Impulse Response Measurements Using All-Pass Deconvolution

David Griesinger

Lexicon, Inc.
Waltham, Massachusetts 02154, USA

A method of measuring impulse responses of rooms will be described which uses time reversed electronic reverberation from a single pulse as the excitation. The room response, recorded on DAT, is decoded by playing the tape back through the reverberator. The output can be heard, recorded, or analyzed. The method improves S/N by 20 dB and can be implemented with approximately 14 multiples/sample on a PC.

Introduction

It is often desirable to measure the impulse response of a particular system. In many cases the best way to do this is to apply a signal to the system which is close to a delta function - a click - and to measure the resulting output. In room acoustics this is traditionally done by using a blank pistol as the sound source, and recording the impulse response on a tape recorder. In seismology an explosive charge is often used for the same result. A problem with explosives is that they are not very repeatable, and the spectrum of the impulse that they produce is frequently complicated. In addition the high peak power they produce can present problems for the system under test, creating nonlinearities which can change the result of the test. For example, in a hall with an electronic reverberation system the high peak pressure of the pistol will probably overload the electronics, and the response of the electronics will be much lower than if music were used as an excitation.

If we try to replace the explosives with a mechanical or electronic transducer which produces a pulse we have the problem that the peak energy required is too high. If we run the transducer at some reasonable peak power we may not get enough total energy into the system to make useful measurements.

For these reasons, and many others, impulse responses are often measured indirectly - through a technique which employs a stimulus which is not an impulse itself, but which is sufficiently deterministic that the true impulse response can be calculated by some form of convolution between the stimulus and the response.

All indirect methods for measuring impulse responses attempt to increase the power which can be supplied to the system under test by stretching the stimulus out in time, and then using some mathematical technique to remove the artifacts of such a time stretching. All these techniques must make some (sometimes incorrect) assumptions about the system under test, particularly the assumption that the system is LINEAR - that is if we double the strength of the stimulus we will double the strength of the response. The assumption of linearity is essential to the deconvolution process. Another explicit assumption is that the system is time stationary during the length of the stimulus.

TDS and convolution techniques

For example, a linearly increasing swept sine wave can be employed as a stimulus. If we simply multiply the stimulus and response together and integrate the product with some time window the result will be the impulse response of the system under test. This is the technique used in time delay spectrometry (TDS) as developed by Heyser and Crown.

TDS systems require that the length of the stimulus be much longer than the length of the response of the system. Thus if a room being measured has a reverberation time RT of one second, the sine must sweep quite slowly - compared to one second. The system must be time stationary over the length of time of the integration window.

The deconvolution of the swept signal into an impulse response requires very little computation. However any non-linearities in the system under test will result in a reduction of apparent signal to noise ratio.

As another example of a swept sine technique, a swept sine can also be deconvolved into an impulse response through deconvolution (phase shift) techniques as shown by Berkhout. This technique has been extensively used in seismology and in room acoustics. Berkhout, Gade, and Yamasaki have used this technique. One of the major advantages of using a swept sine as a stimulus is that a transducer can be built which produces a constant amplitude signal. There is no difference between the peak power and the average power that the amplifier needs to supply to the transducer. In seismology the transducer can even be made with a motor driving the ground through a rotating weight or some form of crank. There are however several disadvantages to swept sine techniques.

With convolution techniques we need not use a swept sine as an input signal. Gaussian noise can be also effective if exactly the same noise signal can be used when the response of the system is deconvolved. We can even use recorded music as a stimulus in room acoustics, and deconvolve the response with the original recording. The amount of processing required is very large. (This technique does not work well with live music because of the extended nature of musical sources.)

Systems which use convolution to obtain the impulse response can use a sweep of any length, with longer sweeps resulting in a higher average power into the system under test, and correspondingly greater signal to noise ratio in a noisy environment. The disadvantage here is the need for a great deal of processing power to obtain the impulse response, and one must assume linearity and time invariance.

MLS - Maximum Length Sequences

Schroeder has found that a noise signal derived from a maximal length sequence (MLS) has some particularly desirable properties as a stimulus. An MLS sequence is generated by a feedback shift register which is clocked at a frequency high compared to frequencies of interest. The output is an apparently random sequence of ones and zeros. The pattern repeats with a period given by the clocking frequency divided by 2 to the power of the length of the shift register. The resulting output has the spectrum of white noise.

The advantage of an MLS sequence as a stimulus is that since the stimulus originates as only ones and zeros, the multiplies required to deconvolve the response can be replaced by additions. An algorithm for this deconvolution similar to the fast Fourrier transform has been developed, called the fast Hadamard transform. This algorithm can run quite quickly in a high powered PC, and several commercially available systems measure room acoustics this way.

MLS systems use a noise-like signal as a stimulus. The transducer must be capable of a linear response to a signal with a high peak to average ratio. The stimulus is also usually made to have a white frequency spectrum, which can cause severe overload problems with the amplifier or transducer at high frequencies if sufficient power is to be obtained at low frequencies. In general room noise tends to be either pink or brown in spectrum, and a stimulus of white noise is not a good match. A pink spectrum has equal energy in each octave band, and is about 2OdB louder at IOOHZ than a white spectrum of equal power. It is possible to pass the output of the MLS generator through a pink filter before exciting the room, but there may be undesirable signal to noise consequences when such a signal is deconvolved unless it is depinked first.

The major problem with MLS technique is the assumption of linearity and time independence. The stimulus sequence must be at least as long as the system response. This means the system must be time stationary over the entire duration of the sequence. Rooms are known to contain air currents which can significantly alter the exact standing wave patterns toward the end of the decay-. Air currents may play a role in the limitation of the S/N ratio of an MLS measurement, particularly in large rooms.

Problems with air currents or other movement in the room can be reduced by using a shorter stimulus sequence, as long as the sequence is long enough to include the whole decay of interest. Longer stimuli will give a higher signal to noise ratio if the system is truly stationary, and if S/N is limited by acoustic noise and not by problems with linearity. However in our experience with MLS in unoccupied halls adequate acoustic signal to noise is seldom a problem. Usually the apparent signal to noise ratio is limited to 30 to 5OdB depending mostly on the amount of direct sound in the impulse response. Changing the sequence length or the acoustic power from the transducer has little effect on the S/N achieved. The limit to the S/N appears to be artifacts due to non-linearity or air movement.

Schroeder has shown that in a simple system which is dominated by one impulse - such as a room response close to the source position - the affect of non linearity is to produce artifacts which appear as spurious reflections at particular time delays. With a complex impulse response, such as that from a hall well beyond the critical distance, such spurious reflections blend into a continuous background noise which cannot be reduced by increasing the length of the stimulus, or by averaging several measurements.

In our experience with MLS methods the apparent S/N in the deconvolved output depends strongly on the complexity of the impulse response. If one simply connects the output of an MLS measurement system to its input, a very sharp impulse results, with on the order of 8OdB S/N. When one applies the stimulus to a loudspeaker the loudspeaker non linearity is added to the electrical non linearity of the converters, and the S/N is degraded. It is important to realize that with MLS there is only a certain amount of signal energy available in the deconvolution. The noise energy is also relatively constant, and depends on the overall amount of non linearity. *If* the signal (the impulse response) is spread over a large amount of time its apparent amplitude relative to the noise level will decrease. Schroeder has shown S/N ratios in real halls of 40 to 5OdB, and these values seem to be typical of MLS measurements made by John Bradley and others. However, far back in a hall where the impulse response is particularly broad the apparent S/N can be lower, in the 3OdB range.

The problem with MLS is that you are trying to detect a signal which is quite small - the decaying tail of reverberation in a room - in the presence of a much louder signal. In MLS the signal and the decay from that signal are both present in the room at the same time, and you are relying on the linearity of the system to be able to sort them out. It is not helpful that both the stimulus and the impulse response have spectra which are similar to noise.

The traditional method of measurement - the pistol - has no such problem. The excitation is quite brief. Once it is gone the response of the system can be measured with a device of low linearity but high signal to noise ratio, such as an analog cassette recorder. The pistol response is also unaffected when the system is time-variant. You will accurately record the response of the system at the time the measurement was made. If you want to know the average response of the system you can shoot the pistol several times and average the energy in the resulting curves.

MLS can be improved by adjusting the bandwidth of the stimulus. As mentioned above substituting a pink stimulus for a white one is not necessarily an improvement. In the presence of non-linearities in the system it is possible that the distortion components of a pink room response will produce artifacts which have a white frequency spectrum upon deconvolution. Thus artifacts due to the non-linearity would tend to preferentially contaminate data at higher frequencies. When the deconvolved impulse response is "de-pinked" these added noise components would be further enhanced, causing a decrease in effective S/N.

As with all broad-band systems the affect of non-linearity on the results can be reduced by reducing the bandwidth of the system. If one octave band filtered data is desired for example, the stimulus to the loudspeaker can be filtered with a one octave filter, and the response of the room similarly filtered before it is deconvolved. The filtering will remove many of the spurious harmonics and cross products before they can degrade the S/N of the deconvolution. So far as I know commercially available MLS systems do not do this. They attempt to measure a broad-band impulse response and then filter it into bands later. It is not possible to increase the S/N by filtering the room response from a broad band excitation before deconvolution. Much of the non-linearity is likely to be produced in the loudspeaker, and this will already be present in the room response before filtering takes place.

The All-Pass Technique

For many reasons it would be useful to have an alternative measurement method which is not strongly dependent on system linearity, is kind to the transducers, and is quick to compute. The all-pass technique is such a method. It uses a particular type of filter to stretch a stimulus pulse out in time, in such a way that compressing it back is mathematically simple. The amount of stretching is essentially arbitrary, which allows a great deal of flexibility in choice of stimulus length versus maximum power.

All Pass Filters

An all pass filter affects only the phase of the input.   The output of the filter has the same frequency spectrum as the input signal.   The time delay of different frequencies however can be made different. For example a pulse applied to a bank of analog phase shift networks will emerge as a chirp - a type of swept sine. Any all-pass filter has an interesting property. If we record the output of the filter and then turn the tape over and play it backwards back into the filter the time dispersion caused by the filter will be exactly removed, Another way of saying this is if the filter affects only the time that different signals emerge from it, then if we run time backwards the filter will exactly undo itself.

The Schroeder all pass network has some particular advantages as a time stretcher. As you can see from figure I the Schroeder all pass consists of a delay line with both positive feedback and negative feedforward. When the right balance between feedback and feedforward is made the response becomes precisely all pass - that is there is no effect on the long term frequency response of a signal passing through. In figure I this balance is maintained for any absolute value of $g < 1$. The time response of such an all-pass with positive g is a sharp impulse in the negative direction, followed by a decaying series of impulses in the positive direction. The output looks and sounds like a form of reverberation. Several Schroeder allpass networks can be placed in series to develop a complex reverberant pattern. If we adjust the gain of the networks so that the negative and the positive peaks are equal in amplitude the network will maximally reduce the peak power of an impulse. This occurs when

$$G^2 + g - 1 = 0,$$
which implies $g = 0.62$

If we put several networks in series, each network will decrease the peak value of the impulse by 0.62. For '5 networks in series, this is a decrease in peak height of 26dB. The actual decrease is somewhat less than this, because there is a probability of some of the later impulses overlapping, and causing the height to increase again. Figure 2 shows the response of a seven segment allpass filter (g=0.62, delays increasing by pi/2 each section) to a single sample impulse. As you can see there is some redundancy. This signal was generated by a C language program operating on a single sample click. If a pink pulse is used overlap becomes more likely. To form a room stimulus the signal shown in figure 2 is reversed in time by another C program, and played into the room through a loudspeaker.   When the stimulus is sent through an Ariel DSP 32C board, recorded on a SONY D3 DAT recorder, played back into the Ariel board, and decoded by the allpass program the result is shown in figure 3. Notice there are discrete artifacts about 55dB below the peak of the impulse due to lack of linearity in the converters. Ringing in the anti aliasing filters is also visible.

As can be seen from figure I the allpass network can be implemented with little computation in a computer. The form Schroeder uses requires three multiplies per network, but both two multiply and one multiply versions of the network exist. See Moorer. The C program which implements the seven segment allpass filter uses ordinary floating point multiplies and runs in approximately two times real time on a PC with a 48kHz sample rate. On a DSP chip there are no problems running a much more complex filter in real time and in stereo.

We currently use a reverberation processor to do the deconvolution when we make measurements, because it is simpler and quicker than using the computer. generate a pink pulse using Hypersignal software on a PC. This pulse is shown in figure 4. The resulting pop is then reversed in time, and sent out the Ariel D/A converter to a Lexicon 480 which has been programmed to contain a 5 element allpass network. The output of the 480 is sampled inside the 480, and reversed in time with the 480 sampler. The sample is triggered every 4 seconds, and the output is recorded on a DAT recorder. The waveform is similar to figure 2, but backwards.

If one simply plays the DAT tape back into the 480 processor the output shown in figure 5 results. Figure 6 shows this impulse on a log scale after high pass filtering. Note the pulse is cleanly reproduced, with some noise and artifacts. The advantage to using a 480 processor is that the room response can simply be played back through the processor to deconvolve it in real time.

The time reversed filtered stimulus from the 480 is played from a DAT recorder into the room. The response of the hall is then recorded on another DAT recorder, in stereo. Typically an omni and a figure of eight microphone make one pair, and a dummy head makes another microphone pair. With a portable DAT recorder and two sets of microphones a great deal of hall data can be recorded in a short period of time. I walk around the room, making binaural and soundfield measurements at various points. The speaker is then moved, and the process is repeated. Since the excitation we use is relatively brief - 0.75 seconds or so - the microphones can be hand-held with no loss in S/N ratio.

The hall data on the DAT is played back into the 480 and deconvolved in real time. The resulting decoded impulses can be listened to, entered into a computer for analysis, or recorded on another DAT recorder for distribution or later use. The deconvolved tape is quite interesting to hear. It sounds as if the room is being excited by repeated pistol shots. Any room noise or voices are smeared out in time, although with some care the voice announcements on the tape can be understood. With a pink pulse and an equalized speaker the room response has a lot of bass, which sounds wonderful on playback.

Advantages of the allpass technique

The allpass technique allows the stimulus to be of any length. If the stimulus has an effective reverb time which is shorter than the reverb time of the room you do not get into problems from the linearity of the transducers. With no room response artifacts of the non linearity appear as spurious reflections both before and after the main pulse. These artifacts are not a problem in practice, since they can be edited away from the beginning of the impulse response, and they are usually swamped by the room response once the impulse has started. These artifacts are typically 40 to 5OdB below the main pulse. They are clearly audible on the decoded tape, where they sound like the person shooting the gun draws breath just before firing. Figure 7 shows a binaural impulse response recorded in a small chapel. The build up of the artifacts before the next pulse arc clearly visible. For a particularly simple room response - a small room close to the speaker - the artifacts sound discrete and rather grating. Once again however they can be edited out.

Thus by using a stimulus which is shorter than the room response we get data which is limited in S/N by the power we can put into the room, and not by the linearity of the measurement system or the stability of the room. If the stimulus is shorter than the speed with which air currents cause the room response to become chaotic air movement will also produce no error. The allpass system is capable of producing impulse response curves with more range than MLS systems, and rooms with subtle but audible dual-slope decays can be measured.

The data shown in figures 8 and 9 were recorded with an early version of the system where no attempt was made to optimize the power delivered to the loudspeaker. A white spectrum excitation was passed with no equalization in the loudspeaker. The S/N is adequate for many purposes.

Naturally since the stimulus must be shorter than the room response the power which can be put into the room will be limited by the power the transducer can supply during the stimulus. The transducer must also be able to handle a relatively large peak to average power, just as in MLS. In the current system a pink spectrum (obtained from using a pink pulse as an input) can be fed either directly into a 500 watt amplifier, or through an equalizer which gives the speaker a flat power response from 100 to lOkHz. We are using a dodecahedral speaker with 12 5" drivers. With this system we are able to obtain signal to noise ratios at 250 to 500 Hz of about 65dB 100 feet back in a large concert hall. The speaker has its main resonance at about 30OHz, and the sound is very loud at this frequency with a pink excitation. Schroeder integrals at this location in this hall look flat to at least -5OdB. When the speaker is equalized for flat response from IOOHZ to lOkHz the S/N in the 30OHz band is about lOdB worse, with Schroeder integrals accurate to about -4OdB. Figure 10 shows a Schroeder integral in this hall from the unequalized speaker. The equalized speaker gives the more generally useful data, because the accurately pink response gives third octave impulse responses with high S/N at all frequencies.

It sounds spectacular when you listen to a binaural impulse response through headphones! It should be possible to increase the signal to noise ratio by averaging many measurements, as long as the room remains stationary. We have had no need to implement this yet, since the current system gives good results with single measurements.

The equalized speaker produces pink impulse response measurements of high quality.

It is possible to shift these responses back to a white frequency response with a filter, and then use them to convolve speech or music. The results are quite natural, with very little sound characteristic of the loudspeaker, and no contamination f rom noise at low f requencies.

The current stimulus is about .75 seconds long. It is possible to make a much shorter impulse when it is desired to measure a room with a time variant reverberation system. However there will be a penalty in S/N.

Stimulus Acceptability

A disadvantage of the current allpass system is that the stimulus is rather grating. With a 500 watt amplifier the loudspeaker is quite loud. Something about reverse reverberation sounds particularly unnatural, and the gradual shift from pink noise to an increasingly loud signal with decreasing time diffusion is quite unpleasant. We could reverse the system by playing the stimulus forward and deconvolving the response backward. We have not needed to do this yet, but I suspect the general public would prefer the sound.

The stimulus could be made more acceptable by making it noise like and continuous. We could do this with a very long allpass network excited with multiple impulses and fed to a circular buffer. The room response could be also integrated into a circular buffer. With this system an impulse response could be measured in an occupied hall. Such a system has no theoretical advantage over a similarly repeated MLS system, but might be easier to implement. Since the stimulus would be longer than the response of the room the same linearity problems would occur for both methods.

Speed Accuracy

A minor disadvantage of the All-Pass technique is the need to maintain speed accuracy in the recorders. Since you are deconvolving a signal of about I second in length with a sample rate of 48kHz you should be accurate in speed to about 2 parts in 100,000. This accuracy is well within the stability of DAT recorders, but is beyond their variation from unit to unit. Thus the same DAT that recorded the stimulus must be used to play it back, and the DAT which records the room response must be used to play it back into the decoder. This has not been a problem in practice. If there is a mix-up somewhere the data is still recoverable. If a playback machine with an adjustable clock is available (and the setability of the clock is fine enough) it is possible to adjust the sampling rate by ear for best results. A tape from an unknown DAT can be decoded in this way.

Mathematical Accuracy Needed

The version of the all-pass program we are using in the 480 was not written to take maximal advantage of the scaling of the fixed-point calculations. However artifacts from round off error are typically much lower than artifacts due to non linearities in the loudspeaker and the converters. The 480 is an 18 bit fixed point processor. As can be seen from the figures artifacts and not noise limit the system even when only the A/D and D/A converters are used. The program written in *C* uses floating point for the deconvolution and is essentially noise free. When an impulse is deconvolved without A/D or D/A conversion there are no artifacts at all at the 16bit level, and an input pulse of one sample is exactly reproduced.

Conclusions

A method *of* impulse response measurement has been described which has some useful advantages when compared to current techniques. It uses a stimulus which can be stretched to be of any length, which allows short high amplitude stimuli to be used in situations where time variance is a problem. Since the stimulus can be shorter than the response, loss of S/N due to non-linearity can be avoided. This allows impulse response with unusually wide dynamic range to be made. The technique allows complete deconvolution with very simple mathematics, which means any PC with a floating-point processor can deconvolve the room response in near real time, even at a digitization rate of 25 to 48kHz. With a DSP co-processor or a reverberation processor such as the Lexicon 480 deconvolution can be done in stereo and in real time at 48kHz.

It might be possible to supply a version of the Lexicon 480 software or a copy of the C program to researchers interested in using this technique.

References

M. Schroeder, "Natural Sounding Artificial Reverberation" JAES 10(3) 219-223 (1962)

A. J. Berkhout, D. de Vries, M.M. Boone "A New Method to Acquire Impulse Responses in Concert Halls" JASA 68(l), July 1980 pl79-183

H. Alrutz, M. Schroeder "A Fast Hadamard Transform Method for the Evaluation of Measurements using Pseudorandom Test Signals" Ilth ICA Paris 1983

J. Borish "Electronic Simulation of Auditoria" CCRMA report STAN-M-18 Stanford University May 1984

H. Tachibana, Y. Yamasaki, et.al. "Acoustic survey of auditoriums in Europe and Japan" J. Acoust. Soc. Japan 10, 2 (1989) P73-85

J. Moorer "About this Reverberation Business" Computer Music Journal v3 #2 P13-28 (1979) Reprinted in Foundations of Computer Music ed Curtis Roads and John Strawn, MIT press 1985 pp 605-639

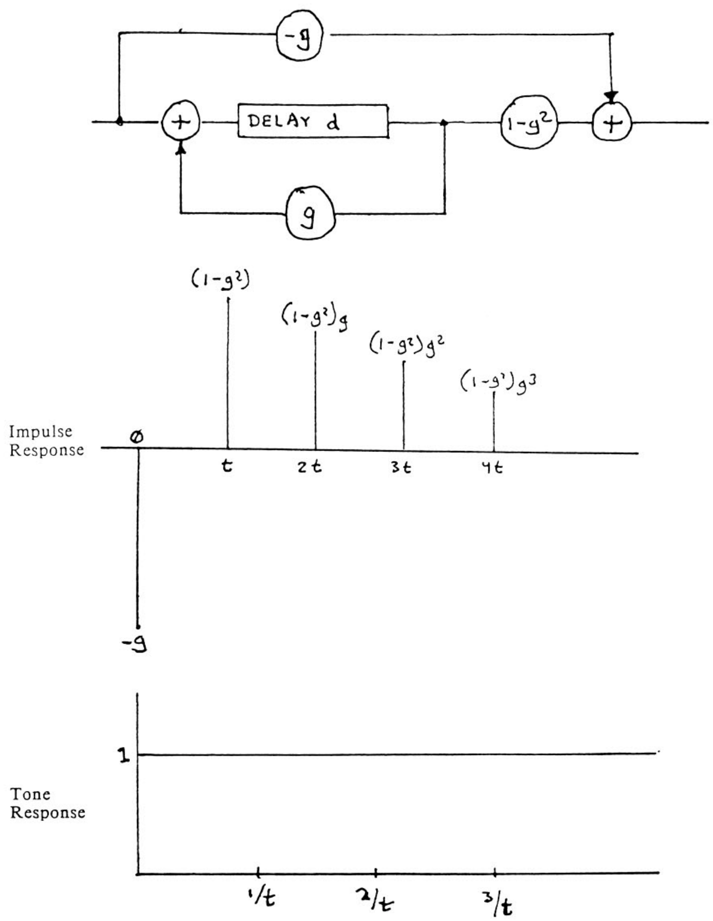

Figure 1 Schroeder All-Pass network in a three-multiply form, with the impulse response and frequency response.

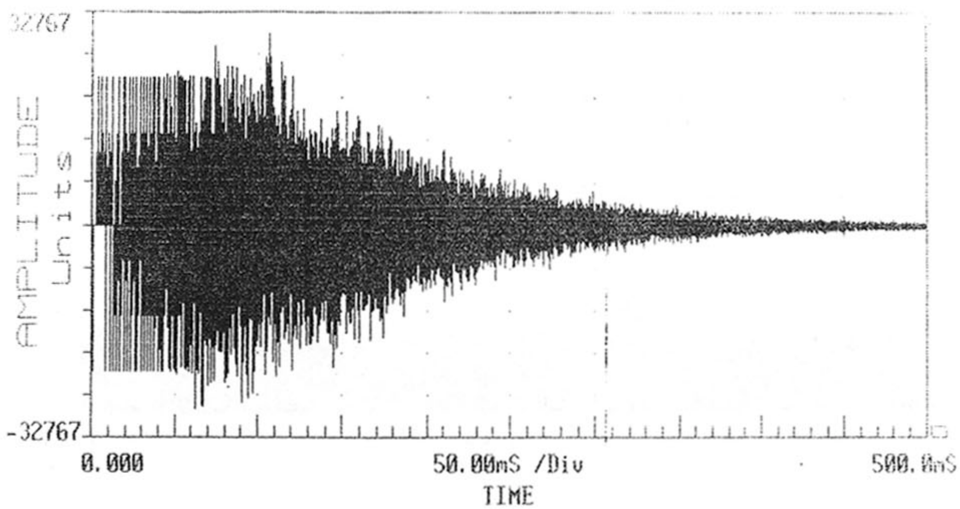

Figure 2 Response of a seven section all-pass network to a single sample input pulse. 50ms per major horizontal division, linear vertical scale. This signal is reversed in time and played into the room under test

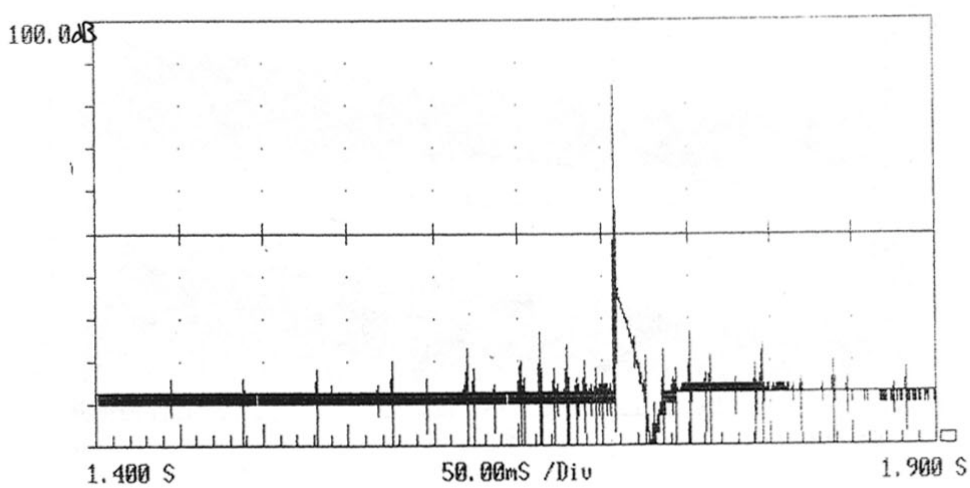

Figure 3 The result of decoding the signal in figure 2. The signal was d/a converted by an Ariel DSP32C and recorded on a SONY D3 DAT. It was then played back into the Ariel and decoded. DC offset and various artifacts due to non-linearity can be seen. Vertical axis is 10dB per step, 50ms per major division horizontal.

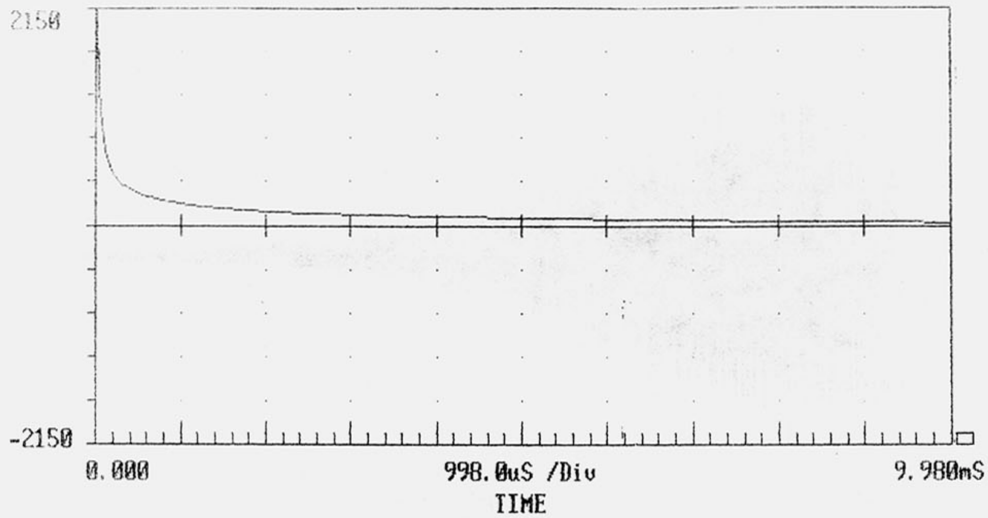

Figure 4 A single sample click after passing through a pink filter. 1ms per major division horizontal, linear vertical scale. This pop is reversed in time and used as an input to a 5 section all-pass in a Lexicon 480.

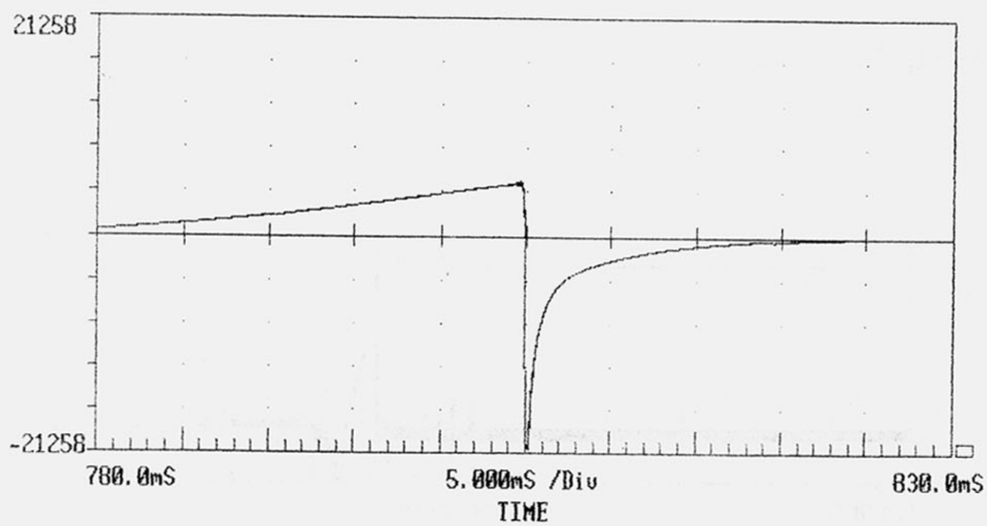

Figure 5 The result of deconvolving the 480 signal after recording and playing back on a DAT recorder. Linear vertical scale, 5ms per major division horizontal

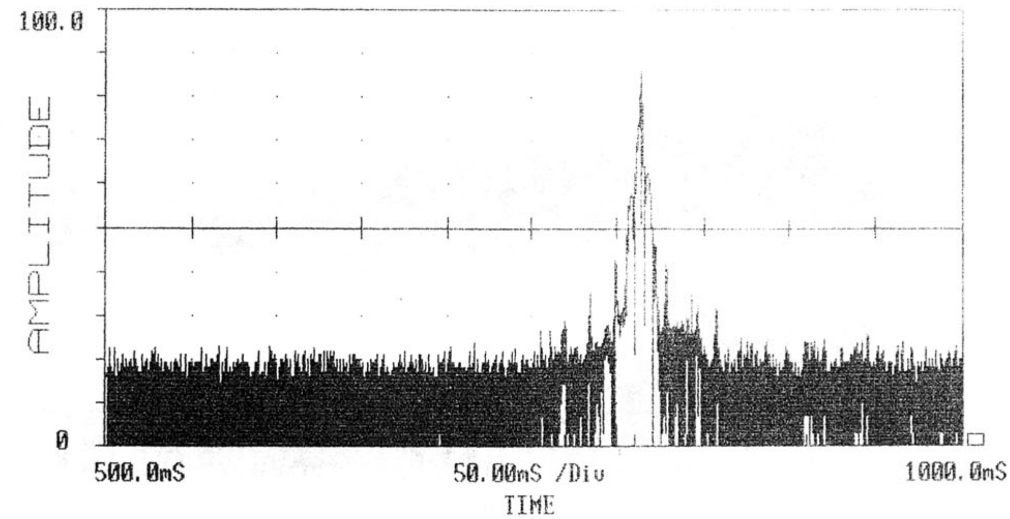

Figure 6 The same signal as figure 5 on a Log scale, 10dB per vertical division, 50ms per major horizontal division

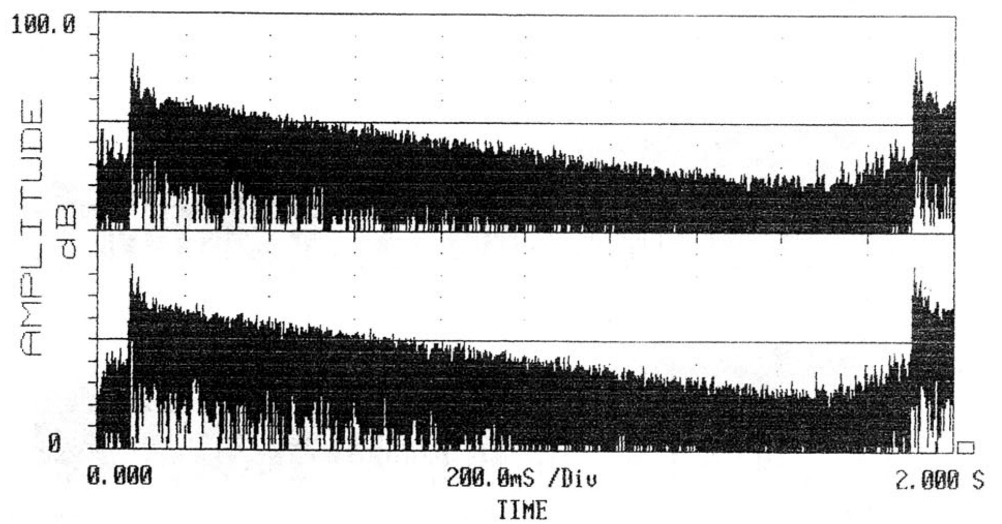

Figure 7 Binaural impulse response of Lindsay Chapel Microphone 15 feet from loudspeaker. Note the build up of energy before the next pulse starts. White stimulus, 10dB per vertical division

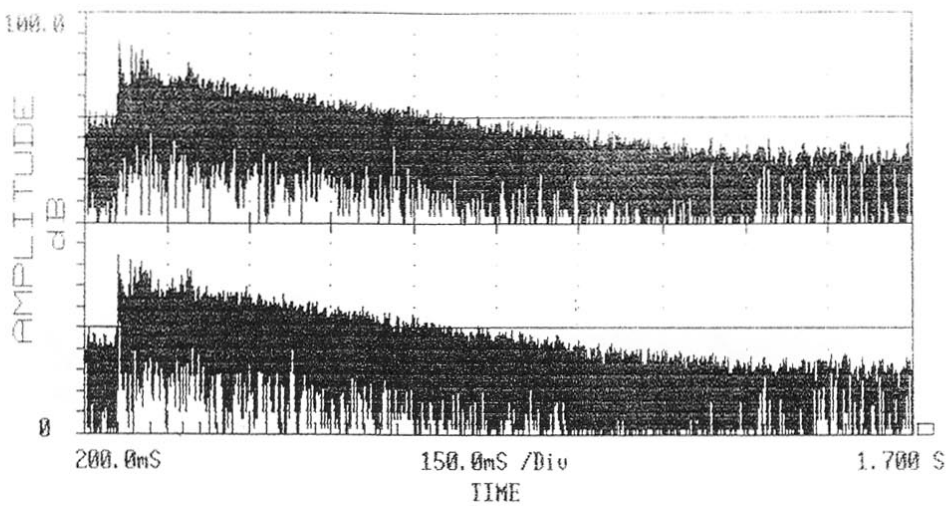

Figure 8 Binaural impulse response of Kresge Auditorium
MIT, row N. Unoccupied. White stimulus
10dB per division

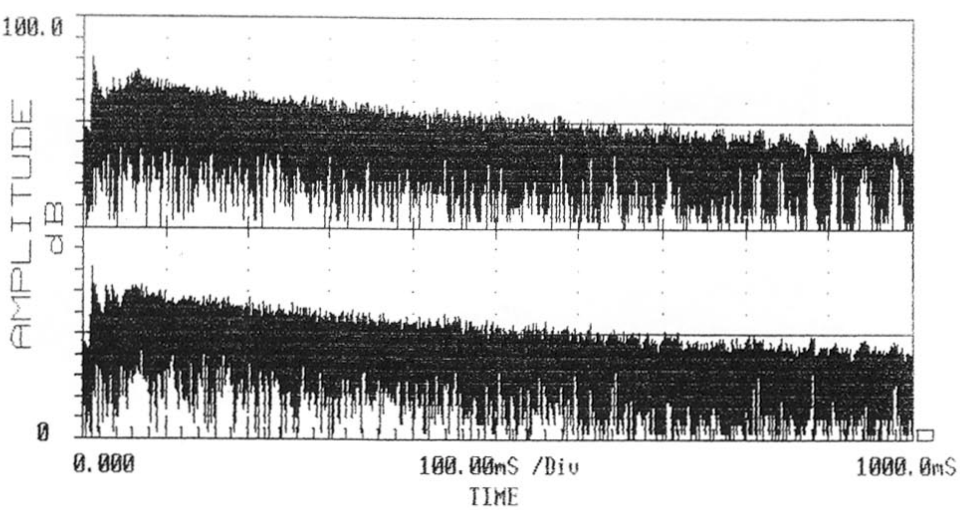

Figure 9 Binaural impulse response of Jordan Hall, balcony
front. Unoccupied. White stimulus
10dB per division

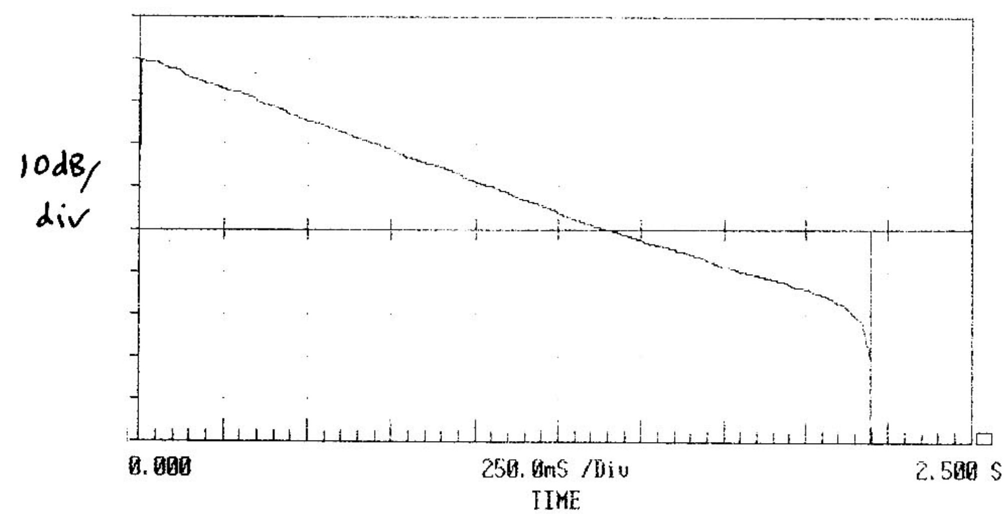

Figure 10 Schroeder integral from an impulse response made in the 250-1.2kHz band with an unequalized loudspeaker and a pink stimulus. Note the > 50dB range. 10dB per vertical division, 250ms per major horizontal division.