Bell Labs worked on the problem of characterizing various impulse responses as to their invertibility. Here the assumption is that of knowing, or being able to determine accurately the actual impulse response of a given room and microphone placement

The traditional approach in measuring the impulse response of a room is to excite the room with an impulse, but an impulse creates difficulties related to generation, spectral content, initial overload, and signal to noise ratio. A more robust technique for the measurement of impulse response was addressed in 1983 by Borish and Angell [10] of Stanford. They propose the use of pseudo-random noise as excitation, allowing greater control over the power and spectrum of excitation. The impulse response of the room can then be extracted by cross correlating the input noise with the room output This technique requires no knowledge of the room or microphone technique, but does require the excitation noise to be recorded alo with the room response.

REVERBERATION MODELS

A survey of the models used for simulating reverberation seems relevant to the approach taken in reverberation removal. As was mentioned in the introduction, the events occuring during reverberation can be divided into two sections; the early reflections, and the reverberant tail. The transition is ill-defined, because it takes place smoothly as the reflections occur closer together, eventually fusing until discrete echos are undetectible. Figure 1 shows a typical reverberant impulse response.

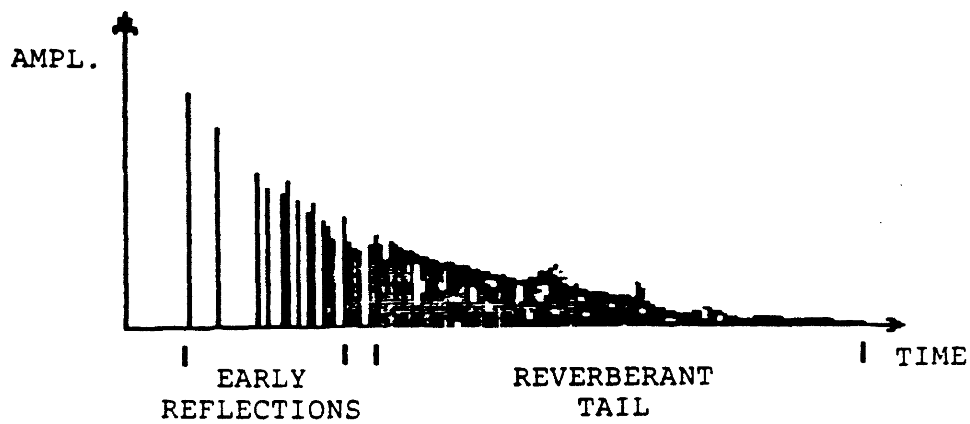


Figure 1 Room Impulse Response

It would seem that the ideal way to model reverberation would be to simulate the reflectivity of each wall, and trace wavefronts around the room until the energy has dissipated sufficiently. This points to a system of recursive filters, corresponding to the delays and decays encountered in a real room. Moorer [11] of CCRMA and IRCAM states, however, that simulating reverb with finite systems of recirculating delays cannot ever correspond to the reflection patterns of real rooms. His explanation for this disparity is diffusion, which seems to consistently confound any purely geometric approach to wave problems. Irregular walls, with frequency dependent reflectivity patterns, are the principle contributors to the diffuse quality of reverberation. Indeed in simple, rectangular, highly reflective rooms, where the geometric models are the most correct, the reverberant sound is not as 'normal' to our ears.

The solution of Moorer and others seems to be to model the earl reflections explicitly, using an appropriate geometric model, then switch to a more complex set of filters for the diffuse reverberant tail. Figure 2 shows a block diagram of such a reverberator. There are many opinions on how to simulate the diffuse late-reverberant portion. It has even been suggested that convolution of the input signal with exponentially damped uncorrelated white noise produces a very pleasant and believable reverberation pattern.

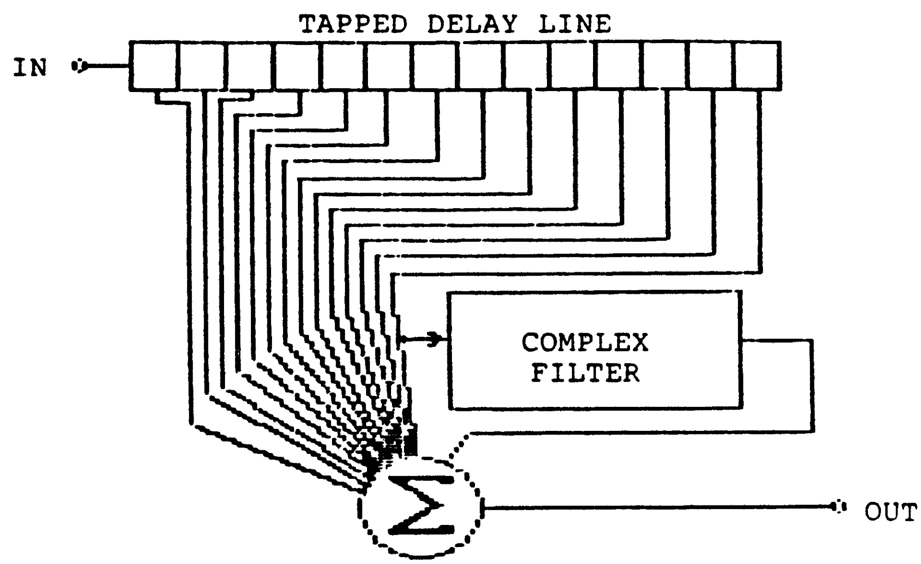


Figure 2 A Reverberation Simulator

In modeling the early reflections, a technique is commonly employed which involves computing each reflection as if it had originated from another source. Borish [12][13] and Allen [14] call this technique 'image modeling', and the secondary sources are calle source images. Moorer calls them phantom sources, and another commo term is virtual source. The order of a virtual source is equal to the number of walls that the actual source wave reflects from to generate a particular source image. Figure 3 shows a source in a room and some of its images.

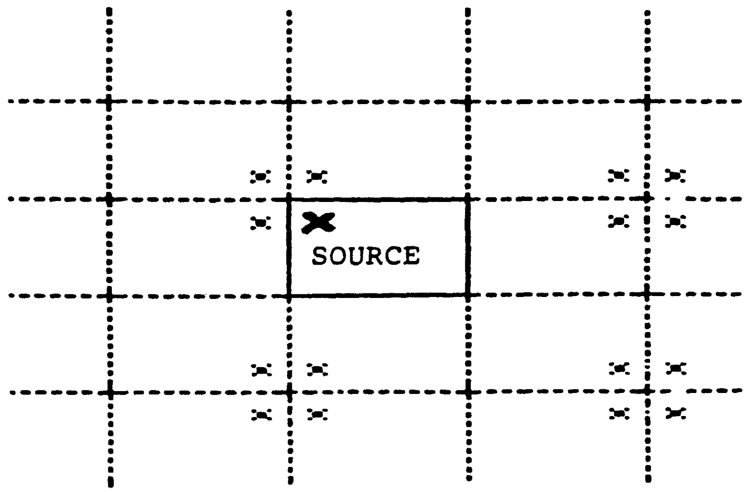


Figure 3 The Image Model

Borish [15] states that since the traditional stereo and monaural playback configurations cannot accurately reproduce the placement of early reflections, steps are taken to remove them using microphone placement. Directional microphones are employed to rejec the sound arriving from the rear, and thus sound arriving from the direction of the source is emphasized. His decision in designing a digital early reflection enhancement device was to simulate only the first few echos, and depend on the recording to contain the diffuse reverberation. He based this on his assumption that the early reflections are deemphasized by the recording process.

SYSTEM MODELS

The system composed of the source, the room, and the recording chain is a complex one, and must be modeled sufficiently if the problem is to be solved. Here various models for the entire system are presented. The problem of removing the effects of a filter without any knowledge of the filter itself is a classic adaptive systems problem, and typically some reference signal is required to derive the desired response. As a result, the decision was made to assume a two microphone system, with one microphone placed close to the source, and the other used as a reference. Figure 4 shows a typical recording setup.

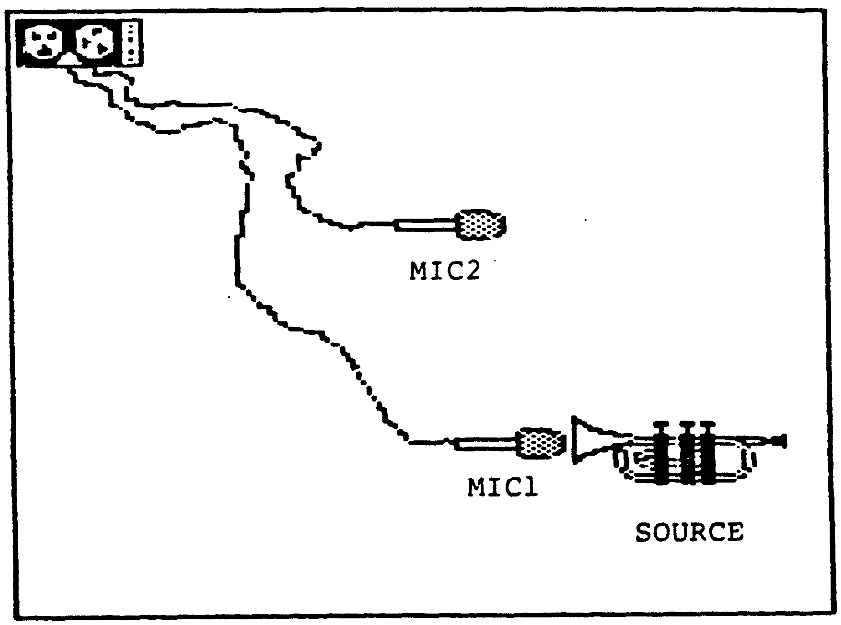


Figure 4   Recording setup

The first model contains the most assumptions. So many assumptions are made, in fact, that the reverb removal solution is trivial, and does not require an adaptive filter. Figure 5 shows th block diagram for the first model.

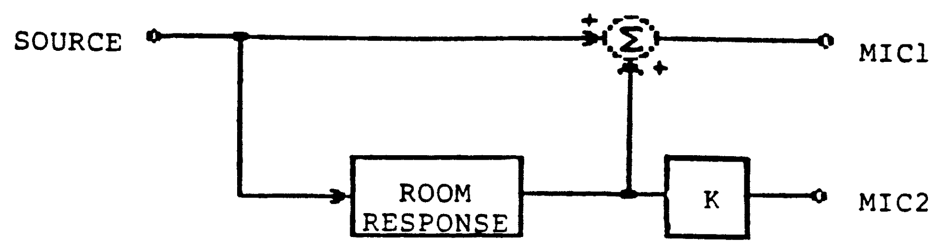


Figure 5 Simplest Model With Most Assumptions

The assumptions for the simplest model are;

1. The signal in mic1 is composed of source and reverb.
2. The signal in mic2 is composed only of reverb.
3. The reverb signal in mic2 is the same as that in mic1, except for perhaps a scale factor.

Under these assumptions, the solution is trivial;

Mic1 = source + reverb

Mic2 = k(reverb)

Form the difference signal given by;

Diff = Mic1 - (Mic2)/k

= (source + reverb) - (k(reverb))/k

= source

The problem with these assumptions is that basically none of them are true. The actual conditions are better described by the next model.

The antithesis of the last model is the one which makes no assumptions except linearity throughout the signal path. Each possible modification that the signal can go through on its way to the adaptive filter is included. Figure 6 shows the block diagram o this model.

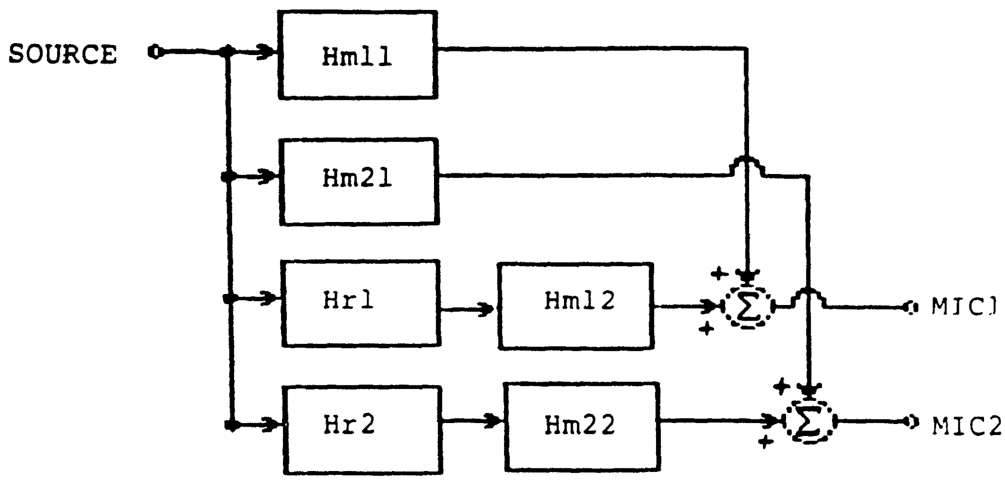


Figure 6 Model With The Least Assumptions

The transfer functions Hm11 and Hm21 are the filtering operations that take place on the signal on its way through the two microphones (this also includes any filtering in the amplification and recording processes). Hr1 and Hr2 are the impulse responses of the room from the source to the microphones. Hm12 and Hm22 are the filtering operations that take place on the reverberated signals as they enter the microphones and pass through the remaining signal path. It should be pointed out that Hm12 and Hm22 are quite different from Hm11 and Hm21, as the off-axis arrival of reverberant signals will cause acoustic filtering to occur that is different frc that acting on a signal arriving on-axis (the source).

The microphone signals are given by;

$$Mic1 = source*(hm11+hr1*hm12)$$

$$Mic2 = source*(hm21+hr2*hm22)$$

Since Hm11 is the natural filtering that would occur to the signal if the room effects were not present, it is not absolutely necessary to remove it. But the Mic1 signal contains a sum of filtered versions of the source. The problem is to remove at least the effects of hr1*hm12 from the Mic1 signal. Or more specifically, using both signals, somehow arrive at hr1*hm12 and deconvolve by this filter. Deriving the ideal deconvolution form in the frequency domain;

$$\text{Mic1} = \text{SOURCE}(\text{Hm11}) + \text{SOURCE}(\text{Hr1})(\text{Hm12})$$

$$= \text{SOURCE}\ (\text{Hm11}+\text{Hr1}(\text{Hm12}))$$

Thus the required filter for inversion is;

$$\text{Hrecovery} = 1/(\text{Hm11}+\text{Hr1}(\text{Hm12}))$$

And of course, there is no guarantee as to the stability of this filter, if it could be found.

A slightly relaxed version of the no assumptions model is shown in Figure 7.

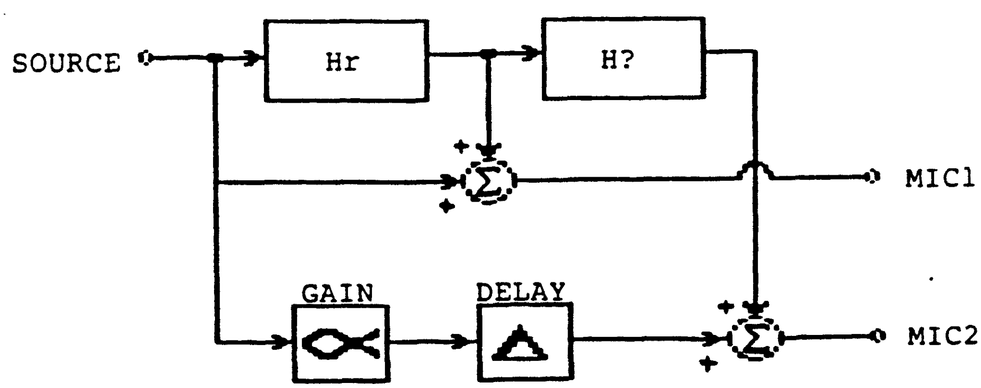


Figure 7 A More Relaxed Model

The assumptions for the diagram of Figure 7 are;

1. The source appears in both microphone signals, but is greatly attenuated (20 to 60 dB) in the mic2 signal.

2. There is some fixed delay associated with the mic2 signal component.

3. The reverb signals in both mics are related by some (unknown) filter function.

4. Any off-axis and proximity effects of the microphones are negligible.

Hr is the impulse response from the room to the microphone 1 location. H? is the unknown relation between the reverberation in mic2 and that in mic1. The alpha delta branch denotes attenuation and delay simply due to the different signal path lengths.

The microphone signals are now;

$$\text{Mic1} = \text{source} + \text{source}*\text{hr}$$

$$= \text{source}*(\delta + \text{hr})$$

$$\text{Mic2} = \text{Delayed and attenuated source} + \text{source}*\text{hr}*\text{h?}$$

In the frequency domain;

$$\text{Mic1} = \text{SOURCE} + \text{SOURCE}(\text{Hr})$$

$$= \text{SOURCE}(1 + \text{Hr})$$

$$\text{Mic2} = \alpha(\text{SOURCE})Z^{-\Delta} + \text{SOURCE}(\text{Hr})(\text{H?})$$

$$= \text{SOURCE}\ (\alpha Z^{-\Delta} + \text{Hr}(\text{H?})$$

So $$\text{Hrecovery} = 1/(\alpha Z^{-\Delta} + \text{Hr}(\text{H?}))$$

Again the stability of the required filter inversion is questionable.

The final system model takes advantage of a recurring characterization found in the literature. As already mentioned numerous times in this paper, the reverberant tail has a noise-like quality. Benade [4], Allen et. al. [8], Moorer [11], and Borish [1i all make mention of the noise-like and uncorrelated nature of the diffuse part of the reverberation. The multi-microphone reverb removal scheme described in the HISTORY OF PAST ATTEMPTS section of this paper relied on the reverberation in two microphones being uncorrelated with themselves and each other. Thus, based on the literature, this seems like a good assumption to make.

I offer a further heuristic justification of the reverb-as-noise assumption. When examined, the image model presents similarities to a computer generated random number sequence. Many computer pseudo-random number generators use a modulo-type operation to randomize and keep the numbers within a certain range. An example random number generator forms a number by adding pi to the last random number, raising it to some power, then taking the fractional part. A popular integer type of random number generator forms random numbers by multiplying the last random number by a multiplier (where the multiplier is large enough to ensure overflow). If the product of the multiplier and the last random number, modulo the word length, is numerically prime, a pseudo-random number sequence is generated. The impulse response obtained from the image model can be formed by incrementing by the speed of sound times the sampling period, modulo the wall dimensions. Thus a modulo type psuedo-random impulse response is not particularly counter intuitive.

The block diagram of the reverb-as-noise model is shown in Figure 8. The simplicity of the model indicates that solution might be simpler also. H1 and H2 are the filter responses from the source to mic1 and mic2 and through the remainder of the signal chain. The reverb is modeled as two independent additive noises.

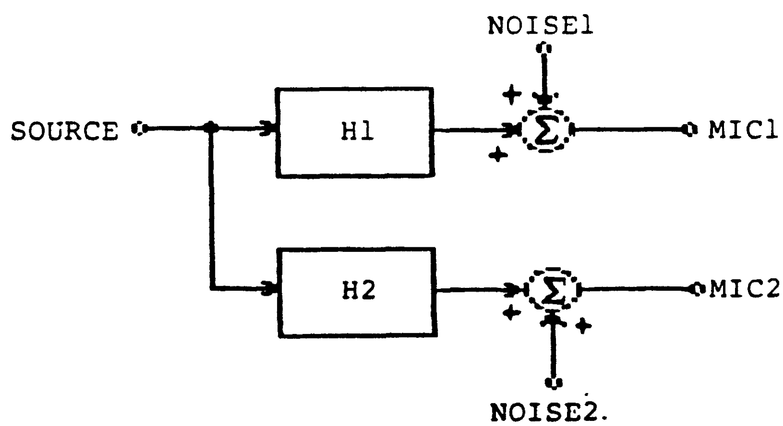


Figure 8 Model For Reverb-As-Noise Assumption

ADAPTIVE FILTER CONFIGURATION

Under the reverb-as-noise assumption, the problem of reverberation removal becomes one of noise removal. Any filter configuration which adapts itself to reinforce that part of both signals which is mutually correlated, while cancelling the part whic is uncorrelated, will accomplish the task to some degree. Noise cancellation configurations and algorithms are abundant, as this is one of the earliest and most common applications for adaptive filter [16].

Figure 9 shows the configuration used for adaptive reverb removal. Inside the box is an FIR tapped delay filter, a linear combiner whose inputs are the delayed samples of the input signal. The arrow leading from the error signal through the box denotes some algorithmic adaptation scheme. The filter adapts so that the filtered signal is as close as possible to the component of the desired signal which is mutually correlated with the signal being filtered. Thus the desired signal is selected to be the mic1 (close microphone placement) signal, and the input to the filter is the mic (what we call the reference mic) signal.

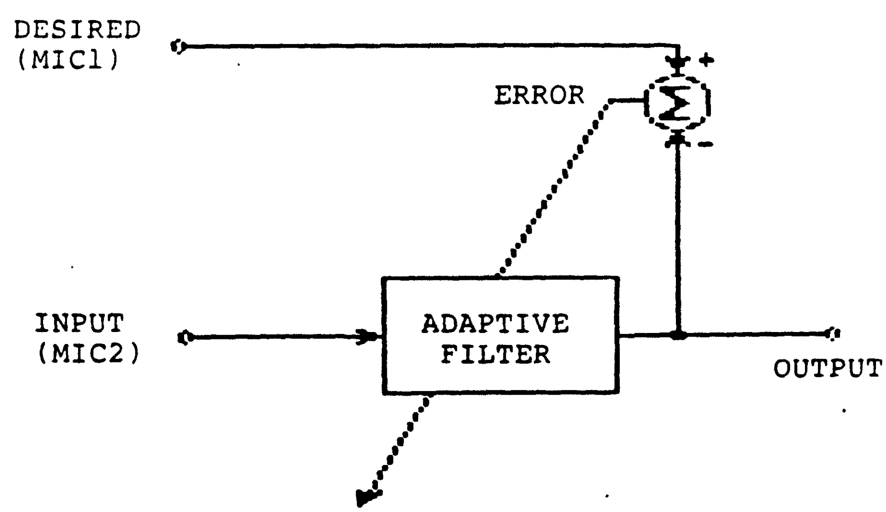


Figure 9 Adaptive Filter Configuration

An interesting filter results when both signals are composed entirely of mutually uncorrelated noise. Under this special input signal circumstance, the optimum weight vector of the filter is the zero vector. Intuitively, this says that the best job that a linea filter can do (in a least squares sense) to make a signal look like signal which is completely uncorrelated with it is to provide no output at all.

Proof;

(Vectors are notated with an arrow overscore)

Denote two uncorrelated noises as n_1 and n_2.

Denote their power spectral densities as σ_1^2 and σ_2^2.

The desired signal is $d_k = n_1$

The input signal vector is $\overline{x_k} = \overline{n_2}$

The weight vector is $\bar{w}$

The error signal is $e_k = d_k - \bar{w}^T(\overline{x_k})$

$= n_1 - \bar{w}^T(\overline{n_2})$

The error squared signal is $e_k^2 = n_1^2 - 2n_1\bar{w}^T\overline{n_2} + \bar{w}^T\overline{n_2}^2\bar{w}$

The mean square error is

$$xi = E[n_1^2] - 2\bar{w}^T E[n_1\overline{n_2}] + \bar{w}^T E[\overline{n_2 n_2^T}]\bar{w}$$

$$= \sigma_1^2 - \emptyset + \sigma_2^2\bar{w}^T\bar{w}$$

Which is minimum at $\bar{w}^* = \emptyset$, with a value of σ_1^2

For this project, the decision was made to take the output of the adaptive filter block as the system output, because complete noise elimination is possible under the above condition of uncorrelated noise inputs. In the presence of desired signal input, the filter output should adapt to the characteristics of the correlated part of the desired signal. Thus the best results could be achieved if both microphone signals contain significant amounts c the direct source, while being placed far enough apart to ensure a high relative uncorrelation for the room impulse response.

The algorithm used for this research is the Widrow Hoff [17] Lr adaptive algorithm, chosen primarily for its ease and speed of implementation. The LMS algorithm uses the error squared signal as an estimate of the mean square error. This yields an estimate for the gradient which is twice the negative of the current error times the current input vector. Thus no elaborate matrix computations ar required, multiple estimates need not be taken, and it is not necessary to perturb the weights for gradient measurement. It was assumed that if good results could be achieved using LMS, on future research more exotic algorithms could be tried with the confidence that the overall approach is sound.

The weights can be started in any state without affecting the eventual optimum solution (although certainly the initial value affects the time to reach the optimum vector). Since the event bei tested was known to begin with source signal, the weights were all set to zero except the first one, which was set to one. With this initial weight vector, some signal will appear at the output from t first incidence of input signal, and thus a transient attack would not be lost.

TEST SIGNAL CONSTRUCTION

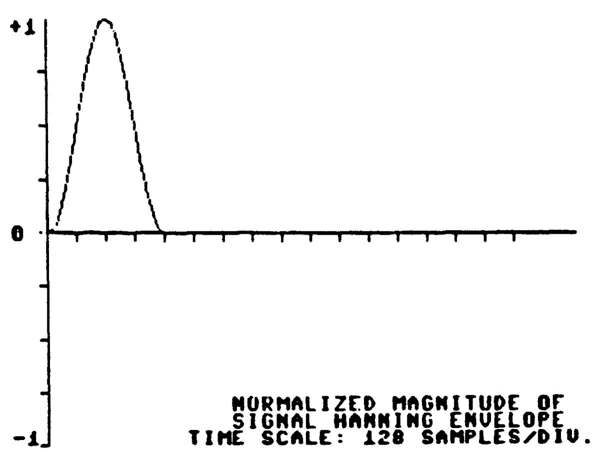
To construct appropriate test signals, there were two primary concerns;

1. Generate an event that has most of the important characteristics necessary to simulate real conditions.

2. Keep the event short enough to allow large amounts of processing and evaluation to be done.

The duration of the entire signal event was chosen to be a very short 2048 samples, with the expectation that this would not be long enough for an appropriate evaluation. Surprisingly, the filter was able to adapt quickly enough, and the 2048 length trial was adaquate for testing. Keeping to this short length allowed FFT evaluation of the entire event, which proved quite useful for comparing different filter lengths and mu values.

The source portion of the event was chosen to be 512 samples (the first 1/4 of the total event). A Hanning window was used as the source signal envelope, because of its good behavior under the FFT. The reverberant noise signal begins at sample 256, when the source is at its maximum volume, and persists until the end of the 2048 sample frame. The envelope for the reverberation begins as a Hanning shape then merges into a damped exponential at sample 530. By the 2047th sample, this exponential has damped to .0166 times the maximum noise envelope value. Care was taken to match both the levels and the derivatives of the Hanning and exponential curves at their junction point, so no artifacts of discontinuity could arise. Figure 10 show the envelopes of the source and the noise. Figure 11 shows the resultant envelope.



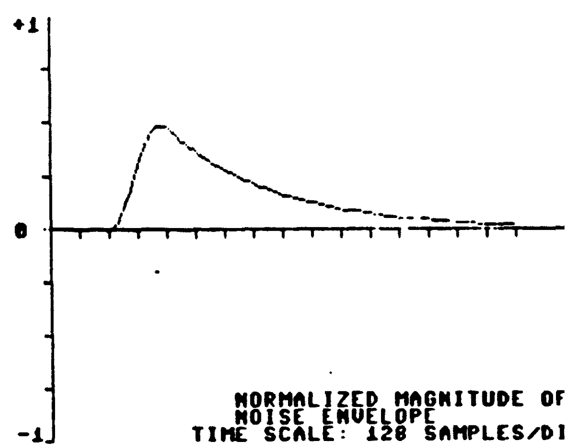


Figure 10 Source Envelope Reverb Envelope

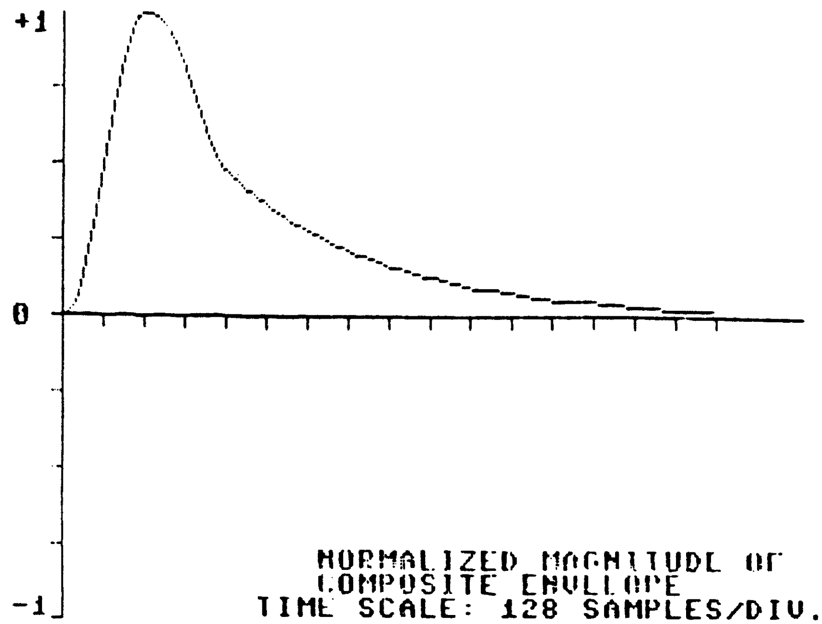


Figure 11   Resultant Signal Envelope

The source signal was chosen to be complex enough to adequately test the filter, but simple enough to easily identify subtle changes using the FFT. A square-like wave was built using three sinusoids, yielding 3 distinct peaks on the FFT. All signal plots throughout the paper are normalized to the peak of the source signal. The actual data was packed into an 8 bit signed binary format, so the peak value is 127. The mic1 source signal component is;

$$(\text{Hanning Window})*(.7\sin(.1n)+.2\sin(.3n)+.1\sin(.5n))$$

The windowed square wave source component of the mic1 signal is shown in Figure 12, a 512 point FFT of the source event is shown in Figure 13, and a 2048 point FFT (equivalant to zero padding by a factor of 3), is shown in Figure 14. These transforms will form part of the basis for filter evaluation later in the paper.

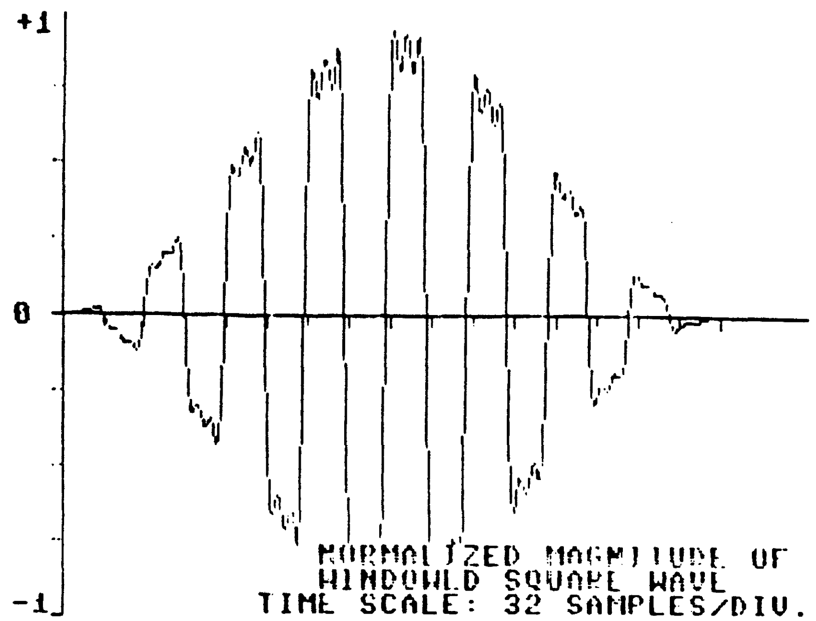


Figure 12 Source Signal Mic1

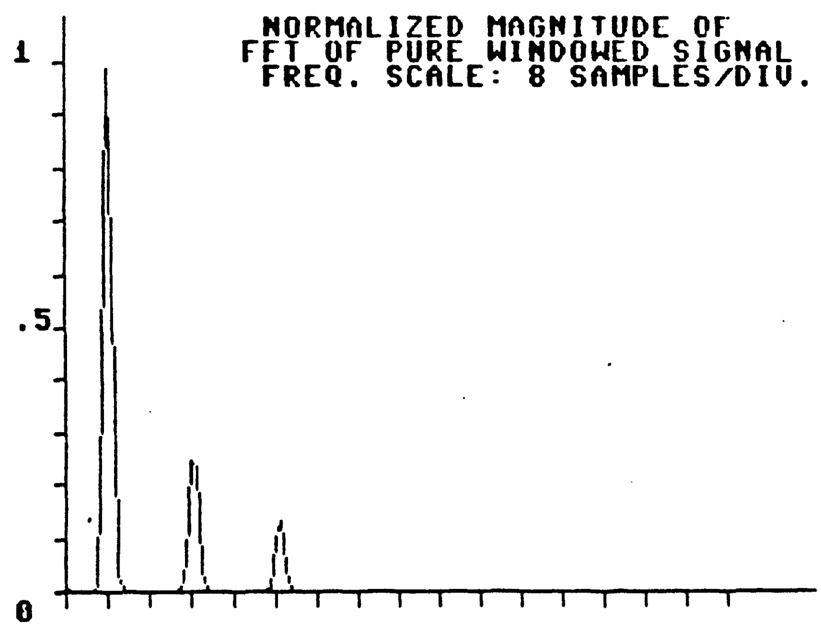


Figure 13 Short Term FFT Of Source Signal

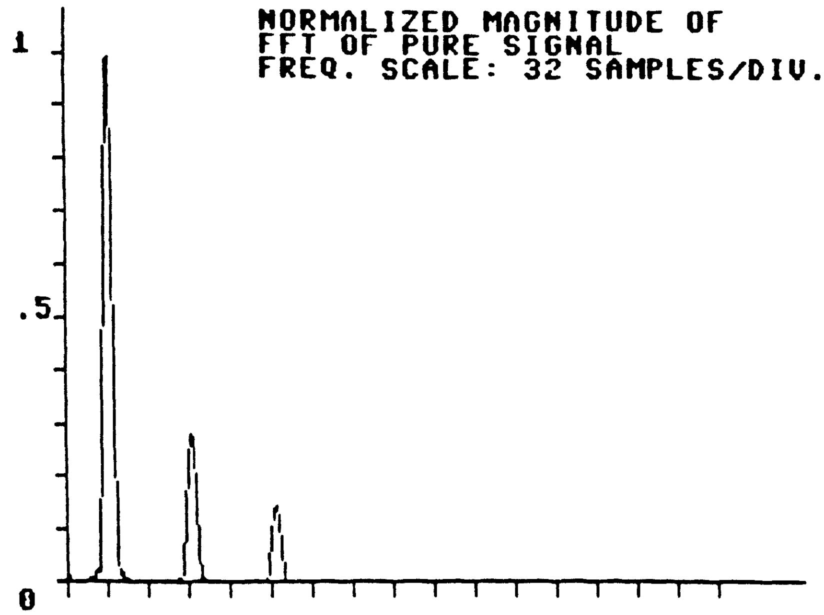


Figure 14 Zero Padded FFT Of Source Signal

The final mic1 signal was formed by adding the Hanning windowed square wave to the Hanning/exponential windowed noise. The noise wa generated as a standard uniform computer pseudo-random number sequence. The noise signal component was scaled to peak at 1/2 of the source signal amplitude. Figure 15 shows the final resultant test mic1 signal, and Figure 16 shows a 2048 point FFT of the mic1 signal.

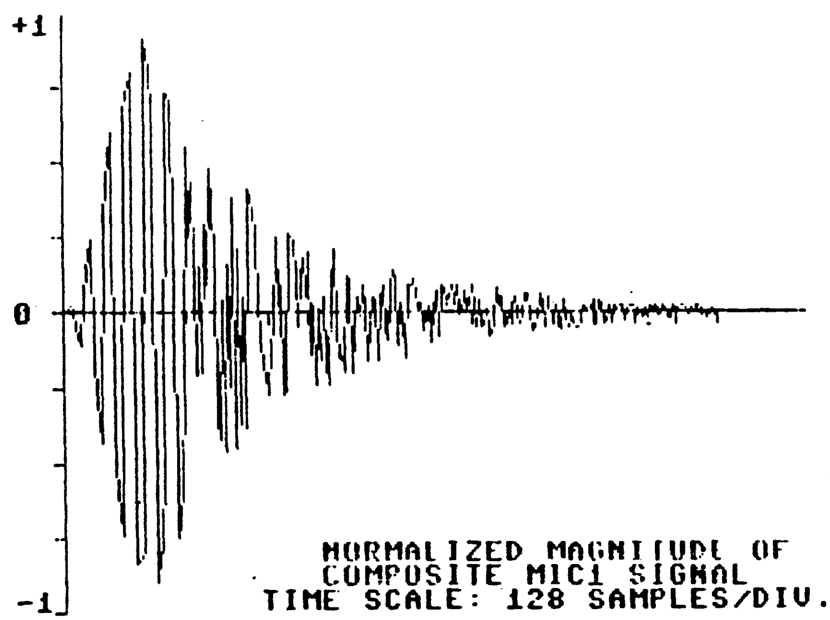


Figure 15   Total Composite Mic1 Signal

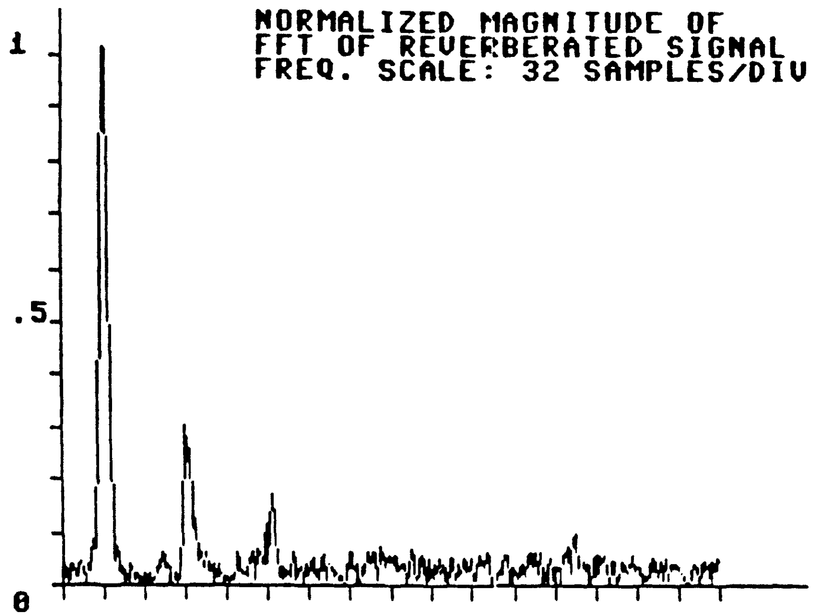


Figure 16   FFT Of Composite Mic1 Signal

An important figure of merit when discussing reverberation is the reverb time, defined as the time required for the energy in the reverberant tail to drop to some fraction of the peak excitation energy. RT60, the time until a 60 dB intensity reduction is reached, is commonly used in acoustics. For the purposes of this study, RT40 was selected, because the 8 bit signed data format only affords 42 dB of resolution. For the mic1 signal, the RT40 was mathematically determined to be 2079 samples. An experimental trial was run by taking a 20 sample running variance of the actual mic1 data file, and the RT40 was determined to be 2030 samples.

The mic2 signal is very similar to that of mic1, but appropriate realistic differences were introduced. The source signal component contains the same three sinusoids, but with reduced amplitudes. This simulates the filtering that might take place due to air absorption, microphone off-axis and proximity effects, and other signal chain effects. The entire source event, both sinusoids and window, were also delayed by seven samples. This is to account for the longer path length from the source to mic2, versus the short path from the source to mic1. The seven sample delay would translate to about 8 inches of path length difference at 10kHz sampling rate.

The source component of the mic2 signal is;

(Hanning W dow)*(.3sin(.1(n-7))+.15sin(.3(n-7))+.03sin(.5(n-7)))

The reverb envelope is identical to that of the mic1 reverb, beginning at the same time. As per our assumption, the uniform noi: sequence is uncorrelated with that in the mic1 signal. Figure 17 shows the total mic2 signal, and Figure 18 shows a 2048 point FFT of the mic2 signal. Note that the highest frequency sinusoid peak is completely lost in the reverberant noise floor.

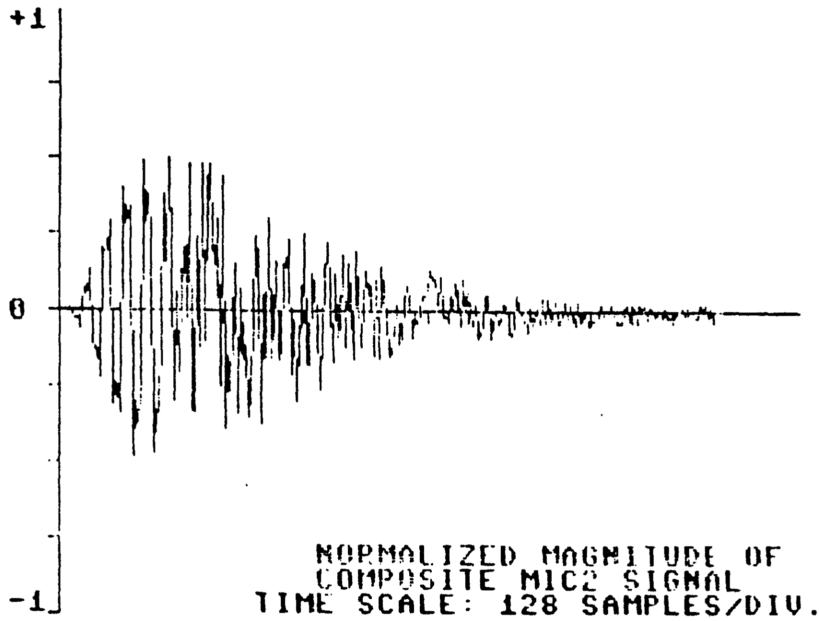


Figure 17 Total Composite Mic2 Signal

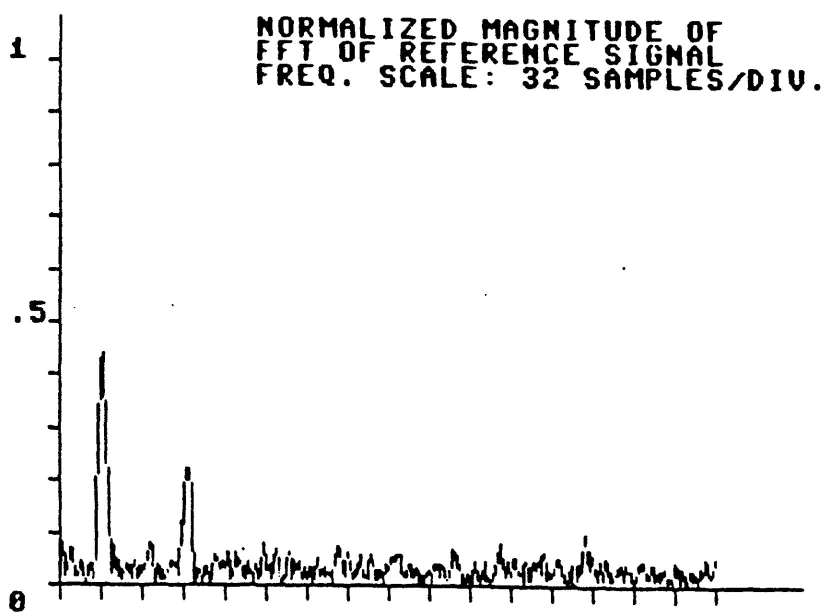


Figure 18 FFT Of Composite Mic2 Signal

FIRST FILTER TRIALS

The first adaptive filter attempted was an arbitrarily chosen 11 weight, causal (no pre-delay in desired signal) filter. The maximum stable value of the adaptation constant mu is given to be the recipro of the largest eigenvalue of the input autocorrelation matrix. A mor conservative estimate uses the reciprocal of the trace of this matrix Since the signals used in all of the simulations were 8 bit signed numbers, the peak value could be as great as 127. This number square 16129, and in the case of a perfect square wave, can be taken to be t maximum signal power. If this value is assumed to be the dominant eigenvalue of the input autocorrelation matrix, a very unconservative estimate for the maximum mu value of 6E-5 results. The first test va for mu was chosen to be 5E-5.

Running a trial with the above mu value of 5E-5 yielded unsuitab results. The filter weights diverged beginning at about the 190th sa and the output was an unstable oscillation. Mu was adjusted to 4.5E- and another, somewhat more successful trial was run. The results wer still unsuitable, with the filter output clipping (exceeding the peak +/-127 by over 400 at one point) a total of 81 times. The norm of th weight vector at the 2047th sample was .0457, which was very close to predicted value of 0 for the pure noise input case.

An actual analysis of the specific signal properties was done, u the peak portion of both pure signals (the noise power at any point i the reverberant tail is much less than the source signal power at its peak). Forming the 11x11 input autocorrelation matrix yields;

$$
R = \begin{bmatrix}
2252 & 2223 & 2137 & 1999 & 1821 & 1610 & 1381 & 1148 & 921 & 712 & 528 \\
2223 & 2252 & 2223 & 2137 & 1999 & 1821 & 1610 & 1381 & 1148 & 921 & 712 \\
2137 & 2223 & 2252 & 2223 & 2137 & 1999 & 1821 & 1610 & 1381 & 1148 & 921 \\
1999 & 2137 & 2223 & 2252 & 2223 & 2137 & 1999 & 1821 & 1610 & 1381 & 1148 \\
1821 & 1999 & 2137 & 2223 & 2252 & 2223 & 2137 & 1999 & 1821 & 1610 & 1381 \\
1610 & 1821 & 1999 & 2137 & 2223 & 2252 & 2223 & 2137 & 1999 & 1821 & 1610 \\
1381 & 1610 & 1821 & 1999 & 2137 & 2223 & 2252 & 2223 & 2137 & 1999 & 1821 \\
1148 & 1381 & 1610 & 1821 & 1999 & 2137 & 2223 & 2252 & 2223 & 2137 & 1999 \\
921 & 1148 & 1381 & 1610 & 1821 & 1999 & 2137 & 2223 & 2252 & 2223 & 2137 \\
712 & 921 & 1148 & 1381 & 1610 & 1821 & 1999 & 2137 & 2223 & 2252 & 2223 \\
528 & 712 & 921 & 1148 & 1381 & 1610 & 1821 & 1999 & 2137 & 2223 & 2252
\end{bmatrix}
$$

The cross correlation vector of the desired signal and the input signal is;

$$P^T = \begin{bmatrix} 2853 & 3271 & 3680 & 4061 & 4390 & 4644 & 4805 & 4860 & 4805 & 4644 & 4390 \end{bmatrix}$$

With the inverse of the input autocorrelation matrix, and the desired/input crosscorrelation vector, the optimum weight vector can easily be found;

$$W^* = R^{-1}P$$

$$= \begin{bmatrix} -8.54 & 12.01 & 6.25 & -6.90 & -3.78 & -1.54 & 1.58 & -5.71 & 8.49 & 8.30 & -7.92 \end{bmatrix}$$

The minimum mean square error at this optimum weight vector valu given by;

$$xi_{min} = E[d_k^2] - P^TR^{-1}P$$

$$= 10800-10560 = 240$$

This implies that an average absolute error of 15.5 in 128, or a 12%, could be achieved. In order for this to occur, however, the sig would have to be in steady state at their maximum values long enough allow the weight vector to adapt to the optimum value. The time cons of adaptation is an inverse function of the eigenvalues and the adaptation constant, and the very small (effectively zero) eigenvalue indicate that it would take forever to actually reach the optimum wei vector.

The trace of the R matrix is 11(2252)= 24772. Using this for a conservative estimate for the largest eigenvalue yields a mu value of 4E-5. The actual vector of eigenvalues was determined for the R matr

$$EV = \begin{bmatrix} -2.48 & -2.46 & -.337 & .439 & .632 & .819 & 1.318 & 15.78 & 413 & 4457 & 19888 \end{bmatrix}$$

Obviously this matrix is rather ill-conditioned, with the larges eigenvalue equal to 19888, and the smallest one equal to -2.48. The input autocorrelation matrix is by definition positive semi-definite, the small negative eigenvalues must be interpreted as being caused by rounding errors, and thus the R matrix can be assumed to have only fo non-zero eigenvalues;

$$EV = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 15.78 & 413 & 4457 & 19888 \end{bmatrix}$$

Using the largest eigenvalue of 19888 leads to a maximum mu esti of 5.2E-5. As was stated above, however, this range proved to be unsuitable, even if stable. Continuing with the mathematical analysi inverting the R matrix yields;

$$R^{-1} = \begin{bmatrix}
-.38 & .34 & .48 & -.29 & .048 & -.23 & -.21 & -.16 & .33 & .61 & -.51 \\
.34 & .005 & -.90 & .30 & -.12 & .55 & .27 & -.13 & .062 & -.85 & .61 \\
.49 & -.90 & .35 & .000 & -.02 & -.13 & .41 & .54 & -1.0 & -.06 & .33 \\
-.30 & .30 & .000 & .41 & -.10 & -.39 & -.51 & .33 & .55 & -.13 & -.16 \\
.05 & -.12 & -.02 & -.10 & .47 & .02 & -.26 & -.51 & .41 & .27 & -.21 \\
-.23 & .55 & -.13 & -.40 & .02 & .42 & .02 & -.40 & -.13 & .55 & -.26 \\
-.21 & .27 & .41 & -.51 & -.26 & .02 & .47 & -.10 & -.19 & -.12 & .05 \\
-.16 & -.13 & .55 & .33 & -.51 & -.40 & -.10 & .41 & .000 & .30 & -.29 \\
.33 & .06 & -1.0 & .55 & .41 & -.13 & -.02 & .000 & .35 & -.90 & .005 \\
.61 & -.85 & -.06 & -.13 & .27 & .55 & -.12 & .30 & -.90 & .005 & .34 \\
-.51 & .61 & .33 & -.16 & -.21 & -.26 & -.05 & -.29 & .48 & .34 & -.39
\end{bmatrix}$$

The second trial was done using a mu value of 5E-6, which is approximately 1/10 of the maximum stable mu value. The results were strikingly better. The filter output only exceeded the 127 bit maximum range once (by 5), and the time domain waveforms appear much more well behaved. Figure 21 shows the output signal of the second filter trial. Again the reverberant signal is attenuated greatly, and Figure 22 shows a better waveform, although still not very squarelike. The RT40 (defined in the section on test signal generation), is 1066, which is a reduction by a factor of 1.9 as compared to the original reverb time.

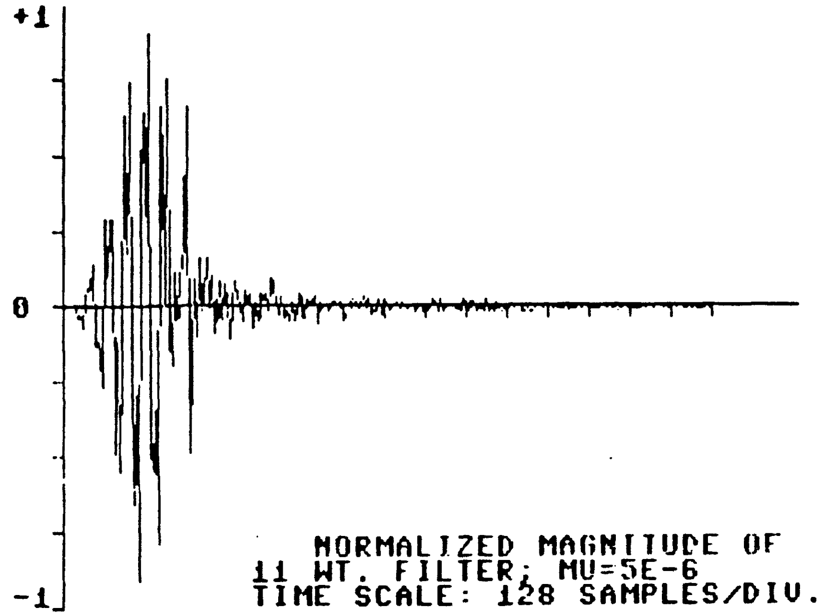


Figure 21 Second Trial Output

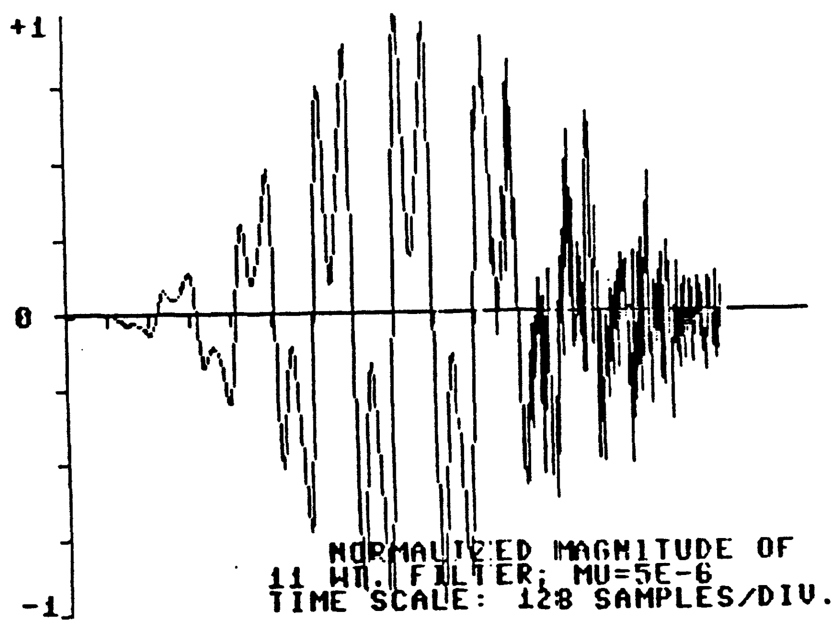


Figure 22 Close Up Of Source Signal Block

The resultant filtered signal output for the first 11 weight filter trial using a mu value of 5E-5 is shown in Figure 19. It can easily be seen that the reverberant tail is greatly reduced from the original (compare to Figure 15). Figure 20, however, points up the distortion to the signal even in the source only portion, which should appear more like a square wave. When the reverberant noise entered at sample 256, the filter output reacted violently, creating more distortion than in the original mic1 signal. Based on the time domain results, along with the clipping data, no further analysis of this trial was done.

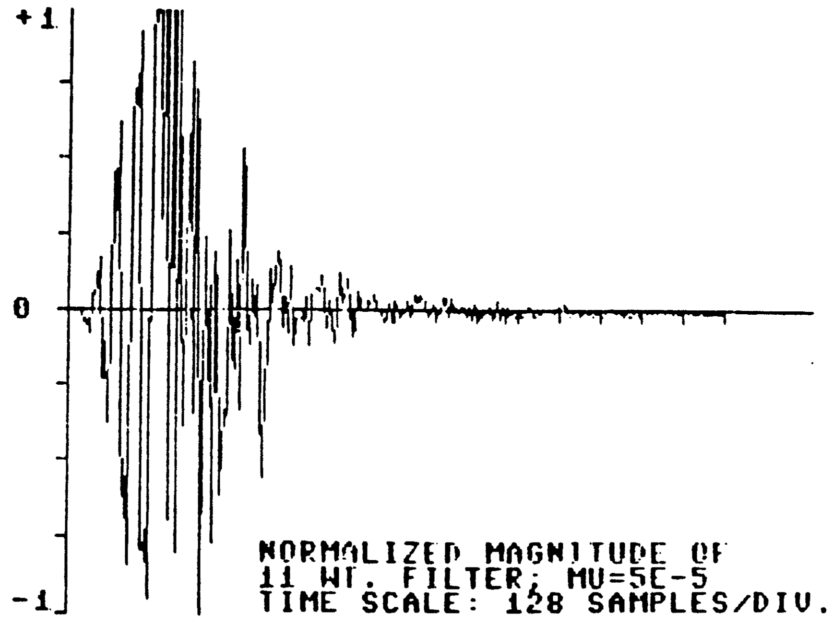


Figure 19 First Trial Output

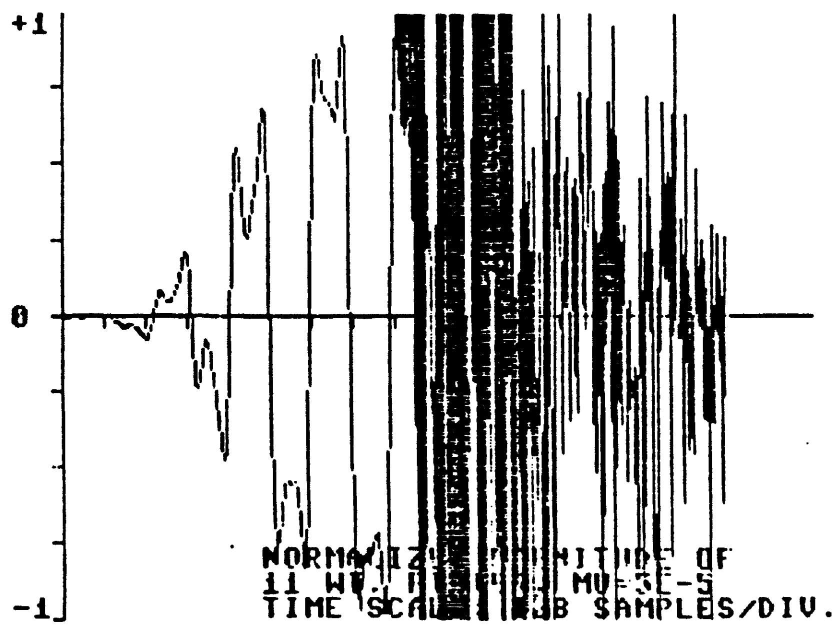


Figure 20 Close Up Of Source Signal Block

Figure 23 shows an FFT of the second trial, mu = 5E-6, 11 weight filtered signal. Clearly the sine peaks were not restored to their desired values, but there is a noticeable decrease in the noise floor. It should be pointed out that the highest sine peak was 'pulled out' of the noise floor quite well, as the signal being filtered is the mic2 signal (called the reference microphone), whose third sinusoid is not even visible in Figure 18.

The first sinusoid peak reflects a -5.54 dB error from the desired value, the second reflects a +3.81 dB error, and the third a -3.098 dB error. The signal peaks were removed, and an average was taken over the squares of the remaining FFT bins [18] to get the average noise power. Comparing this to the average noise in the original signal, the total reverberation noise reduction throughout the entire event is 7.3 dB.

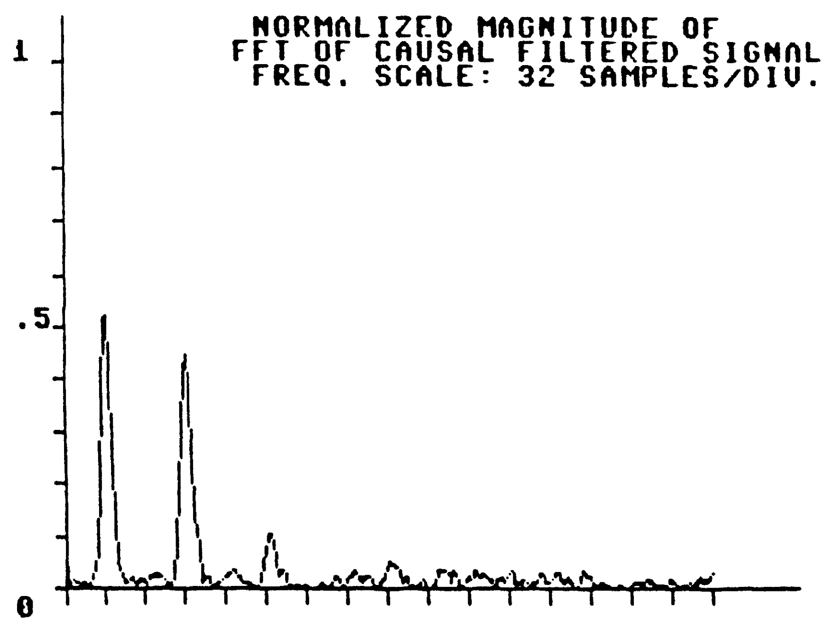


Figure 23 FFT Of Second Trial Output

A third trial was performed with the 11 weight filter, reducing the mu value by another factor of 10 to 5E-7. This proved too slow to react, merely putting out a slightly modified copy of the mic2 signal. The time domain performance was so poor, the data was not analyzed further. Figure 24 shows the filter output for this trial, and Figure 25 shows the source signal block.

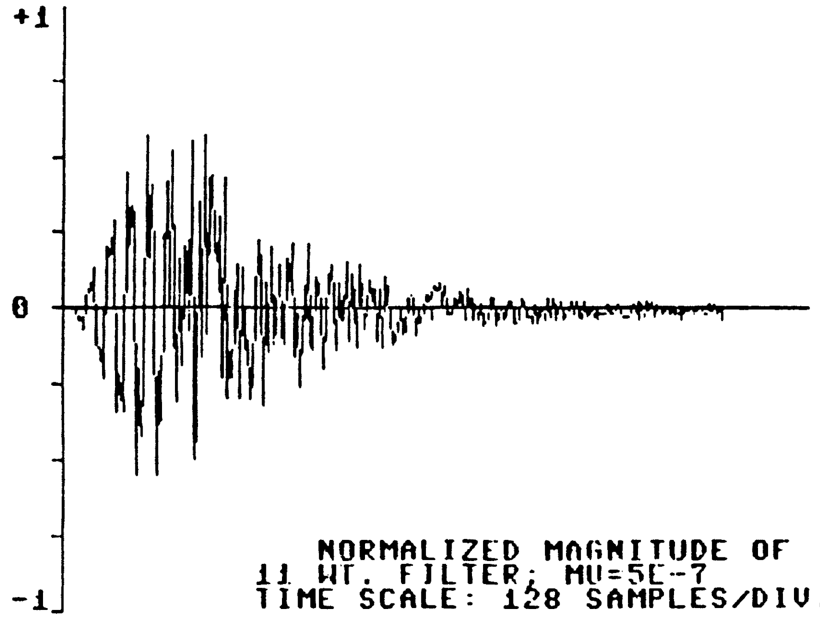


Figure 24 Third Trial Output

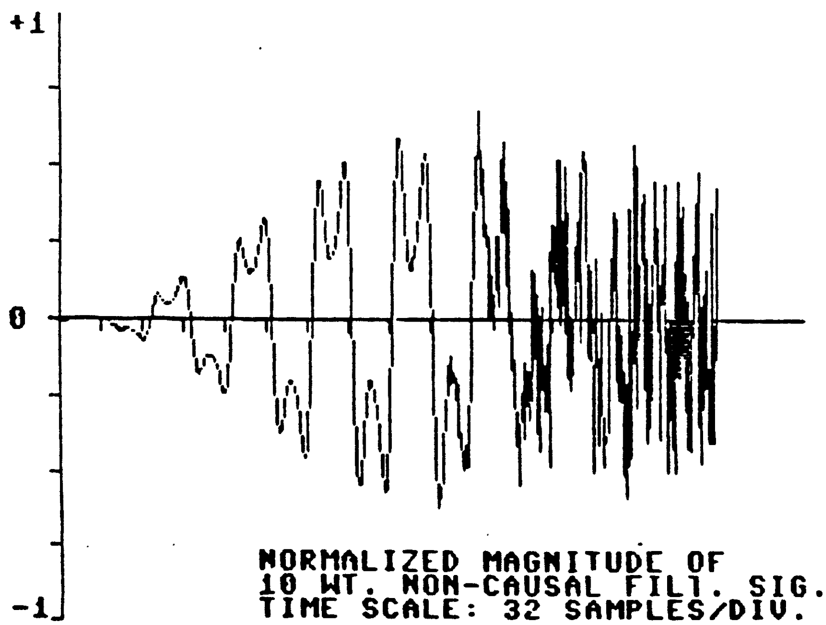


Figure 25 Close Up Of Source Signal Block

SECOND FILTER TRIAL

The second filter used was a 20 weight, non-causal form. A significant advantage is gained if the filter can look into the future, and adjust itself accordingly. This non-causality is implemented by placing a pre-delay of 1/2 the filter length (10 samples in this case) in the path of the desired signal. Figure 26 shows a block diagram of this modified filter form. Only one trial was run, with a mu value of 5E-6. The decision to run with this mu value was based on the favorable results obtained with the first filter using this value. Also the dominance of the R matrix by the one large eigenvalue, indicates that the stable mu for a 20 weight realization is most likely very close to that of the 11 weight filter. Thus a mu equal to 1/10 of the maximum stable value should translate well to a larger filter. Again the weights were started all equal to zero, except the first.

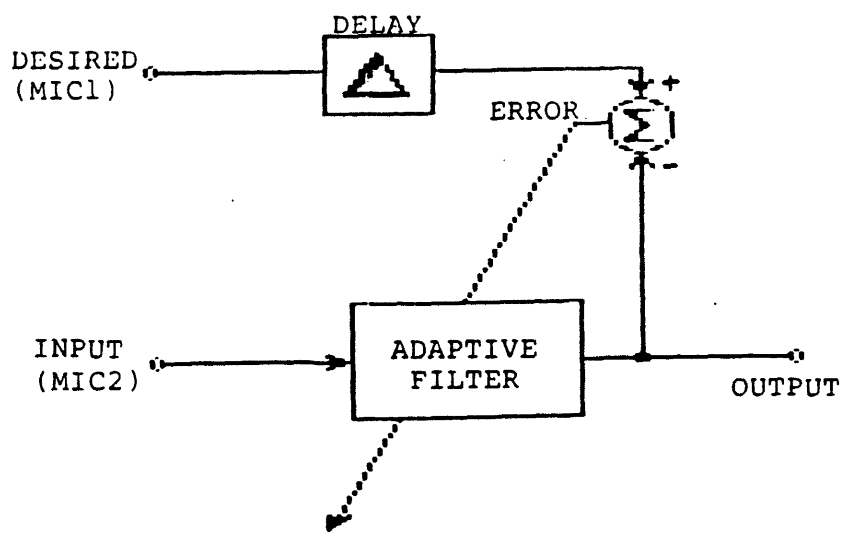


Figure 26   A Non-Causal Filter Realization

The output signal resulting from this filter is quite remarkable. Figure 27 shows the total 2048 length output signal, a Figure 28 shows the 512 sample source signal block. Note that the signal appears much more squarelike, and keeps its shape even when the reverberant noise is introduced. The reverberation reduction i almost as good as the 11 weight filter, with an RT40 of 1386.

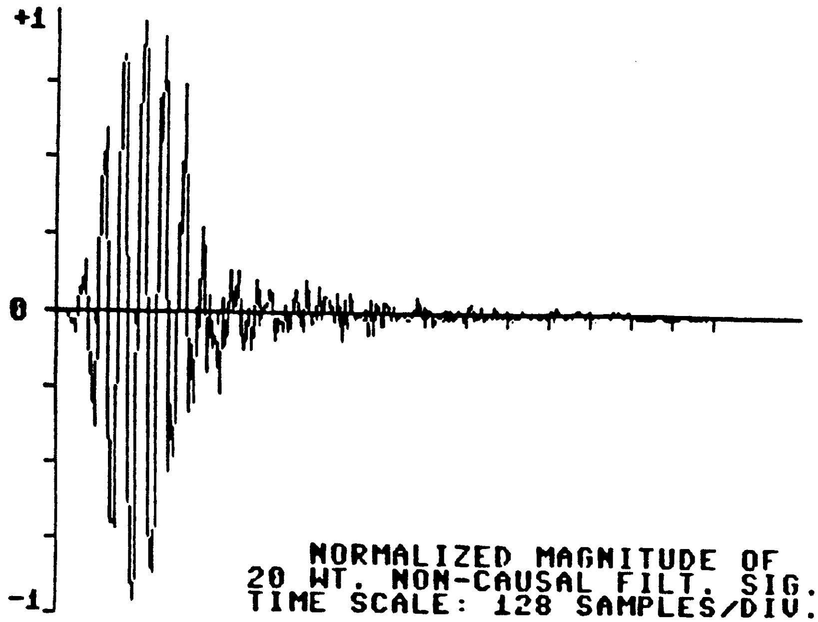


Figure 27   Output From Non-Causal Filter

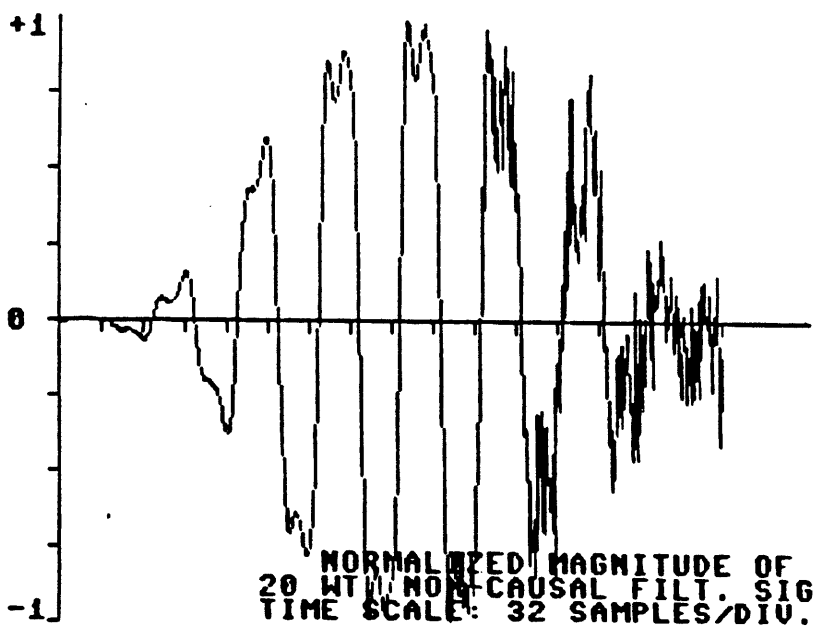


Figure 28   Close up of Source Signal Block

Figure 29 shows an FFT of the non-causal filter output. As expected from the appearance of the time domain signal, the source restoration is quite good. The lowest sinusoid was restored with -.25 dB error, the second one with -.45 dB error, and the third with -8 dB error. The noise power reduction was 6.89 dB.

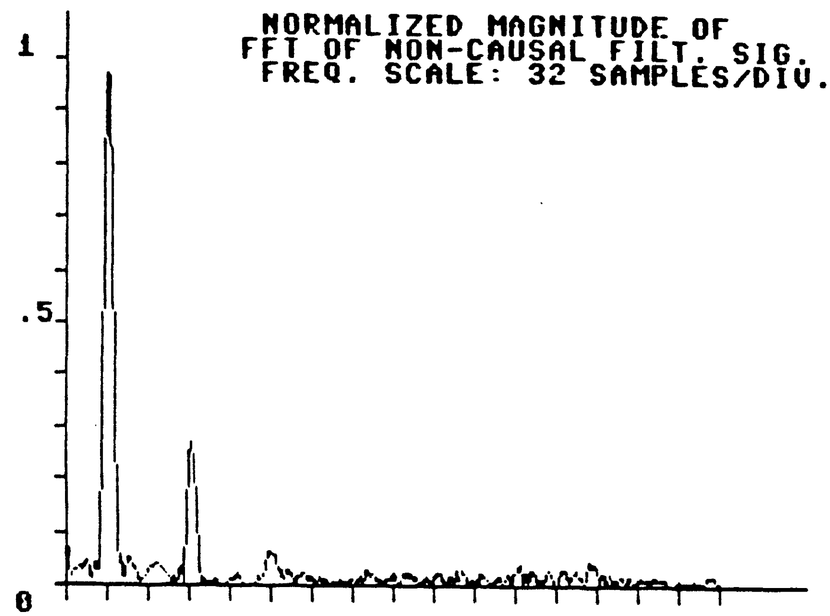


Figure 29 FFT Of Non-Causal Filter Output

Further analysis was done on the output of the 20 weight filter The 2048 sample event was divided into 4 blocks of 512 samples each. Thus the first block contains source signal plus reverberant noise, and the remaining three blocks contain reverberant noise only. An FFT of each block was performed for both the original mic1 signal an the 20 weight non-causal filtered signal. Figure 30 shows these spectral plots for the first three blocks only, as the fourth block noise power is too low to be interesting. The left column is the original mic1 signal, and the right column is the filter output.

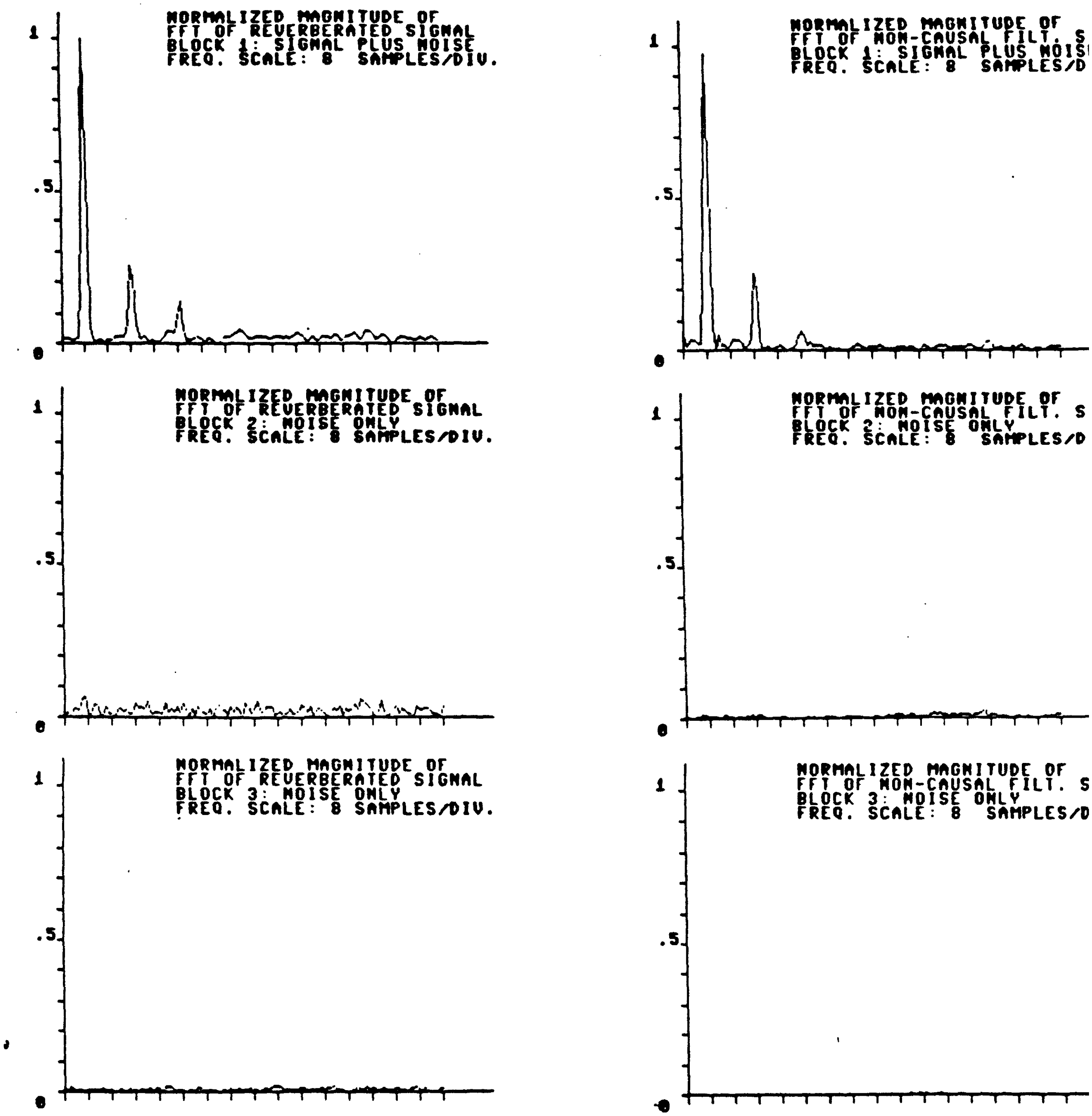


Figure 30 Mic1 Block FFT Plots Filter Block FFT Plots

The noise components of each block were compared. The pure source signal was subtracted from block 1 for this analysis, so any error from the desired signal is also interpreted as noise. The noise gain for the first block is a startling .3524 dB, considering that both noise and error in signal restoration contribute to the noise figure. The noise gain for block 2 is -13.08 dB. The noise gain for block 3 is -12.64 dB, and the noise gain for block 4 is -11.52 dB. Note that the average noise output in block 4 is -42.24 dB over the entire block, which is below the least significant bit resolution of the quantization.

FIRST FILTER REVISITED

Since we have introduced a new set of figures of merit and evaluation criteria in analyzing the non-causal filter with the bloc analysis, it seems appropriate to return to the first filter and evaluate it using this new technique. No plots were made but the significant numbers were computed. The block 1 noise gain was +8.6(dB, due to both source signal error and noise component. The block noise gain was -13.08 dB. The block 3 noise gain was -14.88 dB, an the block 4 noise gain was -11.82 dB. So this filter was better at eliminating the noise during the blocks where noise was the only signal, as should be expected from a shorter filter. But the degradation of the source signal during the first block is objectionable, as was indicated in the Figure 23 FFT plot.

FINAL FILTER TRIAL

There is a natural question to ask after comparing the 20 weig non-causal filter to the 11 weight causal filter. Will a 10 weight non-causal filter yield the extra noise reduction of the shorter filter, while adaquately restoring the signal? The third filter trial was conducted with just such a filter. The mu value of 5E-6 was retained, and the filter was implemented with 10 weights and a sample pre-delay on the desired signal. Figure 31 shows the output of this filter, and Figure 32 shows the close up of the source sign block. The filter clipped 13 times during the event, and the resultant RT40 for the filtered signal was 1071.

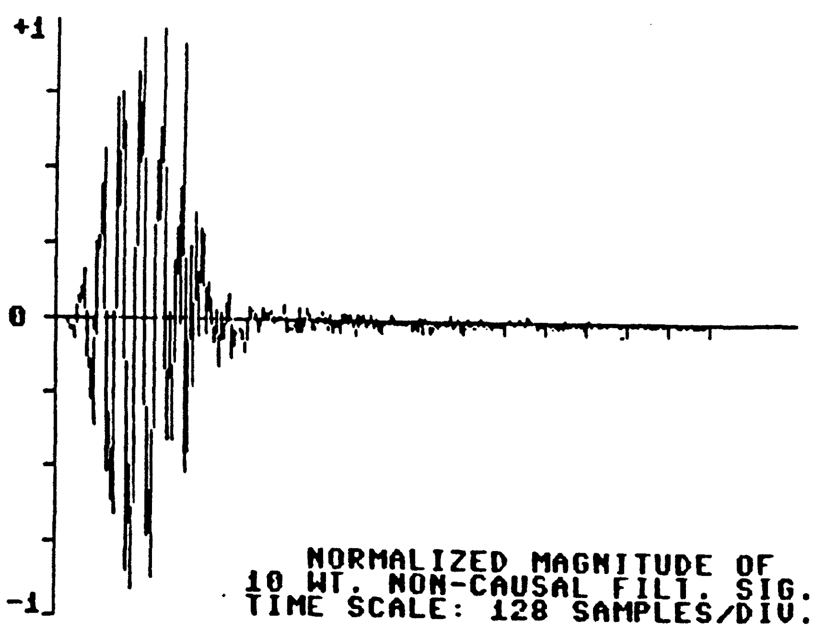


Figure 31 Third Filter Output

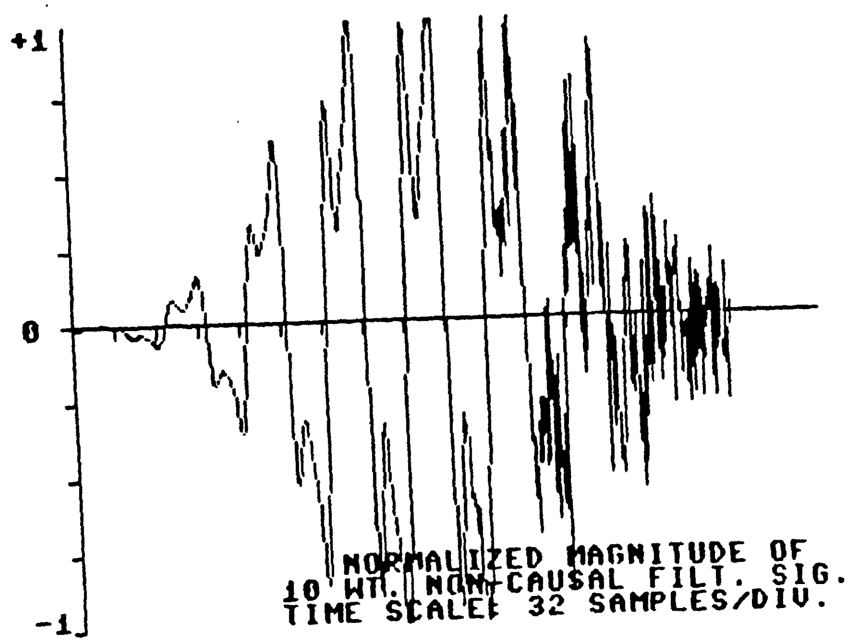


Figure 32 Close Up of Source Signal Block

Figure 33 shows the FFT of the entire signal event. The sinusoids were restored with errors of -2.6 dB for the lowest, +3 dB for the second, and -.0145 dB for the highest. The total event noise reduction was 5.4172 dB.

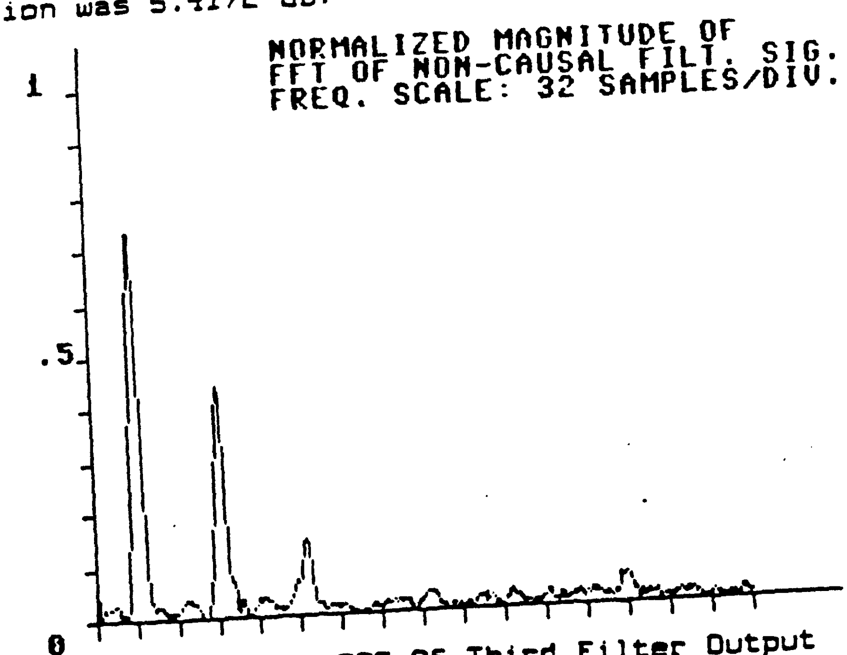


Figure 33 FFT Of Third Filter Output

The block analysis revealed comparable performance to the 11 weight filter. The noise gain for the first block was +4.2 The noise gain for the second block was -12.94 dB. The noise g for the third block was -14.15 dB, and -11.77 dB for the fourt block.

INTERPRETATION OF RESULTS

Table 1 shows the results of the significant filter trials. Included is the 11 weight filter at three values of mu, the 10 weight non-causal filter at one mu, and the 20 weight non-causal filter at one mu.

FILTER	11 Wt. Causal	11 Wt. Causal	11 Wt. Causal	10 Wt. Non-Caus.	20 Wt. Non-Caus.
Mu Val.	5E-5	5E-6	5E-7	5E-6	5E-6
Clipped	81 Times	1 Time	None	None	13 Times
1st Sin Error		-5.54 dB	-34.72 dB	-2.6 dB	-.25 dB
2nd Sin Error		+3.81 dB	+.494 dB	+3.94 dB	-.45 dB
3rd Sin Error		-3.098 dB		-.0145 dB	-8.0 dB
Total Noise Reduc.		6.29 dB		5.42 dB	6.37 dB
Block 1 Noise Reduc.		-8.96 dB		-4.24 dB	-.35 dB
Block 2 Noise Reduc.		+13.08 dB		+12.94 dB	+11.67 dB
Block 3 Noise Reduc.		+14.88 dB		+14.15 dB	+12.64 dB
Block 4 Noise Reduc.		+11.82 dB		+11.77 dB	+11.52 dB
RT40 Reduc. Factor		.5205		.5229	.6768

TABLE 1 Filter Performance Data

In examining the performance of each of the filter configurations, we must review the original goals which were discussed in the introduction. The ideal result is complete removal of all of the effects of the room, with no degradation of the original source signal. When evaluated according to this criterion, all of the filters tested failed miserably. Of, course, any filter for any application probably fails the ideal behavior test.

Even if the filters are evaluated as to whether they removed al reverberant tail effects (after the source signal portion of the event was over), they still fail. The reason for this however, is the gradient search type of algorithm used. The filter can only mov toward the optimal solution in a geometric manner, and thus never actually achieves perfect performance. Perhaps another type of algorithm would move more directly to the zero weight vector, thus allowing rapid shut off of the reverberant tail.

A more relaxed expectation would be simply the removal of some reverberant effects. In this area, all of the filters tested succeeded to some degree, with ranges from 7 to 15 dB of reverb reduction. The simple removal of some reverb, however, is not enoug to lead us to consider a technique successful. The reduction must b significant enough to warrant the expense and time of processing, ar our original prime goal of source integrity must have been met.

In general, the shorter filters were better at removing reverberation for a fixed mu value, as should be expected of short filters (the convergence is related to the number of weights). The general reverb removal performance of the two short filters was nearly identical, at about 13 dB noise reduction (based on the block analysis). The performance of the 20 weight filter was only slight lower than the shorter filters, at 12 dB of reverberant noise reduction.

In terms of the primary goal of restoring the source signal without audible degradation, future research on real recordings wit listening tests will have to be performed before statements can be made with certainty. However, using known psychoacoustic propertie we can at least make some statements about the performance of vario filters. Since the human ear is mostly sensitive to magnitude information, and only weakly sensitive to phase information, plots and data of the source signal were made on a magnitude basis. Usin this information as a basis, only one filter configuration performe adaquately at restoring the signal in the presence of reverberant noise. That filter was the 20 weight non-causal configuration, whi restored the signal with negligible error in the lowest two harmonics, and registered only 8 dB of error in the highest harmoni This 8 dB error takes place in a harmonic which is already 20 dB do from the total signal power, so it is conceivable that masking by t two lower harmonics might make the 8 dB error inaudible. Thus it i reasonable to state that the 20 weight non-causal filter succeeded removing some reverberant effects while leaving the signal relative unaffected.

It could also be stated that, in general, the non-causal filters performed much better at signal restoration. The 10 weight non-causal filter registered source harmonic errors averaging 2.19 dB (absolute error). When compared to the 4.14 dB average error of the 11 weight causal filter, the choice would seem clear between these two filter configurations. Both filters, however, had their greatest error in the first and second harmonics, attenuating the lowest harmonic, and making the magnitude of the second sinusoid much greater. This would easily be interpreted by the human ear as a gross source degradation error, and thus both of these filters must be considered as failing the primary goal of the reverb removal process.

The best performance of the filters at removing reverberation was about 15 dB. Whether this is the best performance that can be achieved is an open question. Widrow et al. [16], however, stated results of 20 to 25 dB of noise reduction on speech signals, so it is reasonable to expect better results in future reverb removal tests.

The resulting reverberation times of the filtered signals were varied, but basically consistant at less than half of the original reverb time. The reverb time reduction is consistant with the reduction of reverb level, as the exponential reverb envelope is simply shifted downward by processing. The reverberation time corresponding to a 15 dB reduction in reverberation level is;

$$-15 \text{ dB} \Leftrightarrow .1778 \text{ factor of volume}$$

$$(\text{Old RTxx})/(\text{New RTxx}) = -\ln(.1778)$$

$$= 1.7271$$

$$\text{New RTxx} = (\text{Old RTxx})/1.7271$$

$$= .579(\text{Old RTxx})$$

Which is consistant with the experimental findings. Thus, if results comparable to Widrow et al. could be achieved, a 25 dB noise reduction should result in a much better reduction in reverb time;

$$-25 \text{ dB} \Leftrightarrow .0562 \text{ factor}$$

$$\text{New RTxx} = (\text{Old RTxx})/2.878$$

$$= .3475(\text{Old RTxx})$$

CONCLUSIONS

It is desirable for a number of reasons to develop techniques for the removal of reverberation from audio signals with specific application to musical sources. The techniques of adaptive filteri and noise cancellation were investigated as to their suitability fo this purpose.

The assumption was made, based on the literature, that the reverberant pattern is significantly uncorrelated from one location to another location within a room. Two test signals were created. One test signal simulated the sound entering a main microphone whic would be placed close to the sound source. The other simulated the sound entering a second microphone, which would contain some source component, but would be placed far enough from the source to pick u a significantly different reverberation pattern.

Adaptive filter configurations based on the LMS adaptive algorithm were run, using three different filters and three differe values for the adaptation constant, mu. All filters eliminated the reverberation by about the same amount. Short filters were found t be unacceptable because of source signal degradation. Non-causal (pre-delay in desired signal path) filters were also found unacceptable for the same reason.

The filter which performed the best was a 20 tap, non-causal configuration. This filter restored the source signal to its original value very closely, and eliminated the reverberant noise b about 13 dB.

SUGGESTIONS FOR FUTURE RESEARCH

It is planned to continue the research in this area. Certainly the next step should be to perform tests on actual digitized sound files. The results from these tests would point to the direction for future research.

Provided that the data from tests on actual sounds proved sufficiently successful, there would be two logical next steps;

1. Perform listening tests to determine the aural success of different configurations of filters. This is critical under our original requirement that the process not interfere with the musical aspects of instrument sound.

2. Investigate the use of different adaptation algorithms. An optimum algorithm might be found which can 'track' the musical characteristics more faithfully, while more quickly converging to the optimum solution in terms of reverberant noise elimination.

The use of more than two microphones, as suggested by Ferrara and Widrow [19], can yield even better performance in a noise cancelling system such as this. This should be investigated further.

Finally, the author hopes that this system can eventually be expanded into a larger recording post-processing system, incorporating both reverberant noise cancellation, and systems to cancel interference from other source instruments within the same room.

REFERENCES

[1] J. Backus, *The Acoustical Foundations of Music.* New York: W. W. Norton, 1969.

[2] J. P. A. Lochner and J. Burger, "The Intelligibility of Speech Under Reverberant Conditions," Acustica 7, pp. 195-200, 1961.

[3] C. W. Sheeline, *An Investigation of the Effects of Direct and Reverberant Signal Interactions on Auditory Distance Perception.* Ph.D. Thesis, Department of Hearing and Speech, Stanford University, November 1982.

[4] A. H. Benade, "From Instrument to Ear in a Room: Direct or Via Recording," Journal of the Audio Engineering Society, vol. 33, no. 4, pp. 218-233, April 1985.

[5] J. Woram, *The Recording Studio Handbook.* Plainview, New York: Sagamore, 1980.

[6] P. R. Cook, "An Analog to MIDI Converter," Senior Design Project Report, Department of Electrical and Computer Engineering, University of Missouri, Kansas City, 1986 Unpublished.

[7] C. Chafe, D. Jaffe, K. Kashima, B. Mont-Raynaud, and J. Smith, "Techniques for Note Identification in Polyphonic Music," Proceedings of the International Computer Music Conference, May 1985.

[8] J. B. Allen, D. A. Berkley, and J. Blauert, "Multimicrophone Signal-Processing Technique to Remove Room Reverberation From Speech Signals," Journal of the Acoustical Society of America, vol. 62, no. 4, pp. 912-915, October 1977.

[9] S. T. Neeley and J. B. Allen, "Invertibility of a Room Impulse Response," Journal of the Acoustical Society of America, vol. 66, no. 1, pp. 165-169, July 1979.

[10] J. Borish and J. B. Angell, "An Efficient Algorithm for Measuring the Impulse Response Using Pseudorandom Noise," Journal of the Audio Engineering Society, vol. 31, no. 7, pp. 478-488, July/August 1983.

[11] J. A. Moorer, "About This Reverberation Business," Computer Music Journal, vol. 3, no. 2, pp. 13-28.

[12] J. Borish, *Electronic Simulation of Auditorium Acoustics*. Ph.D. Dissertation, Department of Electrical Engineering, Stanford University, 1984.

[13] J. Borish, "Extension of the Image Model to Arbitrary Polyhedra," Journal of the Acoustical Society of America, vol. 75, no. 6, pp. 1827-1836, June 1984.

[14] J. B. Allen, "Image Method for Efficiently Simulating Small Room Acoustics," Journal of the Acoustical Society of America, vol. 64, no. 4, pp. 943-950, April 1979.

[15] J. Borish, "An Auditorium Simulator for Home Use," Audio Engineering Society Convention Transactions, October 1983.

[16] B. Widrow et al., "Adaptive Noise Cancelling: Principles and Applications," Proc. IEEE, vol. 63, pp. 1692-1716, December 1975.

[17] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1985.

[18] L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.

[19] E. R. Ferrara and B. Widrow, "Multichannel Adaptive Filtering for Signal Enhancement," IEEE Trans. on Circuits and Systems, vol. CAS-8, no. 6, pp. 606-610, June 1981.